

Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto

Universidade de São Paulo

**Avaliação de Arredondamento de Valores de Atributos  
Contínuos em Diferentes Paradigmas de Aprendizado de Máquina<sup>1</sup>**

**Rogério Nunes Lemos  
José Augusto Baranauskas**

RELATÓRIOS TÉCNICOS DO  
DEPARTAMENTO DE FÍSICA E MATEMÁTICA  
DA FFCLRP-USP

Ribeirão Preto  
Junho/2007

---

<sup>1</sup>Título do projeto originalmente submetido à FAPESP: Avaliação de Arredondamento de Valores de Atributos Contínuos na Indução de Árvores de Decisão

# Avaliação de Arredondamento de Valores de Atributos Contínuos em Diferentes Paradigmas de Aprendizado de Máquina

**Rogério Nunes Lemos**<sup>1,2</sup>  
rnlemos@fmrp.usp.br

**José Augusto Baranauskas**<sup>1</sup>  
augusto@usp.br

<sup>1</sup>Universidade de São Paulo  
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto  
Departamento de Física e Matemática  
Avenida do Café, 3900  
14040-901 - Ribeirão Preto, SP - Brasil

<sup>2</sup>Universidade de São Paulo  
Faculdade de Medicina de Ribeirão Preto  
Avenida do Café, 3900  
14049-900 - Ribeirão Preto, SP - Brasil

---

**Resumo:** Neste trabalho é avaliado o comportamento de diversos paradigmas de Aprendizado de Máquina utilizando arredondamento de valores de atributos contínuos. São analisados cinco indutores em dez conjuntos de exemplos, utilizando arredondamento de 90%, 80%, . . . , 10% de valores do conjunto original, totalizando 441 experimentos (visto que para 1 conjunto de exemplos não foi possível realizar o experimento em 1 dos 5 indutores). Os principais resultados desta pesquisa comprovam que, em geral, há uma redução no tempo de indução. Entretanto, a redução do tempo pode ser acompanhada por um aumento tanto na taxa de erro como no tamanho do classificador induzido. Embora não seja possível aconselhar o arredondamento de valores para todas as situações, para situações individuais sua utilização pode trazer benefícios quanto à redução da taxa de erro e tamanho do classificador.

---

Este documento foi preparado com o formatador de textos  $\text{\LaTeX}$ . O sistema de citações de referências bibliográficas utiliza o padrão *Chicago* do sistema  $\text{\BIBTeX}$ .

Este projeto de iniciação científica conta com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo — FAPESP — sob número 04/10277-0.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Conjuntos de Exemplos</b>	<b>2</b>
<b>3</b>	<b>Indutores</b>	<b>3</b>
<b>4</b>	<b>Algoritmo de Arredondamento</b>	<b>6</b>
<b>5</b>	<b>Metodologia Experimental</b>	<b>10</b>
<b>6</b>	<b>Resultados</b>	<b>11</b>
6.1	Resultados do indutor IB . . . . .	11
6.1.1	Resultados aml-all-completo . . . . .	11
6.1.2	Resultados eucalyptus . . . . .	12
6.1.3	Resultados hepatitis . . . . .	14
6.1.4	Resultados ionosphere . . . . .	15
6.1.5	Resultados lymph . . . . .	17
6.1.6	Resultados sick . . . . .	18
6.1.7	Resultados sonar . . . . .	20
6.1.8	Resultados vehicle . . . . .	22
6.1.9	Resultados vowel . . . . .	23
6.1.10	Resultados wine . . . . .	25
6.1.11	Discussão . . . . .	26
6.2	Resultados do indutor J48 . . . . .	28
6.2.1	Resultados aml-all-completo . . . . .	28
6.2.2	Resultados eucalyptus . . . . .	30
6.2.3	Resultados hepatitis . . . . .	32
6.2.4	Resultados ionosphere . . . . .	34
6.2.5	Resultados lymph . . . . .	36
6.2.6	Resultados sick . . . . .	38
6.2.7	Resultados sonar . . . . .	40
6.2.8	Resultados vehicle . . . . .	42
6.2.9	Resultados vowel . . . . .	44
6.2.10	Resultados wine . . . . .	47
6.2.11	Discussão . . . . .	49
6.3	Resultados do indutor JRip . . . . .	51
6.3.1	Resultados aml-all-completo . . . . .	51
6.3.2	Resultados eucalyptus . . . . .	54
6.3.3	Resultados hepatitis . . . . .	56
6.3.4	Resultados ionosphere . . . . .	58
6.3.5	Resultados lymph . . . . .	60
6.3.6	Resultados sick . . . . .	62
6.3.7	Resultados sonar . . . . .	64
6.3.8	Resultados vehicle . . . . .	66
6.3.9	Resultados vowel . . . . .	68
6.3.10	Resultados wine . . . . .	70
6.3.11	Discussão . . . . .	72
6.4	Resultados do indutor MLP . . . . .	74
6.4.1	Resultados aml-all-completo . . . . .	74
6.4.2	Resultados eucalyptus . . . . .	74

6.4.3	Resultados hepatitis	76
6.4.4	Resultados ionosphere	78
6.4.5	Resultados lymph	79
6.4.6	Resultados sick	81
6.4.7	Resultados sonar	82
6.4.8	Resultados vehicle	84
6.4.9	Resultados vowel	85
6.4.10	Resultados wine	87
6.4.11	Discussão	88
6.5	Resultados do indutor Naïve Bayes	90
6.5.1	Resultados aml-all-completo	90
6.5.2	Resultados eucalyptus	92
6.5.3	Resultados hepatitis	93
6.5.4	Resultados ionosphere	95
6.5.5	Resultados lymph	96
6.5.6	Resultados sick	98
6.5.7	Resultados sonar	99
6.5.8	Resultados vehicle	101
6.5.9	Resultados vowel	103
6.5.10	Resultados wine	104
6.5.11	Discussão	106
<b>7</b>	<b>Discussão Geral</b>	<b>108</b>
<b>8</b>	<b>Considerações Finais</b>	<b>111</b>
	<b>Referências</b>	<b>111</b>

## Lista de Figuras

1	Modelo do MLP	6
2	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) aml-all-completo	11
3	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) aml-all-completo	12
4	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) eucalyptus	13
5	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) eucalyptus	13
6	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) hepatitis	14
7	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) hepatitis	15
8	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) ionosphere	16
9	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) ionosphere	16
10	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) lymph	17
11	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) lymph	18

12	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) sick . . . . .	19
13	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) sick . . . . .	20
14	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) sonar . . . . .	21
15	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) sonar . . . . .	21
16	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) vehicle . . . . .	22
17	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) vehicle . . . . .	23
18	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) vowel . . . . .	24
19	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) vowel . . . . .	24
20	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) wine . . . . .	25
21	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) wine . . . . .	26
22	Resumo dos resultados do tempo de indução utilizando base 2 de IB . . . . .	27
23	Resumo dos resultados do tempo de indução utilizando base 10 de IB . . . . .	27
24	Resumo dos resultados da taxa de erro utilizando base 2 de IB . . . . .	27
25	Resumo dos resultados da taxa de erro utilizando base 10 de IB . . . . .	28
26	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) aml-all-completo . . . . .	29
27	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) aml-all-completo . . . . .	29
28	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) aml-all-completo . . . . .	30
29	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) eucalyptus . . . . .	31
30	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) eucalyptus . . . . .	31
31	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) eucalyptus . . . . .	32
32	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) hepatitis . . . . .	33
33	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) hepatitis . . . . .	33
34	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) hepatitis . . . . .	34
35	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) ionosphere . . . . .	35
36	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) ionosphere . . . . .	35
37	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) ionosphere . . . . .	36
38	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) lymph . . . . .	37

39	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>lymph</i> . . . . .	37
40	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) <i>lymph</i> . . . . .	38
41	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sick</i> . . . . .	39
42	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sick</i> . . . . .	39
43	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) <i>sick</i> . . . . .	40
44	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sonar</i> . . . . .	41
45	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sonar</i> . . . . .	41
46	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) <i>sonar</i> . . . . .	42
47	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vehicle</i> . . . . .	43
48	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vehicle</i> . . . . .	43
49	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) <i>vehicle</i> . . . . .	44
50	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vowel</i> . . . . .	45
51	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vowel</i> . . . . .	46
52	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) <i>vowel</i> . . . . .	46
53	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>wine</i> . . . . .	47
54	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>wine</i> . . . . .	48
55	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) <i>wine</i> . . . . .	48
56	Resumo dos resultados do tempo de indução utilizando base 2 de J48 . . . . .	49
57	Resumo dos resultados do tempo de indução utilizando base 10 de J48 . . . . .	50
58	Resumo dos resultados da taxa de erro utilizando base 2 de J48 . . . . .	50
59	Resumo dos resultados da taxa de erro utilizando base 10 de J48 . . . . .	50
60	Resumo dos resultados do tamanho do classificador utilizando base 2 de J48 . . . . .	51
61	Resumo dos resultados do tamanho do classificador utilizando base 10 de J48 . . . . .	51
62	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>aml-all-completo</i> . . . . .	52
63	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>aml-all-completo</i> . . . . .	53
64	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) <i>aml-all-completo</i> . . . . .	53
65	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>eucalyptus</i> . . . . .	54
66	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>eucalyptus</i> . . . . .	55

67	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) eucalyptus . . . . .	55
68	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) hepatitis . . . . .	56
69	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) hepatitis . . . . .	57
70	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) hepatitis . . . . .	57
71	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) ionosphere . . . . .	58
72	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) ionosphere . . . . .	59
73	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) ionosphere . . . . .	59
74	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) lymph . . . . .	60
75	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) lymph . . . . .	61
76	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) lymph . . . . .	61
77	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) sick . . . . .	62
78	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) sick . . . . .	63
79	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) sick . . . . .	63
80	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) sonar . . . . .	64
81	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) sonar . . . . .	65
82	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) sonar . . . . .	65
83	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) vehicle . . . . .	66
84	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) vehicle . . . . .	67
85	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) vehicle . . . . .	67
86	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) vowel . . . . .	68
87	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) vowel . . . . .	69
88	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) vowel . . . . .	69
89	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) wine . . . . .	70
90	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) wine . . . . .	71
91	Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 <i>versus</i> conjunto original) wine . . . . .	71
92	Resumo dos resultados do tempo de indução utilizando base 2 de JRip . . . . .	72

93	Resumo dos resultados do tempo de indução utilizando base 10 de JRip . . . . .	73
94	Resumo dos resultados da taxa de erro utilizando base 2 de JRip . . . . .	73
95	Resumo dos resultados da taxa de erro utilizando base 10 de JRip . . . . .	73
96	Resumo dos resultados do tamanho do classificador utilizando base 2 de JRip . .	74
97	Resumo dos resultados do tamanho do classificador utilizando base 10 de JRip . .	74
98	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>eucalyptus</i> . . . . .	75
99	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>eucalyptus</i> . . . . .	76
100	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>hepatitis</i> . . . . .	77
101	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>hepatitis</i> . . . . .	77
102	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>ionosphere</i> . . . . .	78
103	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>ionosphere</i> . . . . .	79
104	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>lymph</i> . . . . .	80
105	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>lymph</i> . . . . .	80
106	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sick</i> . . . . .	81
107	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sick</i> . . . . .	82
108	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sonar</i> . . . . .	83
109	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sonar</i> . . . . .	83
110	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vehicle</i> . . . . .	84
111	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vehicle</i> . . . . .	85
112	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vowel</i> . . . . .	86
113	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vowel</i> . . . . .	86
114	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>wine</i> . . . . .	87
115	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>wine</i> . . . . .	88
116	Resumo dos resultados do tempo de indução utilizando base 2 de MLP . . . . .	89
117	Resumo dos resultados do tempo de indução utilizando base 10 de MLP . . . . .	89
118	Resumo dos resultados da taxa de erro utilizando base 2 de MLP . . . . .	89
119	Resumo dos resultados da taxa de erro utilizando base 10 de MLP . . . . .	90
120	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>aml-all-completo</i> . . . . .	91
121	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>aml-all-completo</i> . . . . .	91
122	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>eucalyptus</i> . . . . .	92

123	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>eucalyptus</i> . . . . .	93
124	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>hepatitis</i> . . . . .	94
125	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>hepatitis</i> . . . . .	94
126	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>ionosphere</i> . . . . .	95
127	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>ionosphere</i> . . . . .	96
128	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>lymph</i> . . . . .	97
129	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>lymph</i> . . . . .	97
130	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sick</i> . . . . .	99
131	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sick</i> . . . . .	99
132	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sonar</i> . . . . .	100
133	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>sonar</i> . . . . .	101
134	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vehicle</i> . . . . .	102
135	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vehicle</i> . . . . .	102
136	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vowel</i> . . . . .	103
137	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>vowel</i> . . . . .	104
138	Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>wine</i> . . . . .	105
139	Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 <i>versus</i> conjunto original) <i>wine</i> . . . . .	105
140	Resumo dos resultados do tempo de indução utilizando base 2 de NB . . . . .	106
141	Resumo dos resultados do tempo de indução utilizando base 10 de NB . . . . .	107
142	Resumo dos resultados da taxa de erro utilizando base 2 de NB . . . . .	107
143	Resumo dos resultados da taxa de erro utilizando base 10 de NB . . . . .	107
144	Resumo dos resultados do tempo de indução para a base 2 de todos os indutores	108
145	Resumo dos resultados do tempo de indução para a base 10 de todos os indutores	109
146	Resumo dos resultados da taxa de erro para a base 2 de todos os indutores . . .	109
147	Resumo dos resultados da taxa de erro para a base 10 de todos os indutores . . .	109
148	Resumo dos resultados do tamanho do classificador para a base 2 dos indutores J48 e JRip . . . . .	110
149	Resumo dos resultados do tamanho do classificador para a base 2 dos indutores J48 e JRip . . . . .	110

## Lista de Tabelas

1	Características dos conjuntos de exemplos . . . . .	4
2	Exemplo utilizando a Equação 1 . . . . .	7

3	Exemplo utilizando a Equação 2 na base 2 . . . . .	8
4	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>aml-all-completo</b> . . . . .	11
5	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>eucalyptus</b> . . . . .	13
6	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>hepatitis</b> . . . . .	14
7	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>ionosphere</b> . . . . .	16
8	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>lymph</b> . . . . .	17
9	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>sick</b> . . . . .	19
10	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>sonar</b> . . . . .	20
11	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>vehicle</b> . . . . .	22
12	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>vowel</b> . . . . .	23
13	Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 <b>wine</b> . . . . .	25
14	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>aml-all-completo</b> . . . . .	28
15	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>eucalyptus</b> . . . . .	30
16	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>hepatitis</b> . . . . .	32
17	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>ionosphere</b> . . . . .	34
18	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>lymph</b> . . . . .	36
19	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>sick</b> . . . . .	38
20	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>sonar</b> . . . . .	40
21	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>vehicle</b> . . . . .	42
22	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>vowel</b> . . . . .	45
23	Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 <b>wine</b> . . . . .	47
24	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 <b>aml-all-completo</b> . . . . .	52
25	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 <b>eucalyptus</b> . . . . .	54
26	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 <b>hepatitis</b> . . . . .	56
27	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 <b>ionosphere</b> . . . . .	58
28	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 <b>lymph</b> . . . . .	60

29	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 sick . . . . .	62
30	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 sonar . . . . .	64
31	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 vehicle . . . . .	66
32	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 vowel . . . . .	68
33	Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 wine . . . . .	70
34	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 eucalyptus . . . . .	75
35	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 hepatitis . . . . .	76
36	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 ionosphere . . . . .	78
37	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 lymph . . . . .	79
38	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 sick . . . . .	81
39	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 sonar . . . . .	82
40	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 vehicle . . . . .	84
41	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 vowel . . . . .	85
42	Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 wine . . . . .	87
43	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 aml-all-completo . . . . .	90
44	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 eucalyptus . . . . .	92
45	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 hepatitis . . . . .	93
46	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 ionosphere . . . . .	95
47	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 lymph . . . . .	97
48	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 sick . . . . .	98
49	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 sonar . . . . .	100
50	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 vehicle . . . . .	101
51	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 vowel . . . . .	103
52	Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 wine . . . . .	105

## Lista de Algoritmos

1	Algoritmo de arredondamento proposto por Weiss . . . . .	9
2	Algoritmo final de arredondamento . . . . .	9

# 1 Introdução

Nas últimas décadas, a computação científica e comercial vem gerando uma quantidade enorme de dados. Métodos tradicionais de manipulação de dados, tais como planilhas, consultas em bancos de dados, programas gráficos e processadores de texto são ferramentas úteis para o armazenamento, gerenciamento e a organização de dados e informações. Entretanto, quando se trata de descoberta do conhecimento existente, por exemplo, em um banco de dados, torna-se necessário recorrer a outras estratégias.

A extração semi-automática de conhecimento a partir de grandes volumes (bancos) de dados — KDD (*Knowledge Data Discovery*) — é um ramo de pesquisa em Ciência da Computação. Pesquisas nessa área têm como principais objetivos a aplicação e o desenvolvimento de técnicas e ferramentas que automatizem o processo de manipulação de dados, visando a extração de novas informações úteis. Uma das abordagens utilizada consiste em utilizar algoritmos de Aprendizado de Máquina — AM.

O Aprendizado de Máquina supervisionado é definido por Weiss and Kulikowski (1991) como “Um sistema de aprendizado é um programa de computador que toma decisões baseadas na experiência contida em exemplos solucionados com sucesso.”

No Aprendizado de Máquina supervisionado, cada exemplo  $z$  pode ser descrito por um vetor de valores de características  $x$ , ou atributos, juntamente com o rótulo da classe associada  $y$  ou seja,  $z = (x, y)$ , ficando subentendido o fato que tanto  $x$  como  $z$  são vetores, ou seja,  $\vec{z} = (\vec{x}, y)$ . Para rótulos de classe  $y$  discretos, esse problema é conhecido como *classificação* e para valores contínuos como *regressão*.

O objetivo de um algoritmo de AM, denominado *indutor*, é construir uma hipótese  $h(\cdot)$  que possa determinar corretamente a classe de novos exemplos ainda não rotulados, ou seja, exemplos que não tenham o rótulo da classe. Formalmente, em classificação, um exemplo  $z$  é um par  $(x, y) = (x, f(x))$  onde  $x$  é a entrada e  $f(x)$  é a saída e  $y = f(x)$ . A tarefa de um indutor é, dado um conjunto de exemplos da função  $f(\cdot)$ , induzir uma função  $h(\cdot)$  que aproxima  $f(\cdot)$ , normalmente desconhecida. Neste caso,  $h(\cdot)$  é chamada uma *hipótese* sobre a função objetivo  $f(\cdot)$ , ou seja,  $h(x) \approx f(x)$ .

Um ponto importante com relação à árvores de decisão é que enquanto a maior parte das operações para construí-la cresce linearmente com o número de exemplos de treinamento, o processo de escolha de um atributo contínuo contendo  $d$  valores distintos requer a ordenação desses valores, crescendo como  $d \log_2 d$  (Quinlan 1993). Assim, o tempo requerido para construir uma árvore de decisão a partir de um conjunto de treinamento grande pode ser dominado pela ordenação de atributos contínuos, por exemplo, os algoritmos C4.5 (Quinlan 1993) e J48 (Witten and Frank 1999) fazem uso do algoritmo *quicksort* para ordenar valores contínuos (Cormen, Leiserson, Rivest, and Stein 2002)[Cap. 7], (Wirth 1986)[Cap. 2].

O objetivo deste trabalho consiste na avaliação do arredondamento de valores de atributos em diferentes paradigmas de Aprendizado de Máquina, ou seja, neste trabalho é tratado tanto o aprendizado simbólico supervisionado quanto o não simbólico supervisionado para resolver problemas de classificação. O termo *simbólico* indica que os classificadores devem ser legíveis e interpretáveis por humanos. O termo *supervisionado* sugere que algum processo, às vezes denominado *agente externo* ou *professor*, previamente rotulou os dados. Finalmente, o termo *classificação* denota o fato que o rótulo da classe é discreto, ou seja, consiste de valores nominais sem uma ordem definida. Nesta pesquisa são utilizados o indutores da biblioteca Weka (Witten and Frank 1999) — Waikato Environment for Knowledge Analysis.

O restante deste trabalho está organizado da seguinte forma: Na Seção 2 são descritos os conjuntos de exemplos utilizados nos experimentos realizados. Os indutores utilizados para a realização dos experimentos são descritos na Seção 3. Na Seção 4 é mostrada uma metodologia de arredondamento proposta por Weiss and Indurkha (1998). Na Seção 5 é mostrada a metodologia utilizada no trabalho. Na Seção 6 são mostrados resultados dos experimentos,

bem com uma discussão destes resultados para cada um dos indutores. Uma discussão geral é apresentada na Seção 7. Por último, são relacionadas as Referências Bibliográficas.

## 2 Conjuntos de Exemplos

Os experimentos, reportados nas seções subseqüentes, foram conduzidos a partir de conjuntos de exemplos provenientes de diversos domínios do mundo real. Os conjuntos de exemplos *ionosphere*, *hepatitis*, *lymph*, *sick*, *sonar*, *vehicle*, *vowel* e *wine* foram obtidos a partir do repositório UCI Irvine (Newman, Hettich, Blake, and Merz 1998). O conjunto *aml-all-completo* foi obtido de Golub (1999); o conjunto *eucalyptus* foi obtido em pesquisas na agricultura da Nova Zelândia<sup>2</sup>.

A seguir é fornecida uma descrição, sobre os conjuntos de exemplos utilizados neste trabalho bem como um resumo de suas características.

**aml-all-completo** O problema consiste em distinguir entre a leucemia linfoblástica aguda (*acute lymphoblastic leukemia* - ALL) e leucemia mielóide aguda (*acute myeloid leukemia* - AML) utilizando dados de expressão gênica obtidos por monitoramento de *microarrays* de DNA. No trabalho desenvolvido por Golub (1999) o conjunto de treinamento possui 38 exemplos (27 do tipo ALL e 11 do tipo AML) e o conjunto de teste possui 34 exemplos (20 do tipo ALL e 14 do tipo AML). Todos exemplos são descritos por valores de expressão de 7129 genes. Adicionalmente, outro artigo que utiliza esse conjunto de exemplos é (Gamberger, Lavrac, Zelezny, and Tolar 2004).

**eucalyptus** O objetivo foi determinar quais lotes de sementes em uma espécie de eucalipto são os melhores para a conservação do solo na estação seca. Assim se determina a utilidade de cada exemplo para se atingir tal objetivo, podendo ser “nenhuma”, “baixa”, “média”, “boa” e “melhor”. O conjunto de exemplos possui 736 casos, destes 180 pertencem à classe “nenhuma”, 107 à “baixa”, 130 à “média”, 214 à “boa” e 105 à “melhor”, sendo que cada exemplo é representado por 19 atributos.

**hepatitis** Este conjunto de exemplos foi uma contribuição de Gail Gong (*Carnegie-Mellon University*). O estudo trata de prognósticos e não diagnósticos. Porém com o prognóstico, o diagnóstico é conhecido, e o problema é determinar o resultado provável da doença. Assim o prognóstico pode utilizar a informação conduzida pela decisão (para preparar uma análise histológica) feita durante o diagnóstico. O problema consiste em determinar se um paciente provavelmente irá “viver” ou “morrer”. O conjunto de exemplos contém 155 casos, sendo que 123 são exemplos de pacientes pertencentes à classe “viver”, e 32 pertencentes à classe “morrer”. Cada exemplo possui 19 atributos, sendo 12 com valores “sim” ou “não”, e o conjunto de exemplos possui também diversos valores perdidos.

**ionosphere** Estes dados de radar foram coletados por um sistema em Goose Bay, Labrador. Este sistema consiste de um conjunto de 16 antenas de alta frequência com uma potência total transmitida da ordem de 6,4 Kilowatts. Os alvos eram os elétrons livres na ionosfera. O problema consiste em discriminar entre os retornos “bons” do radar que são aqueles que mostram evidências de algum tipo de estrutura na ionosfera dos retornos “maus” que são aqueles que não mostram a evidências de algum tipo de estrutura na ionosfera. O conjunto de exemplos contém 225 exemplos de retornos “bons” e 126 exemplos de retornos “maus”. Cada exemplo é um vetor de 34 números reais entre -1 e 1. Dois números representam um número de pulso, que correspondem a sinais eletromagnéticos complexos.

---

<sup>2</sup><http://www.cs.waikato.ac.nz/~ml/weka/agridatasets.jar> acessado em 08/05/2007.

**lymph** Estes dados representam o domínio de uma linfografia que foi obtido na *University Medical Centre, Institute of Oncology, Ljubljana*<sup>3</sup>. E foram utilizados originalmente por *I. Kononenko* e *B. Cestnik*. O problema consiste em discriminar entre um achado normal (2 exemplos), com metástase (81 exemplos), com fibrose (4 exemplos) ou nódulo linfático com células malignas (61 exemplos), totalizando 148 exemplos. Cada um dos exemplos possui 18 atributos, sendo 15 discretos e 3 inteiros.

**sick** O conjunto de exemplos de Doenças da Tireóide foi criado em *Garavan Institute, Sydney, Australia*. O objetivo é determinar se um paciente está ou não com alguma doença na tireóide através de 29 atributos. No total são 3772 pacientes, dentro os quais 231 foram classificados como doentes e 3541 classificados com não doentes.

**sonar** Este conjunto de exemplos foi usado por Gorman and Sejnowski (1988) no estudo de classificação de sinais de sonar utilizando uma rede neural. O problema consiste em discriminar entre sinais de sonar que representam um cilindro de metal daqueles que representam uma rocha ligeiramente cilíndrica. O conjunto de exemplos contém 111 exemplos obtidos por varredura de sonar de um cilindro de metal em vários ângulos e sob várias condições; contém também 97 exemplos obtidos por varredura de rochas sob as mesmas condições. Cada exemplo é um conjunto de 60 números reais entre 0 e 1. Cada número representa a energia em uma banda de frequência particular integrada sobre um certo período de tempo. A classe associada com cada exemplo contém a letra “R” se o objeto é uma rocha e “M” se ele é uma mina (cilindro de metal).

**vehicle** Este conjunto de exemplos foi gerado em *Turing Institute, Glasgow, Scotland*. O objetivo deste trabalho foi classificar uma dada silhueta como um dos quatro tipos de veículo, usando um conjunto das características extraídas da silhueta. Foram utilizados no total 18 características dos veículo, num total 846 veículos.

**vowel** O problema consiste em reconhecer uma vogal pronunciada por um locutor arbitrário. Há dez atributos contínuos que são derivados de dados espectrais e três atributos nominais: a identidade do locutor, o sexo do locutor e um atributo adicional que indica se o locutor foi utilizado originalmente para treinar ou testar o classificador. Os exemplos são rotulados em onze classes (devido à normalização realizada). O conjunto de exemplos contém 990 exemplos e cada exemplo possui 13 atributos. Maiores detalhes podem ser obtidos em Turney (1993).

**wine** Estes dados são resultados de uma análise química dos vinhos de uma mesma região da Itália mas derivados de três produtores diferentes. A análise determinou as quantidades de 13 constituintes encontrados em cada um dos três tipos de vinhos. O conjunto de exemplos contém 178 exemplos e cada exemplo possui 13 atributos. Maiores detalhes podem ser obtidos em Forina (1991).

Na Tabela 1 são resumidas algumas características dos conjuntos de exemplos utilizados, dos 10 conjunto selecionados 5 tem predominância de atributos contínuos e 5 de atributos nominais. Para cada conjunto de exemplos são mostrados o número de exemplos (#Exemplos), número de atributos (#Atributos) contínuos ou nominais, número de classes (#Classes), o erro majoritário e se o conjunto de exemplos possui ao menos um valor desconhecido.

### 3 Indutores

A seguir é fornecida uma descrição, sobre os indutores utilizados neste trabalho.

---

<sup>3</sup>Agradecimento a *M. Zwitter* e *M. Soklic* pelos dados.

Conjunto de Exemplos	#Exemplos	#Atributos (cont.;nom.)	#Classes	Erro Majoritário	Valor Desconhecido
aml-all-completo	72	7129 (7129;0)	2	28,95%	não
eucalyptus	736	19 (14;5)	5	70,92%	sim
hepatitis	155	19 (6;13)	2	20,64%	sim
ionosphere	351	34 (34;0)	2	35,90%	não
lymph	148	19 (3;16)	4	45,27%	não
sick	3772	29 (7;22)	2	6,12%	sim
sonar	208	60 (60;0)	2	46,63%	não
vehicle	846	18 (18;0)	4	74,23%	não
vowel	990	13 (10;3)	11	90,91%	não
wine	178	13 (13;0)	3	60,11%	não

Tabela 1: Características dos conjuntos de exemplos

J48 O indutor J48 (Witten and Frank 1999), consiste de uma reimplementação em Java do indutor C4.5 (Quinlan 1993). É membro de uma família mais ampla de algoritmos de AM indutivo conhecida como *Top Down Induction of Decision Trees* – TDIDT (Quinlan 1986). É um algoritmo básico para a construção de árvores de decisão com poda, na qual é conduzida uma busca gulosa (*greedy*), ou seja, o algoritmo não reconsidera escolhas anteriores.

A construção de uma árvore de decisão realiza-se da seguinte forma (Breiman, Friedman, Olshen, and Stone 1984; Quinlan 1986): utilizando o conjunto de treinamento, um atributo é escolhido de forma a particionar os exemplos em subconjuntos, de acordo com valores deste atributo. Para cada subconjunto, outro atributo é escolhido para particionar novamente cada um deles. Este processo prossegue, enquanto um dos subconjuntos contenha uma mistura de exemplos pertencendo a classes diferentes. Uma vez obtido um subconjunto uniforme — todos os exemplos naquele subconjunto pertencem à mesma classe — um nó folha é criado e rotulado com o mesmo nome da respectiva classe.

Quando um novo exemplo deve ser classificado, começando pela raiz da árvore induzida, o classificador testa e desvia para cada nó com o respectivo atributo até que atinja uma folha. A classe deste nó folha será atribuída ao novo exemplo. O J48 usa o critério uso o de razão de ganho de informação para escolher os nós de decisão.

**Instance Based** É um indutor preguiçoso, também denominado IB (*Instance-Based*). Ele é também conhecido como algoritmo  $K$  vizinhos mais próximos (*K-Nearest Neighbors* ou  $K$ -NN) (Mitchell 1998) com  $k = 1$ . A idéia geral é postergar a compilação do conjunto de treinamento, armazenando os exemplos. A classificação de um novo exemplo é efetuada com base no voto dos  $K$  exemplos mais próximos utilizando uma métrica de distância (Aha 1992; Aha 1997).

**Naïve Bayes** Este indutor, também denominado NB, usa a regra de Bayes para calcular a probabilidade de cada classe dado um exemplo, assumindo que os atributos são independentes (Langley, Iba, and Thompson 1992; Heckerman 1996). Formalmente,

$$\begin{aligned}
p(y|\vec{x}) &= \frac{p(\vec{x}|y) \cdot p(y)}{p(\vec{x})} && \text{regra de Bayes} \\
&\propto p(\vec{x}|y) \cdot p(y) && p(\vec{x}) \text{ é igual para todas as classes} \\
&= p(x_1, x_2, \dots, x_m|y) \cdot p(y) \\
&= p(x_1|y) \cdot p(x_2|y) \cdot \dots \cdot p(x_m|y) \cdot p(y) && \text{por independência} \\
&= \prod_{j=1}^m p(x_j|y) \cdot p(y)
\end{aligned}$$

Mesmo que, em domínios reais, os atributos não sejam independentes, o algoritmo é bem robusto a violações da condição de independência. As probabilidades para atributos nominais são estimadas através de contagem. A probabilidade para uma contagem de zero é assumida, como sendo  $1/2n$  para  $n$  exemplos. As probabilidades para atributos contínuos são estimadas assumindo uma distribuição normal e calculando a média e o desvio padrão a partir dos dados. Valores desconhecidos são ignorados, ou seja, eles não participam do produtório.

**JRip** Consiste de uma reimplementação de aprendizagem de regras proposicionais, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), sendo esta proposta por Cohen (1995), como uma versão otimizada do IREP.

O algoritmo é brevemente descrito a seguir:

Inicializar  $RS = \{\}$ , e para cada classe da menos predominante para a mais freqüente, faça:

1. Fase de construção:

Repetir as etapas de construção (a) e (b) até que o tamanho da descrição ( $TD$ ) do conjunto de regras ser maior do que o menor ( $TD$ ) de regra encontrado ou não exista nenhum exemplo positivo, ou a taxa de erro for maior ou igual a 50%.

(a) Etapa de crescimento:

Adicionar antecedentes ou condições à regra até a regra ser considerada perfeita (100% de precisão). O procedimento testa todos os valores possíveis para cada atributos e seleciona a condição com maior ganho de informação.

(b) Etapa de poda:

Realizar a poda incremental de cada regra e permitir a poda de quaisquer seqüências finais dos antecedentes.

2. Fase de otimização:

Após gerar o conjunto inicial de regra  $\{Ri\}$  utilizar os procedimentos (a) e (b) para gerar e podar as regras do conjunto  $\{Ri\}$ . Todas as regras do conjunto  $\{Ri\}$  são analisadas e se existir exemplos positivos não cobertos pelas regras, novas regras são construídas seguindo as etapas anteriores.

3. As regras do conjunto  $\{Ri\}$  que aumentam o ( $TD$ ) do conjunto de regras devem ser eliminadas e as restantes compõem o conjunto resultante.

**Multilayer Perceptron** Também conhecido como (MLP) (Haykin 1998) é uma rede neural artificial treinada pelo algoritmo de *Backpropagation*. Basicamente é composta de uma camada de entrada, uma ou mais camadas escondidas, e uma camada de saída, como mostrada na Figura 1. Cada uma das camadas é composta de uma ou mais unidades simples de processamento de informação denominados neurônios. Os neurônios existentes

na camada de entrada são responsáveis por distribuir os sinais de entrada aos neurônios da camada escondida subsequente, sendo que estes, dependem de somas ponderadas das entradas, transformadas por funções de ativação, podem ser consideradas como extratoras de características. A camada de saída é onde o resultado final é concluído e apresentado.

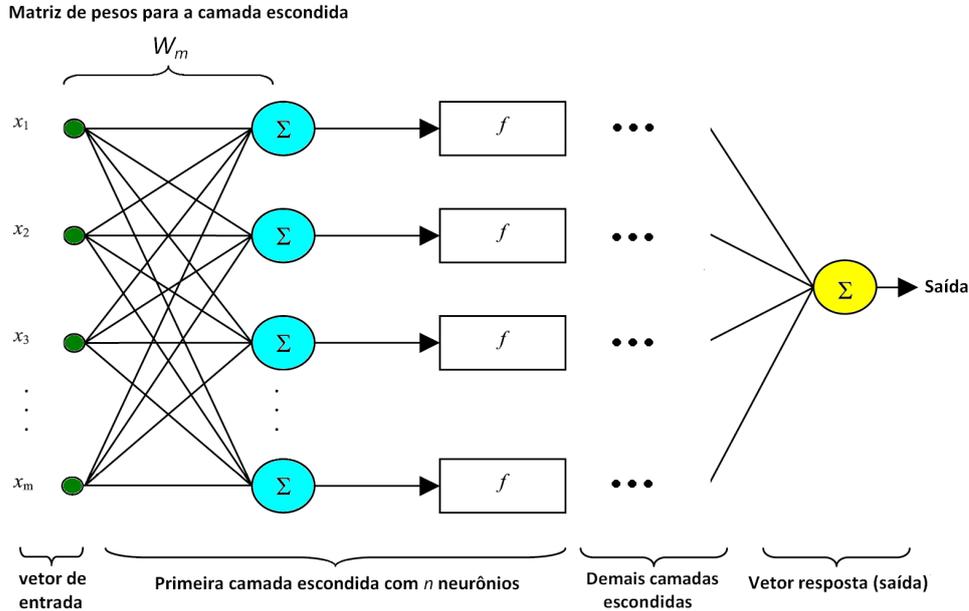


Figura 1: Modelo do MLP

O treinamento com o algoritmo *backpropagation* se da em 2 passos:

- Um exemplo é apresentado à camada de entrada da rede. A atividade resultante flui através da rede, camada por camada, até que a resposta seja produzida pela camada de saída.
- A saída obtida é comparada à saída desejada para esse exemplo particular. Se esta não estiver correta, o erro é calculado. O erro é propagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retropropagado.

Algumas condições de término do treinamento:

- Encerrar após um número máximo de iterações.
- Encerrar quando o erro de treinamento for menor que um determinado valor.
- Encerrar quando a proporção de classificações corretas numa amostra de teste estiver acima de um determinado valor.

## 4 Algoritmo de Arredondamento

Neste trabalho foi avaliada a técnica de arredondamento proposta por Weiss and Indurkha (1998), descrita em maiores detalhes a seguir.

Inicialmente, considere uma variável  $ix$  inteira a ser arredondada e o fragmento de código expresso na Equação 1 onde  $k$  é o número de casas decimais mais à direita do número a ser arredondado. A função  $int(x)$  retorna a parte inteira de  $x$  — por exemplo,  $int(3,0) = 3$ ;

$int(3, 5) = 3$ ;  $int(3, 8) = 3$  — e a função  $mod(x, y)$  corresponde ao resto da divisão inteira de  $x$  por  $y$  — por exemplo,  $mod(10, 3) = 1$ ;  $mod(10, 4) = 2$ ;  $mod(12, 5) = 2$ . Assume-se que a divisão retorna sempre um valor real, mesmo que seus argumentos sejam inteiros — por exemplo  $2/4 = 0,5$ ;  $1/4 = 0,25$ . A variável  $iy$  é inteira.

$$\begin{aligned}
 iy &\leftarrow int(ix/10^k) \\
 \mathbf{if}(mod(ix, 10^k) \geq 10^k/2) &\mathbf{then} \ iy \leftarrow iy + 1 \quad \mathbf{endif} \\
 ix &\leftarrow iy \times 10^k
 \end{aligned} \tag{1}$$

Na Tabela 2 é exemplificado o arredondamento dos números entre 140 e 150 e entre 540 e 550 para valores de  $k$  variando de 1 a 3 utilizando a Equação 1. As três últimas colunas indicam o valor final de  $ix$ .

Valor Inicial $ix$	Valor Arredondado $ix$		
	$k = 1$	$k = 2$	$k = 3$
140	140	100	0
141	140	100	0
142	140	100	0
143	140	100	0
144	140	100	0
145	150	100	0
146	150	100	0
147	150	100	0
148	150	100	0
149	150	100	0
150	150	200	0
540	540	500	1000
541	540	500	1000
542	540	500	1000
543	540	500	1000
544	540	500	1000
545	550	500	1000
546	550	500	1000
547	550	500	1000
548	550	500	1000
549	550	500	1000
550	550	600	1000

Tabela 2: Exemplo utilizando a Equação 1

A Equação 1 pode ser generalizada para qualquer base  $b$  além da base decimal, representada por meio da Equação 2.

$$\begin{aligned}
 iy &\leftarrow int(ix/b^k) \\
 \mathbf{if}(mod(ix, b^k) \geq b^k/2) &\mathbf{then} \ iy \leftarrow iy + 1 \quad \mathbf{endif} \\
 ix &\leftarrow iy \times b^k
 \end{aligned} \tag{2}$$

Na Tabela 3 é exemplificado o arredondamento dos números entre 140 e 150 e entre 540 e 550 para valores de  $k$  variando de 1 a 3 utilizando a Equação 2, considerando a base binária.

Em termos computacionais há interesse em utilizar base binária, ou seja,  $b = 2$  por motivos de eficiência. Na base binária as divisões por 2 (ou potências de 2) podem ser efetuadas por

Valor Inicial $ix$	Valor Arredondado $ix$		
	$k = 1$	$k = 2$	$k = 3$
140	140	140	144
141	142	140	144
142	142	144	144
143	144	144	144
144	144	144	144
145	146	144	144
146	146	148	144
147	148	148	144
148	148	148	152
149	150	148	152
150	150	152	152
540	540	540	544
541	542	540	544
542	542	544	544
543	544	544	544
544	544	544	544
545	546	544	544
546	546	548	544
547	548	548	544
548	548	548	552
549	550	548	552
550	550	552	552

Tabela 3: Exemplo utilizando a Equação 2 na base 2

meio de deslocamento (*shift*) de *bits* à direita e multiplicações por meio de deslocamento de *bits* à esquerda.

Por exemplo, o exemplo seguinte mostra o processo de *shift* para direita e *shift* para a esquerda para o número  $140_{10} = 010001100_2$ . Utilizando *shift* a para direita no número  $140_{10}$  obtém-se  $70_{10} = 001000110_2$ , o que equivale à divisão de 140 por 2; *shift* para a esquerda no número  $140_{10}$  obtém-se  $280_{10} = 100011000_2$ , o que equivale à multiplicação de 140 por 2.

	256 $2^8$	128 $2^7$	64 $2^6$	32 $2^5$	16 $2^4$	8 $2^3$	4 $2^2$	2 $2^1$	1 $2^0$
140	0	1	0	0	0	1	1	0	0
140 com shift para direita = 70	0	0	1	0	0	0	1	1	0
140 com shift para esquerda = 280	1	0	0	0	1	1	0	0	0

O tempo de arredondamento de um grande conjunto de dados é relativamente pequeno, segundo o Algoritmo 1 proposto por Weiss and Indurkha (1998) que descreve o procedimento geral para arredondamento de valores de um atributo, no qual a Equação 2 corresponde às linhas 13–17. Admitindo um número máximo de valores *max* para cada atributo, os valores do atributo são ordenados, para que o número de valores distintos possam ser contados. A ordem é guardada e não são necessárias ordenações adicionais. Começando com  $k = 1$ , o valor de  $k$  é incrementado até o número de valores ser reduzido a um valor menor ou igual ao máximo desejado, *max*. Para que o Algoritmo 1 possa ser aplicado a um conjunto de exemplos, o processo deve ser repetido para cada atributo, como pode ser visto no Algoritmo 2.

Os Algoritmos 1 e 2 foram implementados na linguagem de programação Java (Deitel and Deitel 2005) para a realização de experimentos descritos na Seção 6. Note, entretanto, que as linhas 3 e 5 do Algoritmo 1 são desnecessárias, caso o mesmo seja executado pelo Algoritmo 2.

---

**Algoritmo 1** Algoritmo de arredondamento proposto por Weiss

---

**Require:**  $\{v_i\}$ , conjunto dos valores de um atributo

$max$ , o máximo de valores distintos desejados

$b$ , base a ser utilizada

**Ensure:**  $\{v_i\}$  contendo no máximo  $max$  valores distintos

```
1:  $s \leftarrow 1$ 
2: Se o conjunto  $\{v_i\}$  contém frações, multiplica-se todos os valores por uma constante para
   que se obtenha apenas valores inteiros
3: Ordene os valores  $\{v_i\}$ 
4: loop
5:    $num \leftarrow$  número de valores distintos de  $\{v_i\}$ 
6:   if  $num \leq max$  then
7:     exit loop
8:   end if
9:    $s \leftarrow s + 1$ 
10:  for all valores  $ix \in \{v_i\}$  do
11:    Se  $ix$  negativo, multiplicar por  $-1$ 
12:     $k \leftarrow s$ 
13:     $iy \leftarrow \text{int}(ix/b^k)$ 
14:    if  $(\text{mod}(ix, b^k) \geq b^k/2)$  then
15:       $iy \leftarrow iy + 1$ 
16:    end if
17:     $ix \leftarrow iy \times b^k$ 
18:    Voltar o número  $ix$  para negativo se necessário
19:  end for
20: end loop
21: Dividir todos os valores pela mesma constante utilizada no início para voltar as frações
22: return conjunto arredondado  $\{v_i\}$ 
```

---

---

**Algoritmo 2** Algoritmo final de arredondamento

---

**Require:**  $dataset$ , conjunto de exemplos

$p$ , porcentagem máxima de valores distintos

$b$ , base a ser utilizada

```
1: for all atributo  $v_i \in dataset$  do
2:   Ordene os valores  $\{v_i\}$ 
3:    $num \leftarrow$  número de valores distintos de  $\{v_i\}$ 
4:    $max \leftarrow num \times p$ 
5:   Execute Algoritmo 1 com parâmetros  $\{v_i\}, max, b$ 
6: end for
7: return conjunto de exemplos arredondado
```

---

## 5 Metodologia Experimental

Neste experimento, para os indutores J48 e JRip, foram avaliados tempo de indução e taxa de erro tamanho do classificador usando *10-fold stratified cross-validation* tanto no conjunto original de exemplos (sem arredondamento) como nos conjuntos derivados, obtendo-se média e desvio padrão para o tempo de indução, taxa de erro, e tamanho do classificador para os 10 conjuntos de exemplos. Para os indutores IB, MLP e NB, a metodologia desenvolvida foi a mesma, porém a métrica tamanho do classificador não foi avaliada, pois:

- para o indutor IB o tamanho do classificar é igual ao número de exemplos do conjunto, portanto é fixo para o mesmo conjunto de exemplos;
- para o indutor MLP o tamanho do classificar é igual ao número de neurônios que constitui a rede, por isso é fixo para o mesmo conjunto de exemplos;
- para o indutor NB o tamanho do classificar é fixo para o mesmo conjunto de exemplos.

Esse experimento foi conduzido da seguinte forma: assuma 10 *folds* mutuamente exclusivos. Dos 10 *folds*, foram selecionados 9 *folds* e aplicado arredondamento dos valores somente nestes 9 *folds*; a partir do *fold* remanescente (sem arredondamento) foram avaliados tempo de indução, taxa de erro do classificador e tamanho do classificador para J48 e JRip; e foram avaliados tempo de indução e taxa de erro do classificador para IB, MLP e NB. Esse processo foi repetido um total de 10 vezes, cada vez utilizando um *fold* diferente de teste (sem arredondamento) para todos os conjuntos de exemplos.

Nas seções seguintes é freqüentemente mencionado o Algoritmo 1 por se tratar do algoritmo originalmente proposto por Weiss and Indurkha (1998), embora, em termos computacionais, o Algoritmo 2 tenha sido, de fato, utilizado.

Como já mencionado na Seção 4, o Algoritmo 1 possui o parâmetro ( $p$ ) que indica a porcentagem máxima permitida de valores distintos que são obtidos após aplicação do arredondamento no conjunto original, para cada atributo. Por exemplo, para um conjunto com 2 atributos, sendo o primeiro atributo contendo 100 valores distintos e o segundo atributo contendo 200 valores distintos, após a execução do Algoritmo 1 o conjunto derivado para  $p = 50\%$  terá, no máximo, 50 valores distintos para o primeiro atributo e 100 valores distintos para o segundo atributo.

Nesse experimento foram utilizados os valores de  $p$  iguais a 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20% e 10%, obtendo um conjunto derivado para cada valor de  $p$ . Por exemplo, no caso do sonar esses conjuntos derivados são indicados como sonar-90%, sonar-80%, sonar-70%, sonar-60%, sonar-50%, sonar-40%, sonar-30%, sonar-20% e sonar-10%, respectivamente. De forma análoga essa notação é utilizada para os demais conjuntos de exemplos.

Adicionalmente ao parâmetro  $p$ , os Algoritmo 1 e 2 também possuem o parâmetro  $b$ , que corresponde à base do sistema de numeração. Nos experimentos relatados nesta Seção, foram utilizados os valores de  $b$  iguais a 10 (base decimal) e 2 (base binária).

Nos gráficos apresentados a seguir é mostrada a diferença absoluta em desvios padrões do tempo de indução no eixo vertical do gráfico entre o conjunto original e os conjuntos derivados, ou seja, entre aml-all-completo e aml-all-completo-90%, entre aml-all-completo e aml-all-completo-80% e assim por diante. Quando a barra encontra-se acima de zero significa que o respectivo classificador do conjunto derivado supera o desempenho do classificador do conjunto original; se a barra encontra-se abaixo de zero então o classificador do conjunto original supera o respectivo classificador do conjunto derivado. Quando a altura da barra estiver acima (abaixo) de dois (menos dois) significa que o classificador do conjunto derivado (conjunto original) supera o classificador do conjunto original (conjunto derivado) significativamente, ou seja, nível de confiança de 95% (Rezende 2003; Moses 1986). Analogamente para taxa de erro e tamanho do classificador (quando aplicável).

## 6 Resultados

### 6.1 Resultados do indutor IB

#### 6.1.1 Resultados aml-all-completo

Na Tabela 4 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos aml-all-completo original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
aml-all-completo	1,000 $\pm$ 0,009	1,000 $\pm$ 0,009	9,643 $\pm$ 3,534	9,643 $\pm$ 3,534
aml-all-completo-90%	1,293 $\pm$ 0,015	1,004 $\pm$ 0,012	9,643 $\pm$ 3,534	9,643 $\pm$ 3,534
aml-all-completo-80%	1,291 $\pm$ 0,008	1,004 $\pm$ 0,006	9,643 $\pm$ 3,534	9,643 $\pm$ 3,534
aml-all-completo-70%	1,290 $\pm$ 0,006	0,998 $\pm$ 0,007	9,643 $\pm$ 3,534	9,643 $\pm$ 3,534
aml-all-completo-60%	1,291 $\pm$ 0,013	0,999 $\pm$ 0,010	9,643 $\pm$ 3,534	11,071 $\pm$ 3,386
aml-all-completo-50%	1,289 $\pm$ 0,012	0,994 $\pm$ 0,009	9,643 $\pm$ 3,534	11,071 $\pm$ 3,386
aml-all-completo-40%	1,292 $\pm$ 0,008	0,988 $\pm$ 0,009	9,643 $\pm$ 3,002	11,071 $\pm$ 3,386
aml-all-completo-30%	1,302 $\pm$ 0,017	0,983 $\pm$ 0,008	13,750 $\pm$ 2,843	11,071 $\pm$ 3,386
aml-all-completo-20%	1,306 $\pm$ 0,011	0,975 $\pm$ 0,009	16,607 $\pm$ 4,072	10,893 $\pm$ 3,372
aml-all-completo-10%	1,330 $\pm$ 0,017	0,912 $\pm$ 0,008	29,464 $\pm$ 4,590	18,036 $\pm$ 4,230

Tabela 4: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 aml-all-completo

Na Figura 2 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre aml-all-completo e aml-all-completo-90%, entre aml-all-completo e aml-all-completo-80% e assim por diante, utilizando base 2 e base 10.

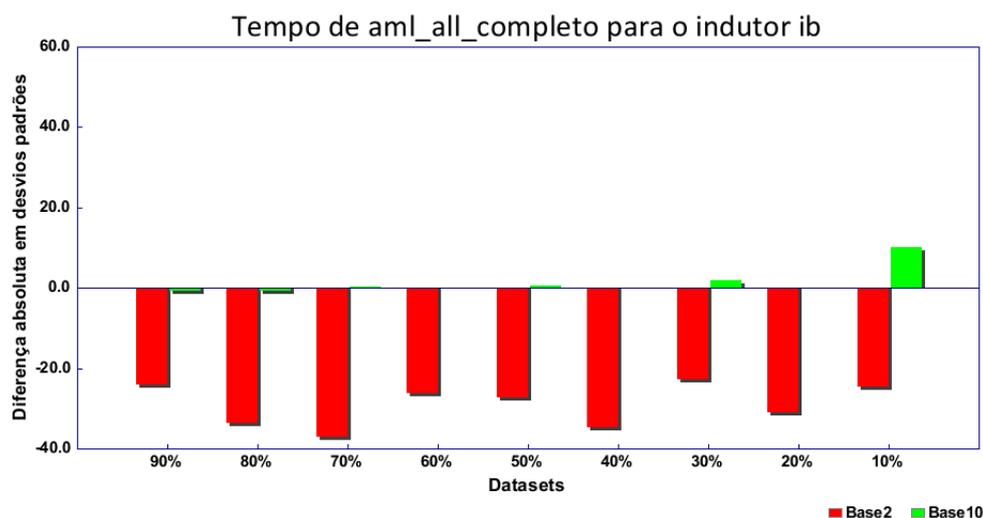


Figura 2: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) aml-all-completo

O tempo de indução aumentou de forma significativa (com grau de confiança de 95%) para todos os conjuntos utilizando arredondamento com base 2. E para a base 10, no geral o tempo

de indução diminui de forma não significativa, exceto para o conjunto `aml-all-completo-10%` que teve uma redução significativa.

Analogamente, na Figura 3 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 9,643% (`aml-all-completo`), para 13,075% utilizando o arredondamento com a base 2, o que representa um aumento de 35,595% da taxa de erro. E aumentou para 11,349% utilizando o arredondamento com a base 10, representando assim um aumento de 17,693%.

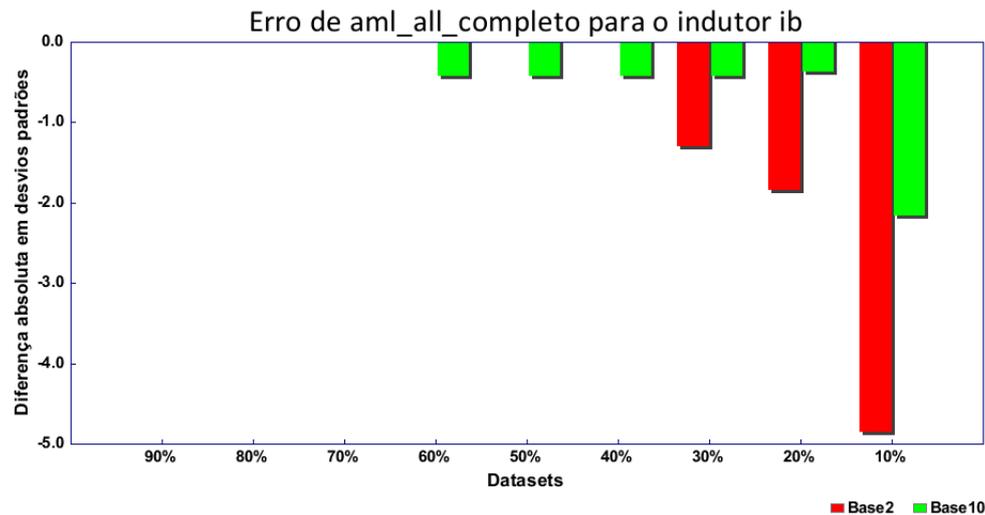


Figura 3: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) `aml-all-completo`

Como pode ser visto, a taxa de erro permaneceu constante ou aumentou para todos os conjuntos de forma não significativa, exceto para `aml-all-completo-10%` que a taxa de erro aumentou significativamente utilizando arredondamento tanto para a base 2 quanto para a base 10.

### 6.1.2 Resultados eucalyptus

Na Tabela 5 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos `eucalyptus` original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Na Figura 4 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre `eucalyptus` e `eucalyptus-90%`, entre `eucalyptus` e `eucalyptus-80%` e assim por diante, utilizando base 2 e base 10.

No geral, o tempo de indução aumentou utilizando arredondamento com as bases 2 e 10, embora de forma não significativa.

Analogamente, na Figura 5 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 48,947% (`eucalyptus`), para 50,230% utilizando o arredondamento com a base 2, o que representa um aumento de 3,574% da taxa de erro. E aumentou para 49,927% utilizando o arredondamento com a base 10, representando assim um aumento de 2,949%.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
eucalyptus	$1,000 \pm 0,047$	$1,000 \pm 0,047$	$48,497 \pm 1,287$	$48,497 \pm 1,287$
eucalyptus-90%	$1,066 \pm 0,047$	$1,017 \pm 0,078$	$48,495 \pm 1,387$	$50,122 \pm 1,393$
eucalyptus-80%	$1,041 \pm 0,043$	$1,248 \pm 0,398$	$48,221 \pm 1,454$	$50,122 \pm 1,393$
eucalyptus-70%	$1,000 \pm 0,047$	$0,983 \pm 0,047$	$49,041 \pm 1,278$	$50,122 \pm 1,393$
eucalyptus-60%	$1,248 \pm 0,460$	$0,975 \pm 0,035$	$49,173 \pm 1,330$	$50,122 \pm 1,393$
eucalyptus-50%	$1,050 \pm 0,078$	$1,041 \pm 0,105$	$49,169 \pm 1,684$	$50,122 \pm 1,393$
eucalyptus-40%	$1,050 \pm 0,088$	$1,000 \pm 0,026$	$49,715 \pm 1,624$	$49,985 \pm 1,379$
eucalyptus-30%	$1,033 \pm 0,080$	$1,000 \pm 0,047$	$51,074 \pm 1,556$	$49,717 \pm 1,496$
eucalyptus-20%	$1,058 \pm 0,065$	$1,000 \pm 0,091$	$53,391 \pm 0,976$	$48,360 \pm 1,637$
eucalyptus-10%	$0,983 \pm 0,047$	$0,967 \pm 0,103$	$53,793 \pm 1,704$	$50,674 \pm 1,877$

Tabela 5: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 eucalyptus

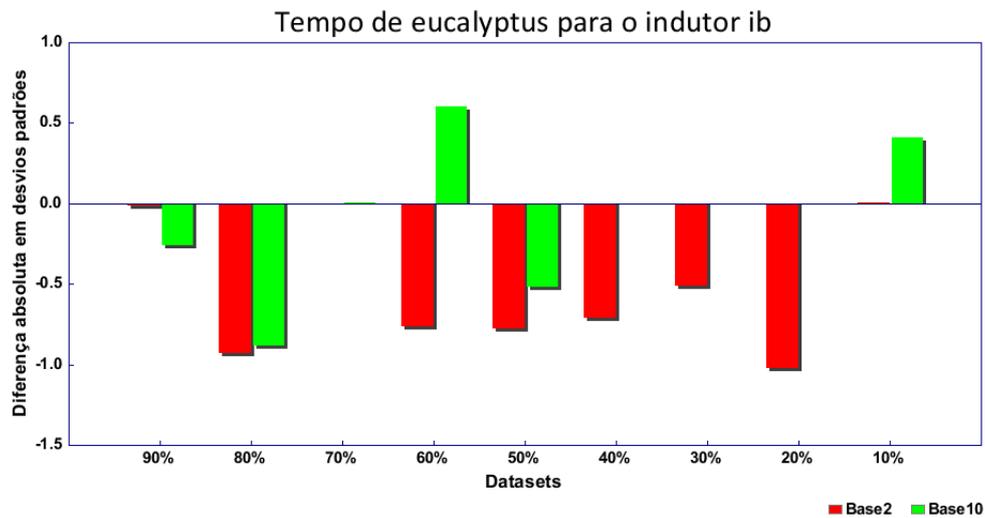


Figura 4: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) eucalyptus

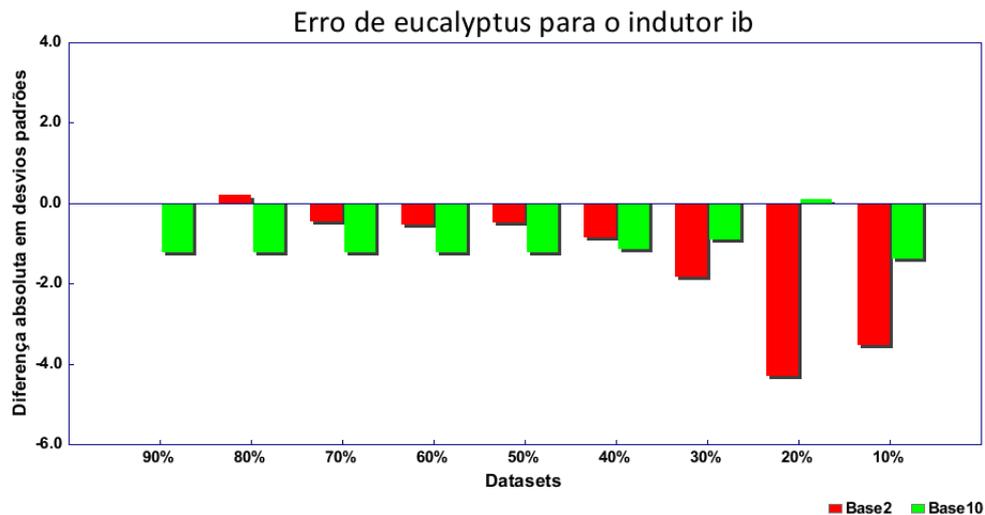


Figura 5: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) eucalyptus

Como pode ser notado, no geral, a taxa de erro aumentou para todos os conjuntos de forma não significativa, exceto para eucalyptus-20% e eucalyptus-10% que a taxa de erro aumentou significativamente utilizando arredondamento com a base 2.

### 6.1.3 Resultados hepatitis

Na Tabela 6 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos hepatitis original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
hepatitis	1,000 $\pm$ 0,070	1,000 $\pm$ 0,070	19,417 $\pm$ 1,768	19,417 $\pm$ 1,768
hepatitis-90%	1,012 $\pm$ 0,052	1,049 $\pm$ 0,177	19,458 $\pm$ 2,285	17,458 $\pm$ 2,010
hepatitis-80%	1,025 $\pm$ 0,060	0,951 $\pm$ 0,102	18,792 $\pm$ 1,900	17,458 $\pm$ 2,010
hepatitis-70%	1,000 $\pm$ 0,039	0,951 $\pm$ 0,060	18,125 $\pm$ 1,699	17,458 $\pm$ 2,010
hepatitis-60%	1,025 $\pm$ 0,060	0,951 $\pm$ 0,083	18,125 $\pm$ 2,205	17,458 $\pm$ 2,010
hepatitis-50%	1,148 $\pm$ 0,193	1,160 $\pm$ 0,452	18,792 $\pm$ 1,900	17,458 $\pm$ 2,010
hepatitis-40%	1,025 $\pm$ 0,083	1,037 $\pm$ 0,255	18,125 $\pm$ 1,699	17,458 $\pm$ 2,010
hepatitis-30%	1,148 $\pm$ 0,193	1,062 $\pm$ 0,325	18,125 $\pm$ 1,699	18,125 $\pm$ 1,699
hepatitis-20%	1,062 $\pm$ 0,086	1,074 $\pm$ 0,210	17,500 $\pm$ 1,786	19,375 $\pm$ 1,378
hepatitis-10%	1,000 $\pm$ 0,039	0,988 $\pm$ 0,101	18,125 $\pm$ 2,394	19,375 $\pm$ 2,419

Tabela 6: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 hepatitis

Na Figura 6 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre hepatitis e hepatitis-90%, entre hepatitis e hepatitis-80% e assim por diante, utilizando base 2 e base 10.

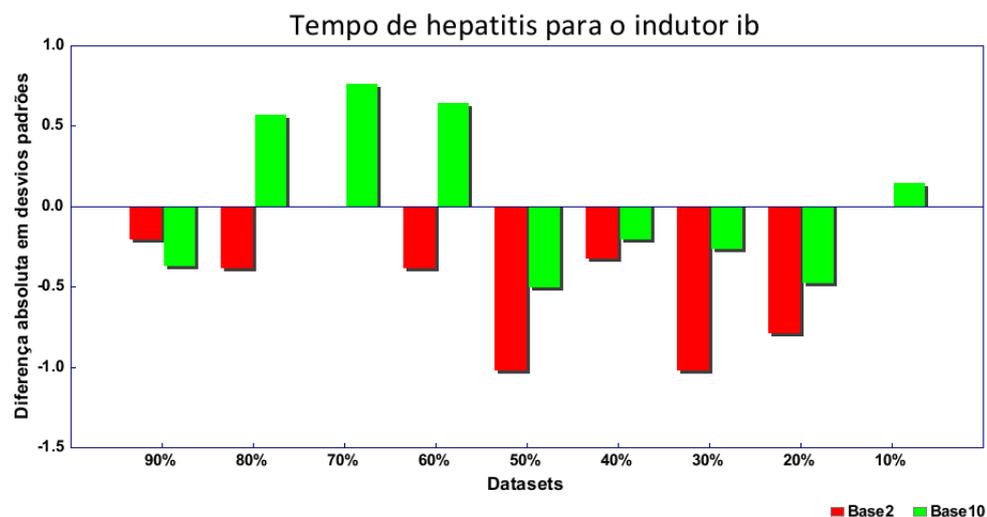


Figura 6: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

Na maioria dos conjuntos o tempo de indução aumentou, sendo de forma não significativa, tanto para base 2 como para a base 10.

Analogamente, na Figura 7 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro diminuiu de 19,417% (hepatitis), para 18,352% utilizando o arredondamento com a base 2, o que representa uma redução de 5,485% da taxa de erro. E reduziu para 17,958% utilizando o arredondamento com a base 10, representando assim uma redução de 7,513%.

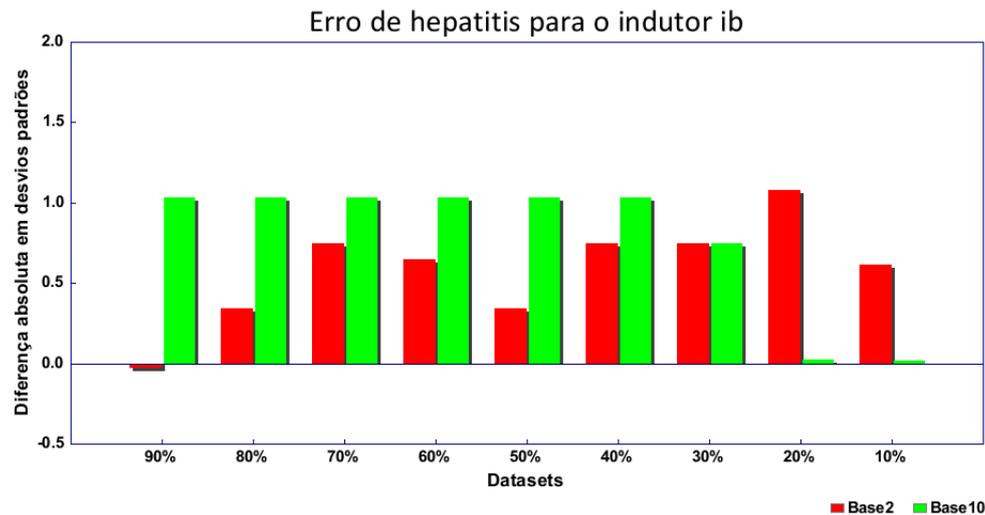


Figura 7: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

Como pode ser observado, no geral, a taxa de erro reduziu para todos os conjuntos de forma não significativa, utilizando arredondamento com a bases 2 e 10.

#### 6.1.4 Resultados ionosphere

Na Tabela 7 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos ionosphere original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Na Figura 8 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre ionosphere e ionosphere-90%, entre ionosphere e ionosphere-80% e assim por diante, utilizando base 2 e base 10.

No geral, o tempo de indução aumentou para todos os conjuntos utilizando arredondamento, sendo de forma não significativa, tanto para base 2 como para a base 10, exceto para o conjunto ionosphere-40%, que o seu tempo de indução reduziu de forma significativa para o arredondamento utilizando base 10.

Analogamente, na Figura 9 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 12,841% (ionosphere), para 13,882% utilizando o arredondamento com a base 2, o que representa um aumento de 8,105% da taxa de erro. E reduziu para 12,705% utilizando o arredondamento com a base 10, representando assim uma redução de 1,056%.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
ionosphere	1,000 ± 0,000	1,000 ± 0,000	12,841 ± 2,146	12,841 ± 2,146
ionosphere-90%	0,960 ± 0,052	0,950 ± 0,053	13,405 ± 2,051	13,119 ± 2,144
ionosphere-80%	1,250 ± 0,460	1,130 ± 0,554	13,405 ± 2,051	13,119 ± 2,144
ionosphere-70%	1,030 ± 0,106	1,090 ± 0,260	13,690 ± 2,252	13,119 ± 2,144
ionosphere-60%	1,300 ± 0,819	1,000 ± 0,082	13,976 ± 2,238	13,119 ± 2,144
ionosphere-50%	1,070 ± 0,082	1,050 ± 0,151	14,540 ± 2,154	13,119 ± 2,227
ionosphere-40%	1,000 ± 0,000	0,930 ± 0,048	13,984 ± 2,318	12,254 ± 2,175
ionosphere-30%	1,180 ± 0,382	0,980 ± 0,123	13,984 ± 2,318	12,262 ± 2,300
ionosphere-20%	1,040 ± 0,070	1,130 ± 0,350	13,690 ± 2,041	12,262 ± 2,300
ionosphere-10%	1,210 ± 0,404	1,060 ± 0,117	14,262 ± 1,713	11,976 ± 2,329

Tabela 7: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 ionosphere

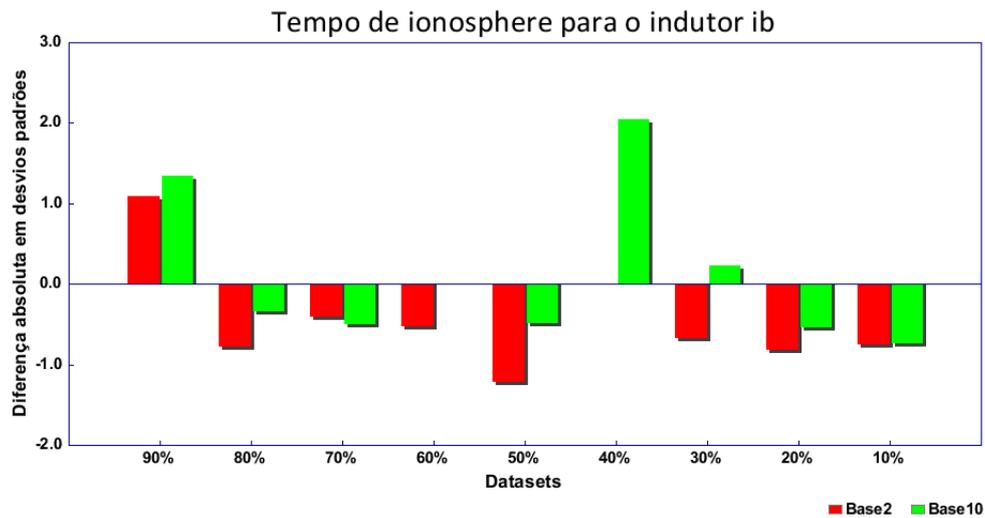


Figura 8: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) ionosphere

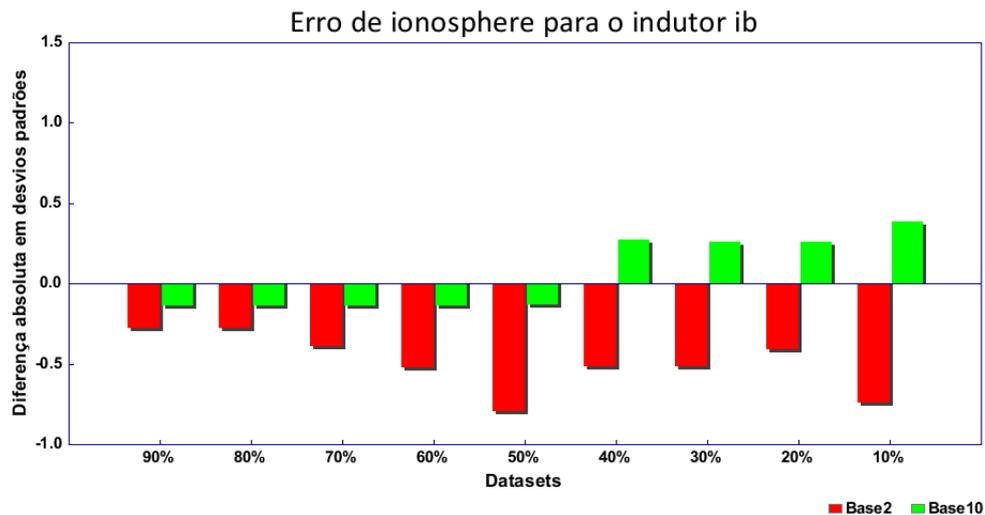


Figura 9: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) ionosphere

Como pode ser visto, no geral, a taxa de erro aumentou para todos os conjuntos de forma não significativa, utilizando arredondamento com a base 2. Já para a base 10, não se observou um padrão, pois alguns conjuntos tiveram a taxa de erro aumentada e outros reduzida.

### 6.1.5 Resultados lymph

Na Tabela 8 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos lymph original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
lymph	1,000 $\pm$ 0,000	1,000 $\pm$ 0,000	19,524 $\pm$ 2,268	19,524 $\pm$ 2,268
lymph-90%	1,217 $\pm$ 0,081	0,983 $\pm$ 0,053	23,762 $\pm$ 2,813	23,714 $\pm$ 2,132
lymph-80%	1,267 $\pm$ 0,117	0,967 $\pm$ 0,070	23,762 $\pm$ 2,813	23,714 $\pm$ 2,132
lymph-70%	1,167 $\pm$ 0,000	0,967 $\pm$ 0,070	23,762 $\pm$ 2,813	23,714 $\pm$ 2,132
lymph-60%	1,200 $\pm$ 0,070	0,950 $\pm$ 0,081	23,714 $\pm$ 2,352	23,714 $\pm$ 2,132
lymph-50%	1,217 $\pm$ 0,081	0,917 $\pm$ 0,088	23,714 $\pm$ 2,352	23,714 $\pm$ 2,132
lymph-40%	1,200 $\pm$ 0,070	0,983 $\pm$ 0,053	23,714 $\pm$ 2,132	23,714 $\pm$ 2,132
lymph-30%	1,100 $\pm$ 0,086	0,983 $\pm$ 0,053	23,714 $\pm$ 2,132	23,714 $\pm$ 2,132
lymph-20%	0,967 $\pm$ 0,070	0,950 $\pm$ 0,081	23,048 $\pm$ 2,353	23,048 $\pm$ 2,353
lymph-10%	0,917 $\pm$ 0,088	0,967 $\pm$ 0,070	23,048 $\pm$ 2,353	23,048 $\pm$ 2,353

Tabela 8: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 lymph

Na Figura 10 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre lymph e lymph-90%, entre lymph e lymph-80% e assim por diante, utilizando base 2 e base 10.

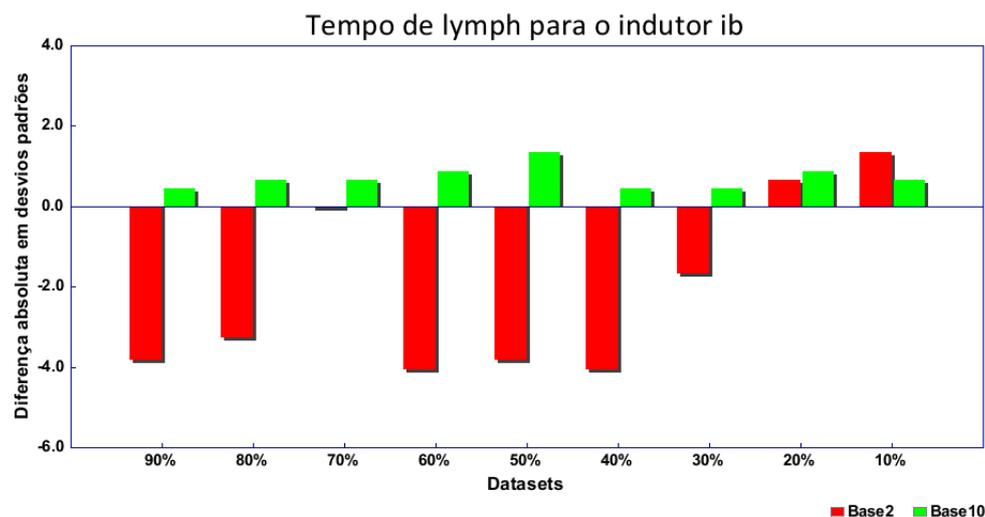


Figura 10: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

No geral, o tempo de indução aumentou para todos os conjuntos utilizando arredondamento com base 2, sendo de forma significativa em 5 dos 9 conjuntos gerados. E para a base 10 o tempo de indução reduziu para todos os conjuntos, embora de forma não significativa.

Analogamente, na Figura 11 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 19,524% (lymph), para 23,582% utilizando o arredondamento com a base 2, o que representa um aumento de 20,785% da taxa de erro. E aumentou para 23,566% utilizando o arredondamento com a base 10, representando assim um aumento de 20,703%.

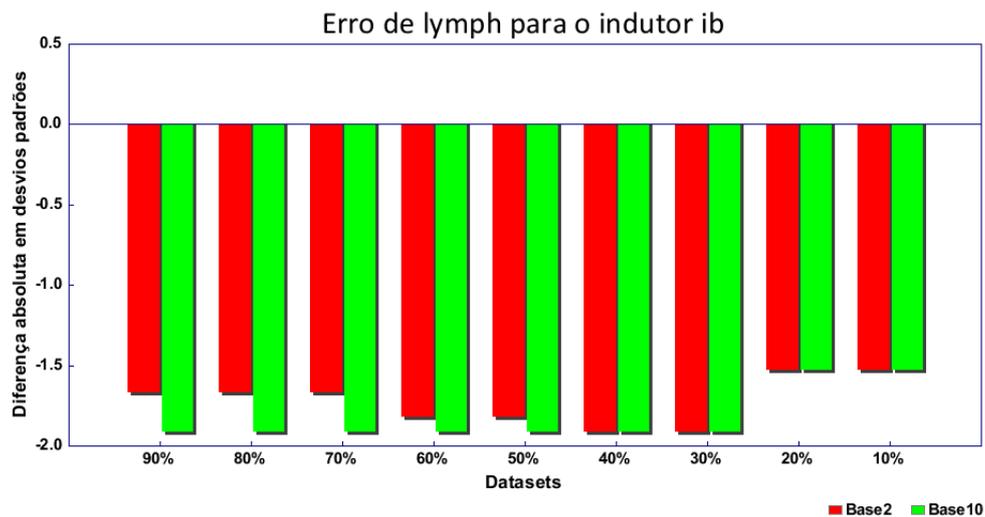


Figura 11: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

Como pode ser notado, a taxa de erro aumentou para todos os conjuntos de forma não significativa, utilizando arredondamento com as bases 2 e 10.

### 6.1.6 Resultados sick

Na Tabela 9 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sick original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Na Figura 12 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre sick e sick-90%, entre sick e sick-80% e assim por diante, utilizando base 2 e base 10.

No geral, o tempo de indução aumentou para todos os conjuntos utilizando arredondamento com base 2, sendo de forma significativa. Já para a base 10, não se observou um padrão, pois

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
sick	$1,000 \pm 0,020$	$1,000 \pm 0,020$	$3,844 \pm 0,397$	$3,844 \pm 0,397$
sick-90%	$1,041 \pm 0,020$	$1,016 \pm 0,020$	$4,003 \pm 0,444$	$6,548 \pm 0,490$
sick-80%	$1,037 \pm 0,017$	$1,016 \pm 0,020$	$4,348 \pm 0,495$	$6,548 \pm 0,490$
sick-70%	$1,033 \pm 0,013$	$1,021 \pm 0,017$	$4,374 \pm 0,495$	$6,548 \pm 0,490$
sick-60%	$1,025 \pm 0,013$	$0,988 \pm 0,000$	$4,427 \pm 0,492$	$6,574 \pm 0,502$
sick-50%	$1,033 \pm 0,013$	$0,996 \pm 0,017$	$4,348 \pm 0,486$	$6,574 \pm 0,502$
sick-40%	$1,033 \pm 0,013$	$0,996 \pm 0,017$	$4,798 \pm 0,517$	$6,574 \pm 0,502$
sick-30%	$1,037 \pm 0,017$	$0,992 \pm 0,013$	$4,931 \pm 0,502$	$6,574 \pm 0,493$
sick-20%	$1,033 \pm 0,013$	$0,988 \pm 0,000$	$7,979 \pm 0,550$	$6,574 \pm 0,493$
sick-10%	$1,041 \pm 0,020$	$1,012 \pm 0,021$	$9,650 \pm 0,450$	$8,060 \pm 0,221$

Tabela 9: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 sick

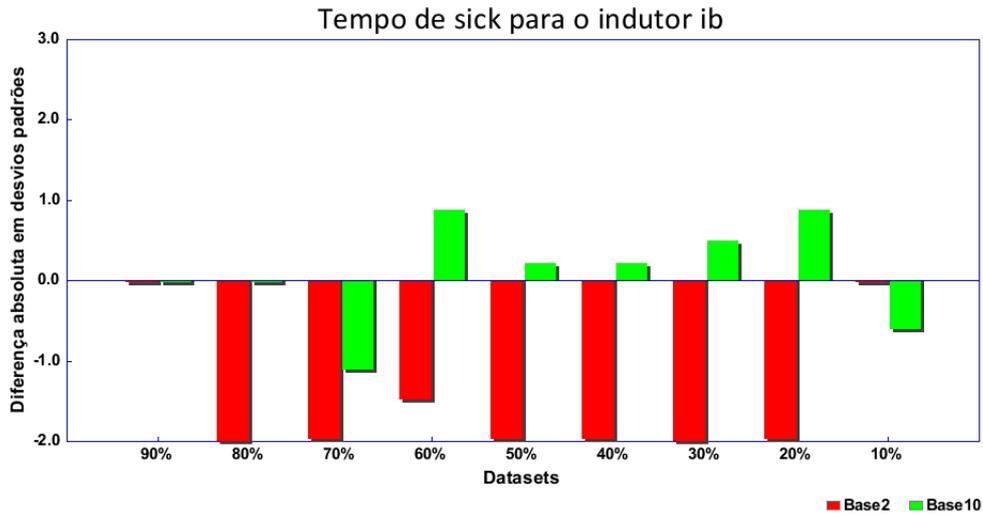


Figura 12: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

alguns conjuntos tiveram a tempo de indução aumentado e outros reduzido.

Analogamente, na Figura 13 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 13,524% (sick), para 13,902% utilizando o arredondamento com a base 2, o que representa um aumento de 2,79% da taxa de erro. E aumentou para 15,056% utilizando o arredondamento com a base 10, representando assim um aumento de 11,33%.

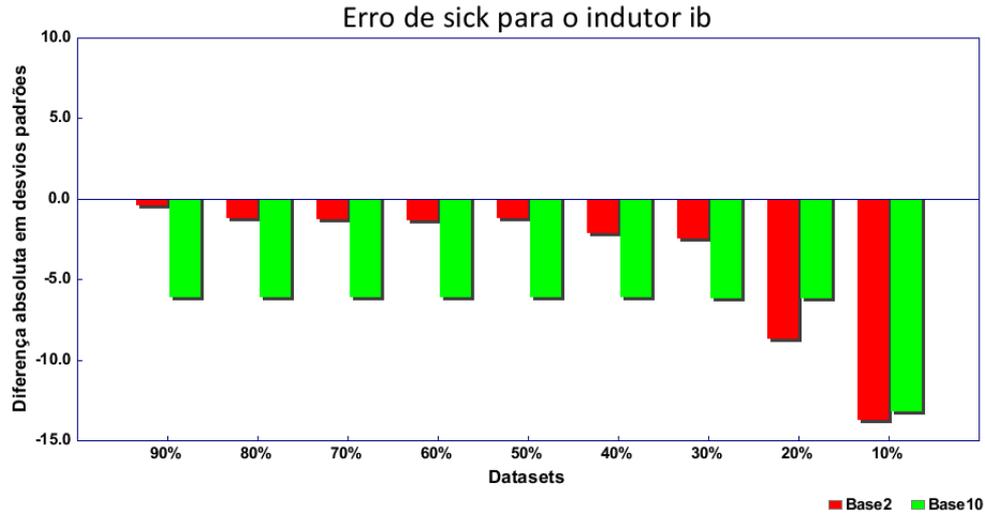


Figura 13: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

No geral, a taxa de erro aumentou para todos os conjuntos utilizando arredondamento com a base 2, sendo de forma significativa para 4 de seus conjuntos, podemos observar que a medida uma menor porcentagem de valores distintos é mantida a taxa de erro aumenta. Já para a base 10, o tempo de indução aumentou para todos os conjuntos de forma significativa.

### 6.1.7 Resultados sonar

Na Tabela 10 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sonar original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
sonar	1,000 $\pm$ 0,125	1,000 $\pm$ 0,125	13,524 $\pm$ 2,059	13,524 $\pm$ 2,059
sonar-90%	0,960 $\pm$ 0,052	0,940 $\pm$ 0,052	14,024 $\pm$ 2,157	15,000 $\pm$ 2,515
sonar-80%	0,950 $\pm$ 0,071	0,940 $\pm$ 0,052	14,024 $\pm$ 2,157	15,000 $\pm$ 2,515
sonar-70%	0,950 $\pm$ 0,071	0,970 $\pm$ 0,095	14,024 $\pm$ 2,157	15,000 $\pm$ 2,515
sonar-60%	0,930 $\pm$ 0,048	0,940 $\pm$ 0,052	13,524 $\pm$ 2,059	15,000 $\pm$ 2,515
sonar-50%	0,960 $\pm$ 0,052	0,930 $\pm$ 0,048	13,524 $\pm$ 2,059	15,000 $\pm$ 2,515
sonar-40%	0,920 $\pm$ 0,042	0,900 $\pm$ 0,000	13,524 $\pm$ 2,059	15,024 $\pm$ 2,845
sonar-30%	0,940 $\pm$ 0,052	0,940 $\pm$ 0,070	14,000 $\pm$ 2,011	15,976 $\pm$ 2,799
sonar-20%	0,930 $\pm$ 0,048	0,900 $\pm$ 0,000	13,524 $\pm$ 2,059	15,952 $\pm$ 2,690
sonar-10%	0,930 $\pm$ 0,048	0,910 $\pm$ 0,032	14,952 $\pm$ 2,123	13,548 $\pm$ 2,640

Tabela 10: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 sonar

Na Figura 14 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre sonar e sonar-90%, entre sonar e sonar-80% e assim por diante, utilizando base 2 e base 10.

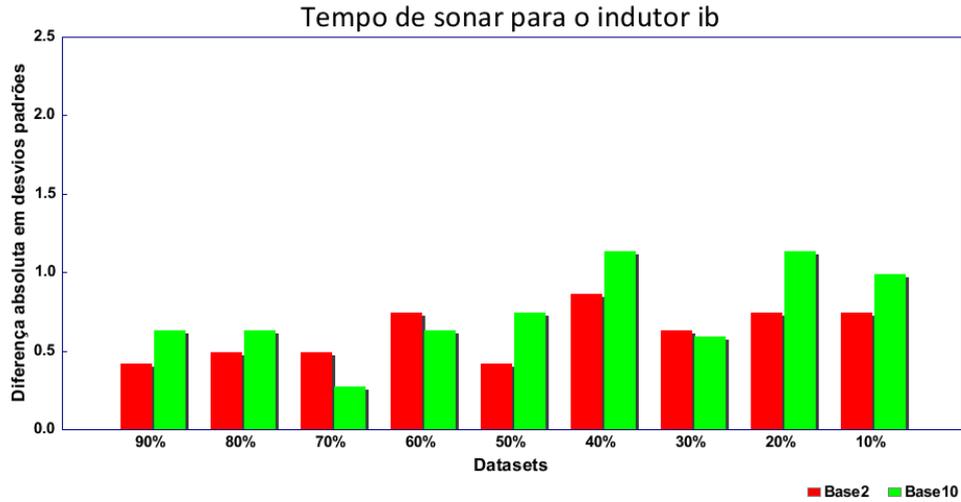


Figura 14: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

O tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo de forma não significativa tanto para base 2 como para a base 10.

Analogamente, na Figura 15 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 13,524% (sonar), para 13,902% utilizando o arredondamento com a base 2, o que representa um aumento de 2,797% da taxa de erro. E aumentou para 15,056% utilizando o arredondamento com a base 10, representando assim um aumento de 11,325%.

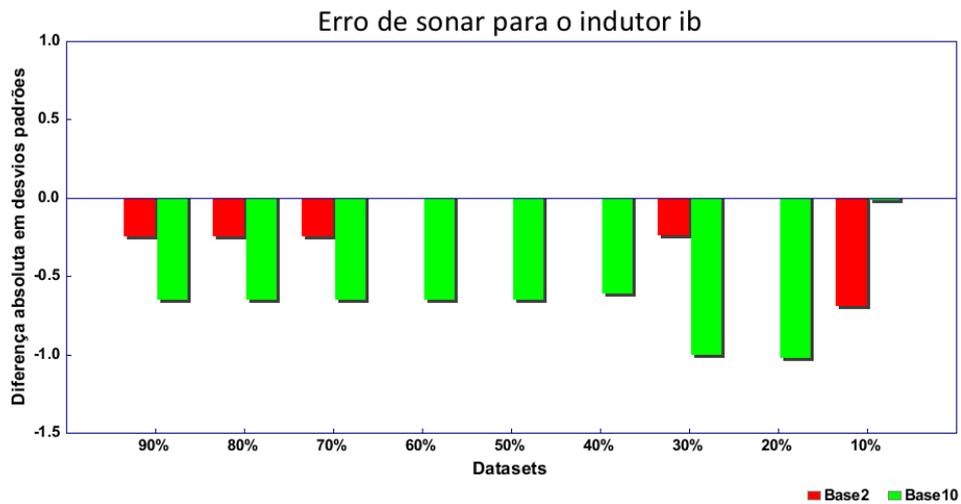


Figura 15: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

Como pode ser observado, no geral, a taxa de erro aumentou para todos os conjuntos de forma não significativa, exceto para sonar-20% e sonar-10% que a taxa de erro aumentou significativamente utilizando arredondamento com a base 2.

### 6.1.8 Resultados vehicle

Na Tabela 11 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *vehicle* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
vehicle	1,000 $\pm$ 0,041	1,000 $\pm$ 0,041	29,562 $\pm$ 1,440	29,562 $\pm$ 1,440
vehicle-90%	1,103 $\pm$ 0,158	0,991 $\pm$ 0,044	32,755 $\pm$ 1,783	34,633 $\pm$ 1,341
vehicle-80%	1,068 $\pm$ 0,045	0,983 $\pm$ 0,045	32,755 $\pm$ 1,783	34,633 $\pm$ 1,341
vehicle-70%	1,068 $\pm$ 0,045	0,983 $\pm$ 0,045	32,989 $\pm$ 1,655	34,633 $\pm$ 1,341
vehicle-60%	1,068 $\pm$ 0,045	0,991 $\pm$ 0,044	32,873 $\pm$ 1,736	34,633 $\pm$ 1,341
vehicle-50%	1,060 $\pm$ 0,044	0,966 $\pm$ 0,041	34,524 $\pm$ 1,920	34,633 $\pm$ 1,341
vehicle-40%	1,068 $\pm$ 0,045	1,009 $\pm$ 0,036	37,120 $\pm$ 1,848	34,633 $\pm$ 1,341
vehicle-30%	1,077 $\pm$ 0,044	0,983 $\pm$ 0,045	39,958 $\pm$ 1,759	34,870 $\pm$ 1,285
vehicle-20%	1,034 $\pm$ 0,027	0,983 $\pm$ 0,045	48,112 $\pm$ 1,837	36,640 $\pm$ 1,344
vehicle-10%	1,060 $\pm$ 0,044	0,966 $\pm$ 0,041	62,301 $\pm$ 2,429	56,501 $\pm$ 1,081

Tabela 11: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 *vehicle*

Na Figura 16 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *vehicle* e *vehicle-90%*, entre *vehicle* e *vehicle-80%* e assim por diante, utilizando base 2 e base 10.

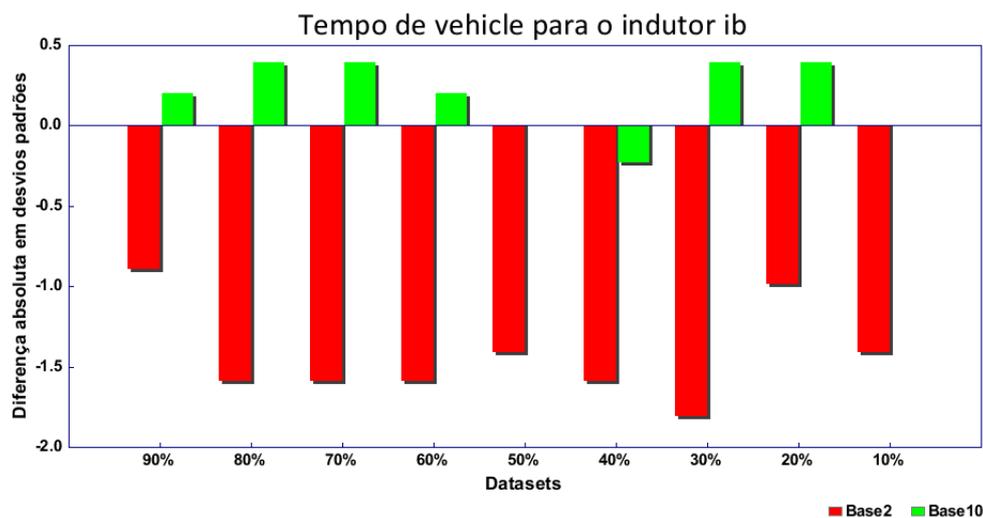


Figura 16: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *vehicle*

Podemos observar que o tempo de indução aumentou para todos os conjuntos utilizando arredondamento com a base 2, e no geral, o tempo de indução reduziu para os conjuntos utilizando arredondamento com a base 10, sendo de forma não significativa para ambas as bases.

Analogamente, na Figura 17 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 29,562% (*vehicle*), para 39,265% utilizando o arredondamento com a base 2, o que representa um aumento de

32,823% da taxa de erro. E aumentou para 37,312% utilizando o arredondamento com a base 10, representando assim um aumento de 26,216%.

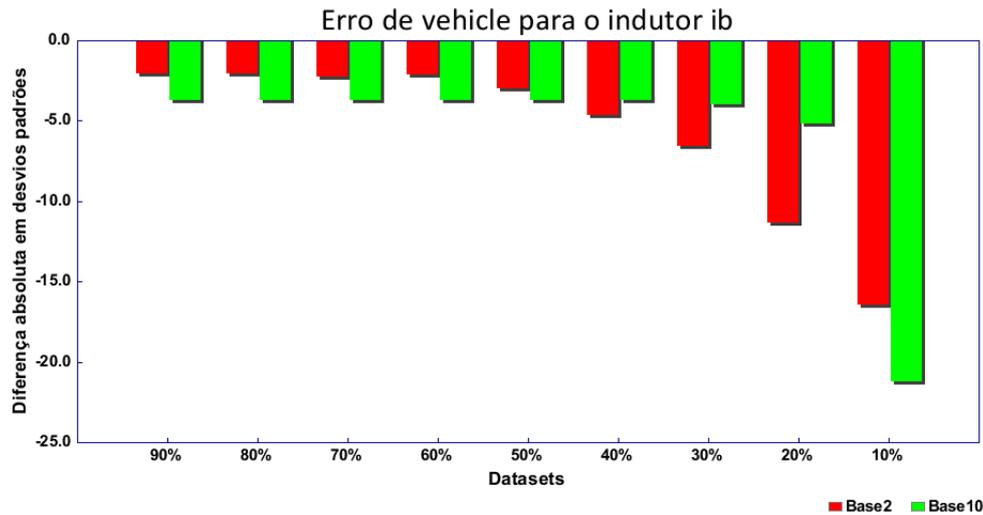


Figura 17: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 versus conjunto original) vehicle

No geral, a taxa de erro aumentou para todos os conjuntos de forma significativa para ambas as bases, podemos observar que a medida uma menor porcentagem de valores distintos é mantida a taxa de erro aumenta.

### 6.1.9 Resultados vowel

Na Tabela 12 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos vowel original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
vowel	1,000 $\pm$ 0,024	1,000 $\pm$ 0,024	0,505 $\pm$ 0,168	0,505 $\pm$ 0,168
vowel-90%	0,992 $\pm$ 0,000	1,023 $\pm$ 0,097	0,505 $\pm$ 0,168	0,505 $\pm$ 0,168
vowel-80%	1,000 $\pm$ 0,024	1,000 $\pm$ 0,024	0,505 $\pm$ 0,168	0,505 $\pm$ 0,168
vowel-70%	1,000 $\pm$ 0,024	0,992 $\pm$ 0,000	0,505 $\pm$ 0,168	0,505 $\pm$ 0,168
vowel-60%	1,008 $\pm$ 0,048	0,985 $\pm$ 0,024	0,505 $\pm$ 0,168	0,505 $\pm$ 0,168
vowel-50%	0,992 $\pm$ 0,000	0,985 $\pm$ 0,024	0,505 $\pm$ 0,168	0,505 $\pm$ 0,168
vowel-40%	0,985 $\pm$ 0,024	0,977 $\pm$ 0,032	0,505 $\pm$ 0,168	0,505 $\pm$ 0,168
vowel-30%	0,992 $\pm$ 0,000	0,977 $\pm$ 0,032	0,505 $\pm$ 0,168	0,808 $\pm$ 0,252
vowel-20%	1,031 $\pm$ 0,082	0,969 $\pm$ 0,037	0,404 $\pm$ 0,165	0,808 $\pm$ 0,252
vowel-10%	1,000 $\pm$ 0,024	0,992 $\pm$ 0,036	0,606 $\pm$ 0,269	0,808 $\pm$ 0,252

Tabela 12: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 vowel

Na Figura 18 é mostrada a diferença absoluta em desvios padrões do tempo de indução

entre o conjunto original e os conjuntos derivados, ou seja, entre vowel e vowel-90%, entre vowel e vowel-80% e assim por diante, utilizando base 2 e base 10.

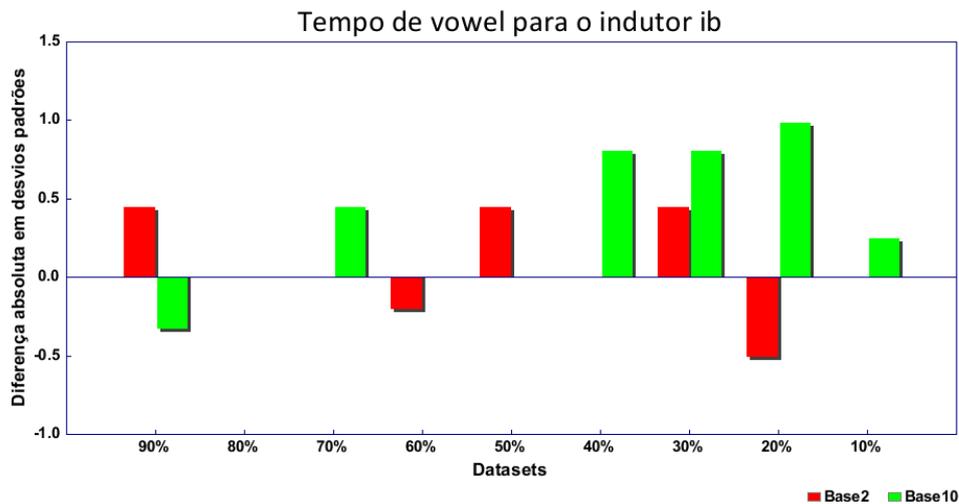


Figura 18: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

No geral, o tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo de forma não significativa (com grau de confiança de 95%), tanto para base 2 como para a base 10.

Analogamente, na Figura 19 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. A taxa de erro permaneceu a mesma entre (vowel), e utilizando o arredondamento com a base 2. E aumentou de 0,505% para 0,606% utilizando o arredondamento com a base 10, representando assim um aumento de 20,000%.

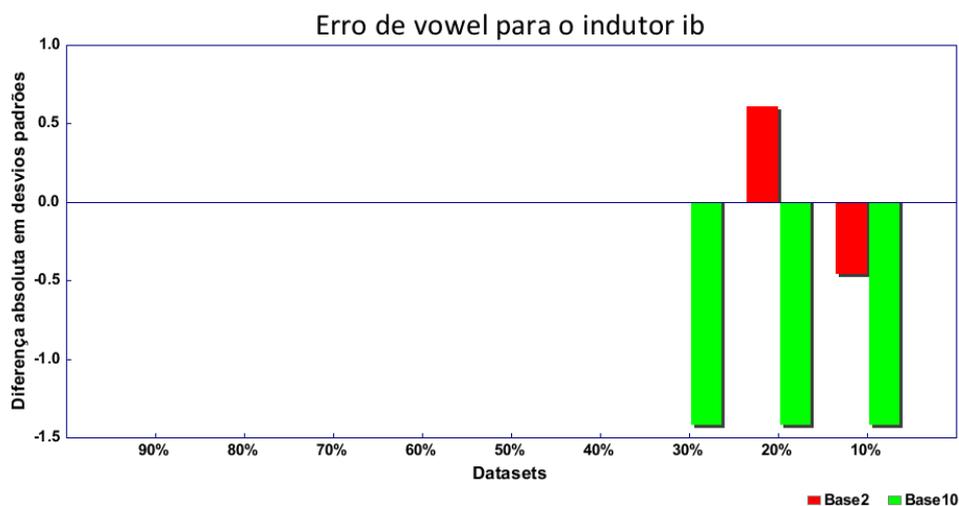


Figura 19: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

No geral, a taxa de erro permaneceu constante para todos os conjuntos, porém para os conjuntos com uma menor porcentagem de valores distintos houve uma variação na taxa de erro, aumentando para alguns e reduzindo para outros, embora de forma não significativa.

### 6.1.10 Resultados wine

Na Tabela 13 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos wine original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
wine	1,000 $\pm$ 0,062	1,000 $\pm$ 0,062	4,510 $\pm$ 1,620	4,510 $\pm$ 1,620
wine-90%	1,066 $\pm$ 0,138	1,077 $\pm$ 0,206	4,510 $\pm$ 1,620	3,954 $\pm$ 1,197
wine-80%	1,022 $\pm$ 0,074	1,022 $\pm$ 0,053	4,510 $\pm$ 1,620	3,954 $\pm$ 1,197
wine-70%	0,978 $\pm$ 0,035	1,022 $\pm$ 0,053	4,510 $\pm$ 1,620	3,954 $\pm$ 1,197
wine-60%	1,022 $\pm$ 0,104	1,000 $\pm$ 0,035	4,510 $\pm$ 1,620	3,954 $\pm$ 1,197
wine-50%	0,945 $\pm$ 0,057	0,945 $\pm$ 0,057	4,510 $\pm$ 1,620	3,954 $\pm$ 1,197
wine-40%	0,978 $\pm$ 0,035	0,989 $\pm$ 0,090	3,922 $\pm$ 1,671	3,954 $\pm$ 1,197
wine-30%	1,044 $\pm$ 0,093	1,011 $\pm$ 0,070	4,510 $\pm$ 1,119	2,810 $\pm$ 0,937
wine-20%	0,978 $\pm$ 0,062	1,011 $\pm$ 0,101	3,366 $\pm$ 1,235	5,621 $\pm$ 1,172
wine-10%	1,033 $\pm$ 0,077	1,143 $\pm$ 0,165	5,000 $\pm$ 1,538	9,543 $\pm$ 2,037

Tabela 13: Tempo de indução e taxa de erro do classificador para o indutor IB utilizando arredondamento com bases 2 e 10 wine

Na Figura 20 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre wine e wine-90%, entre wine e wine-80% e assim por diante, utilizando base 2 e base 10.

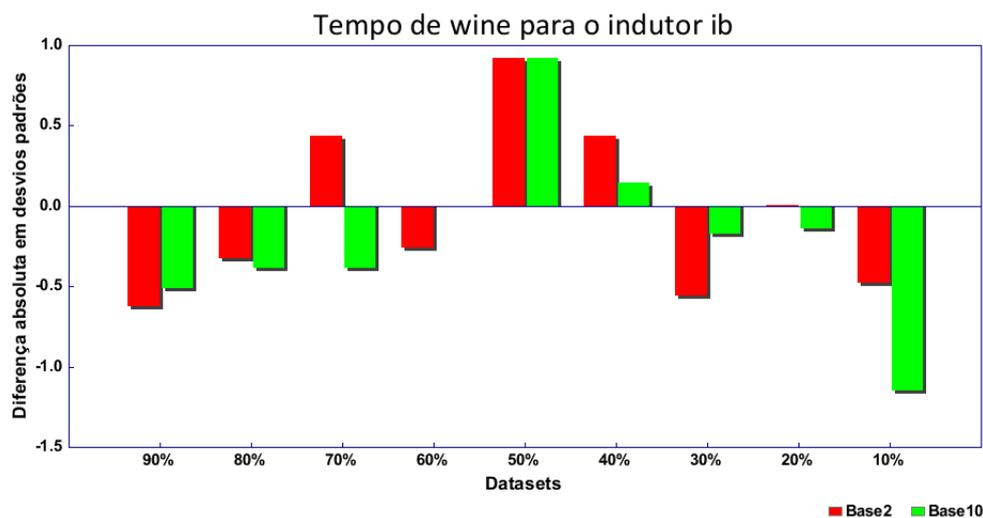


Figura 20: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

O tempo de indução reduziu para a maioria dos conjuntos utilizando arredondamento, sendo de forma não significativa, tanto para base 2 como para a base 10.

Analogamente, na Figura 21 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 4,510% (wine), para 4,372% utilizando o arredondamento com a base 2, o que representa uma redução de 3,060% da taxa de erro. E aumentou para 4,633% utilizando o arredondamento com a base 10, representando assim um aumento de 2,730%.

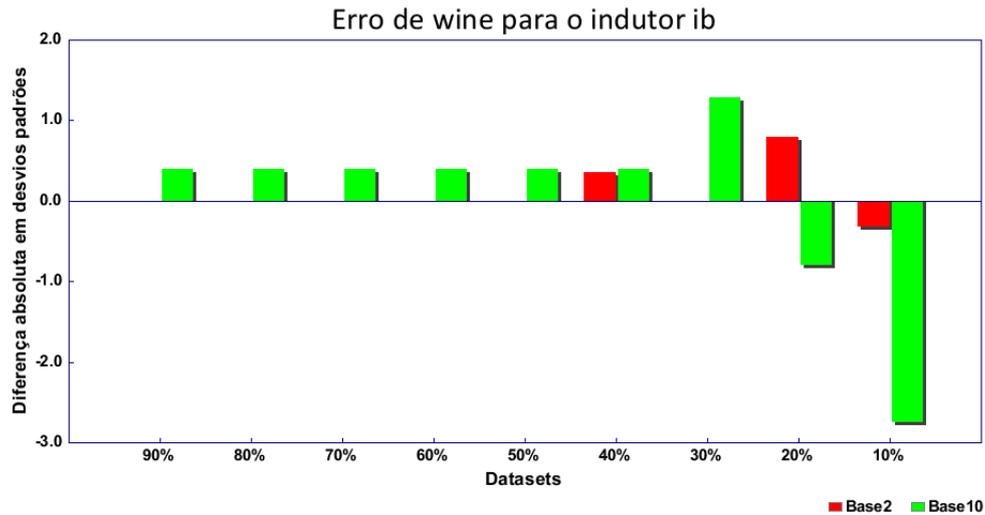


Figura 21: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

No geral, a taxa de erro reduziu para a maioria dos conjuntos de forma não significativa, exceto para wine-10% que a taxa de erro aumentou significativamente utilizando arredondamento com a base 10.

### 6.1.11 Discussão

Nas Figuras 22, 23, 24 e 25 é mostrado um resumo dos resultados obtidos por meio do indutor IB do tempo de indução utilizando base 2, tempo de indução utilizando base 10, taxa de erro utilizando base 2, taxa de erro utilizando base 10, respectivamente. No total são 90 conjuntos (9 conjuntos arredondados  $\times$  10 conjuntos de exemplos). No eixo  $y$  do gráfico é mostrado o número conjuntos que ficaram em cada uma das categorias mostradas no eixo  $x$  do gráfico, que são:

- aqueles que tiveram o tempo significativamente maior para os conjuntos derivados em relação ao conjunto original (barras maior que -2);
- os que tiveram o tempo maior para os conjuntos derivados (barras variando entre -2 e 0);
- os que tiveram o tempo igual para os conjuntos derivados e o conjunto original (barras igual 0);
- os que tiveram o tempo menor para os conjuntos derivados (barras variando entre 0 e 2);
- os que tiveram o tempo significativamente menor para os conjuntos derivados em relação ao conjunto original (barras maior que 2);

Analogamente para a taxa de erro.

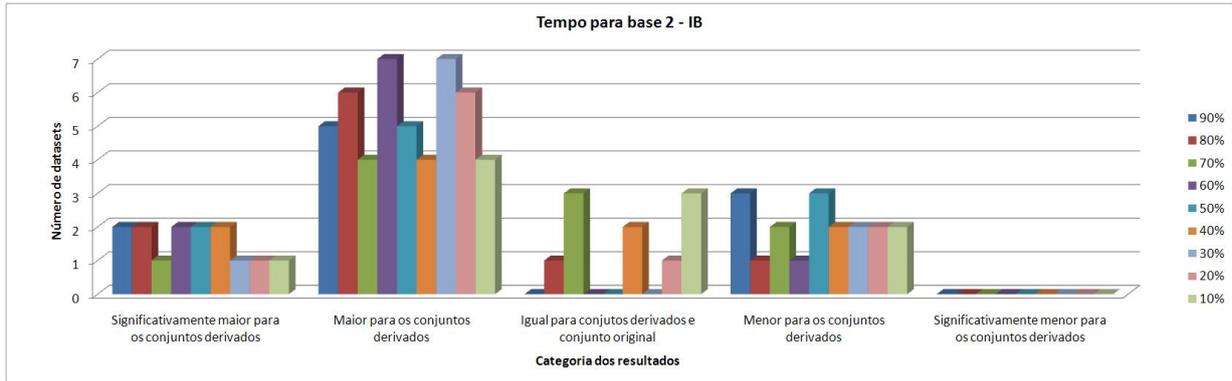


Figura 22: Resumo dos resultados do tempo de indução utilizando base 2 de IB

Como pode ser visto, houve uma tendência do tempo de indução aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa.

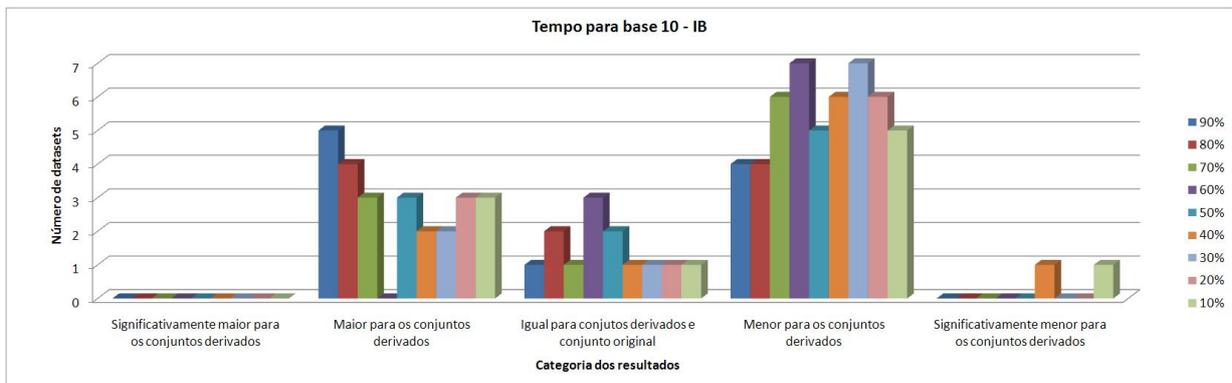


Figura 23: Resumo dos resultados do tempo de indução utilizando base 10 de IB

Como pode ser visto, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma não significativa.

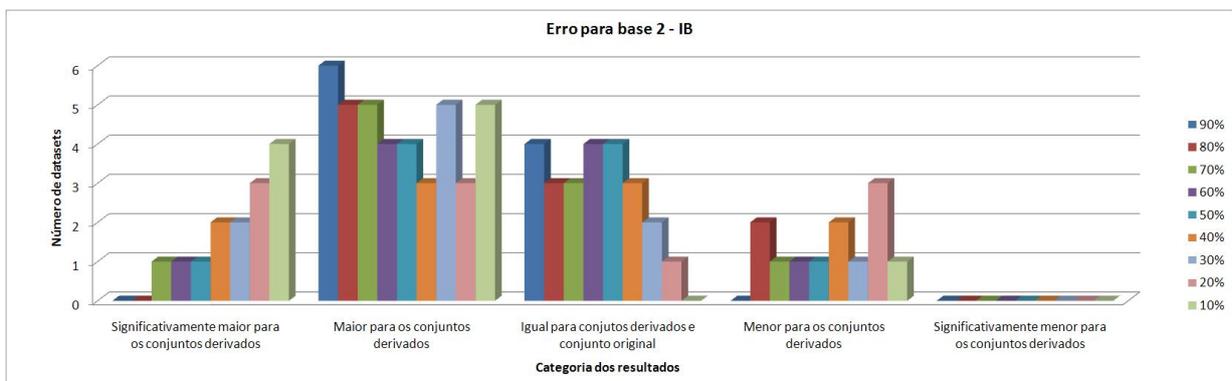


Figura 24: Resumo dos resultados da taxa de erro utilizando base 2 de IB

Como pode ser notado, houve uma tendência da taxa de erro aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa.

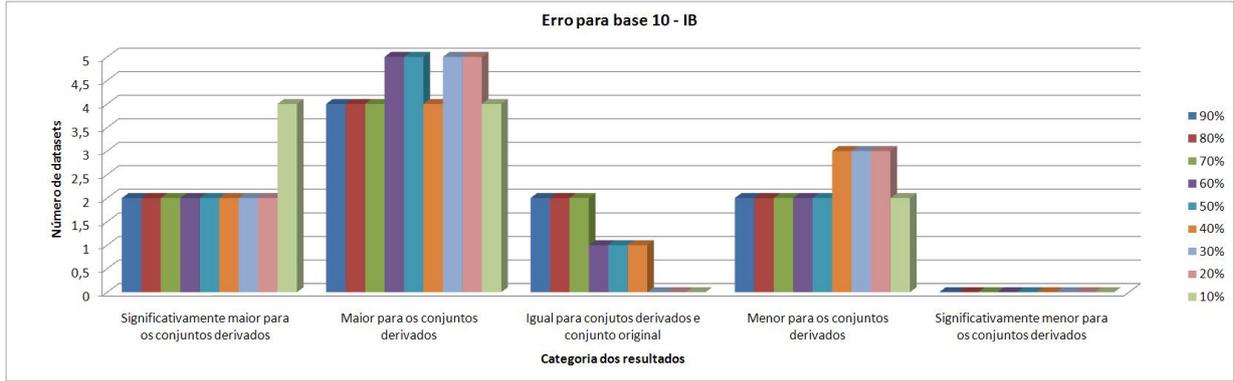


Figura 25: Resumo dos resultados da taxa de erro utilizando base 10 de IB

Como pode ser visto, houve uma tendência da taxa de erro aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa.

## 6.2 Resultados do indutor $J_{48}$

### 6.2.1 Resultados aml-all-completo

Na Tabela 14 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos aml-all-completo original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
aml-all-completo	1,000 $\pm$ 0,114	1,000 $\pm$ 0,114	22,143 $\pm$ 5,954	22,143 $\pm$ 5,954	4,200 $\pm$ 1,033	4,200 $\pm$ 1,033
aml-all-completo-90%	1,177 $\pm$ 0,149	1,000 $\pm$ 0,114	20,893 $\pm$ 5,722	20,893 $\pm$ 5,722	4,200 $\pm$ 1,033	4,200 $\pm$ 1,033
aml-all-completo-80%	1,113 $\pm$ 0,140	0,969 $\pm$ 0,129	20,893 $\pm$ 5,722	22,321 $\pm$ 5,717	4,000 $\pm$ 1,054	4,000 $\pm$ 1,054
aml-all-completo-70%	1,208 $\pm$ 0,283	0,973 $\pm$ 0,163	19,464 $\pm$ 6,073	20,893 $\pm$ 6,105	4,200 $\pm$ 1,033	4,000 $\pm$ 1,054
aml-all-completo-60%	0,927 $\pm$ 0,106	1,265 $\pm$ 0,422	19,464 $\pm$ 6,435	15,357 $\pm$ 6,819	4,200 $\pm$ 1,033	4,200 $\pm$ 1,033
aml-all-completo-50%	1,028 $\pm$ 0,155	0,814 $\pm$ 0,091	19,643 $\pm$ 6,761	16,786 $\pm$ 6,943	4,400 $\pm$ 0,966	4,400 $\pm$ 0,966
aml-all-completo-40%	0,997 $\pm$ 0,130	0,944 $\pm$ 0,162	16,786 $\pm$ 5,052	21,071 $\pm$ 6,443	4,400 $\pm$ 0,966	4,400 $\pm$ 0,966
aml-all-completo-30%	0,873 $\pm$ 0,159	0,952 $\pm$ 0,228	15,357 $\pm$ 5,735	20,000 $\pm$ 6,098	4,400 $\pm$ 0,966	4,600 $\pm$ 0,843
aml-all-completo-20%	0,705 $\pm$ 0,094	0,746 $\pm$ 0,083	14,286 $\pm$ 5,634	22,857 $\pm$ 5,714	4,400 $\pm$ 0,966	5,000 $\pm$ 0,943
aml-all-completo-10%	0,876 $\pm$ 0,133	0,836 $\pm$ 0,048	30,714 $\pm$ 6,393	63,929 $\pm$ 3,021	6,200 $\pm$ 1,033	6,600 $\pm$ 0,843

Tabela 14: Tempo de indução e taxa de erro do classificador para o indutor  $J_{48}$  utilizando arredondamento com bases 2 e 10 aml-all-completo

Na Figura 26 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre aml-all-completo e aml-all-completo-90%, entre aml-all-completo e aml-all-completo-80% e assim por diante, utilizando base 2 e base 10.

No geral, o tempo de indução aumentou de forma não significativa para os conjuntos aml-all-completo-90% até aml-all-completo-60% utilizando arredondamento com bases 2 e 10. Para o restante dos conjuntos, no geral, o tempo de indução diminuiu de forma não significativa, exceto para o conjunto aml-all-completo-20% que teve uma redução significativa, para ambas as bases.

Analogamente, na Figura 27 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 22,143% (aml-all-

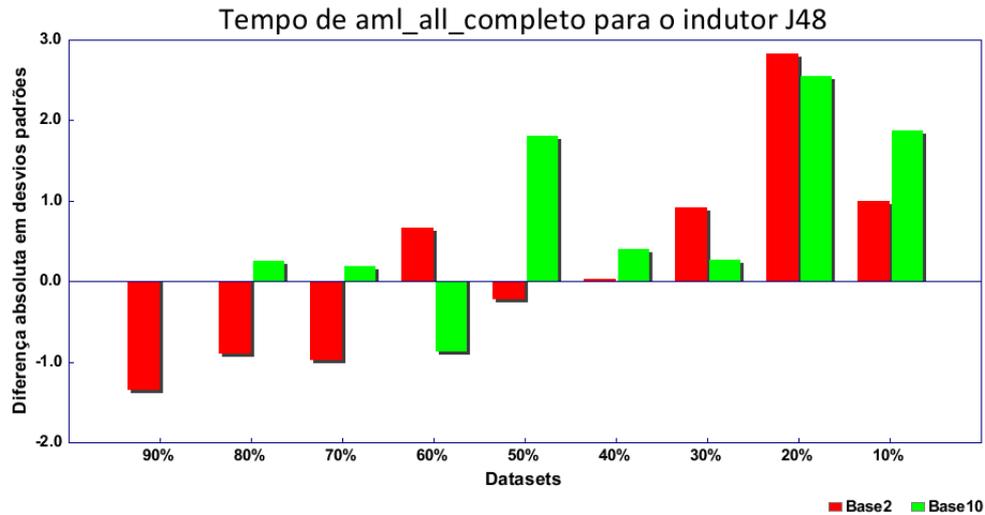


Figura 26: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) aml-all-completo

completo), para 19,722% utilizando o arredondamento com a base 2, o que representa uma redução de 10,932% da taxa de erro. E aumentou para 24,901% utilizando o arredondamento com a base 10, representando assim um aumento de 12,454%.

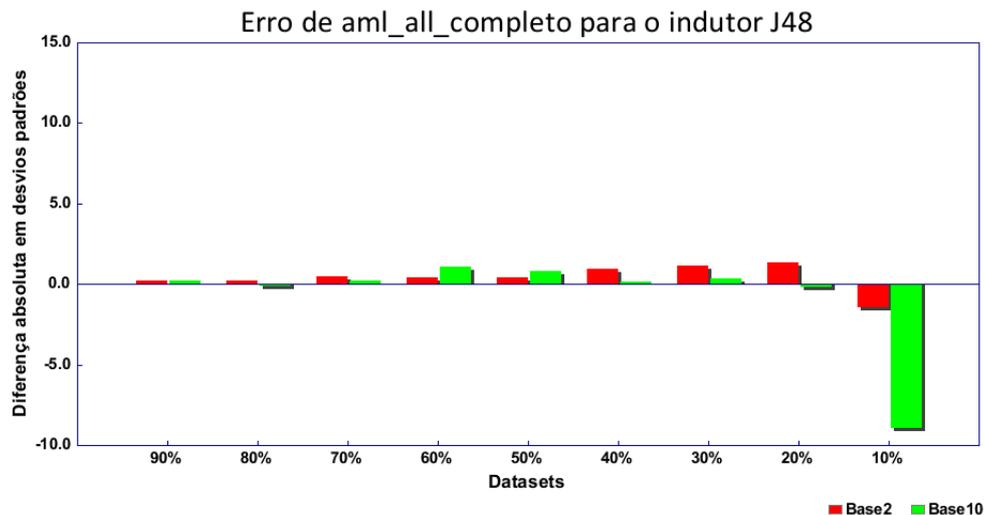


Figura 27: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) aml-all-completo

Como pode ser notado, no geral, a taxa de erro reduziu para os conjuntos de forma não significativa, exceto para aml-all-completo-10% que a taxa de erro aumentou significativamente utilizando arredondamento para a base 10.

Na Figura 28 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico, utilizando arredondamento com base 2 (binária) e base 10 (decimal). Em média, o tamanho da árvore aumentou de 4,200 (aml-all-completo) para 4,489 (aml-all-completo-base2) — média aritmética dos conjuntos arredondados utilizando base 2 — e aumentou para 4,600 (aml-all-completo-base10) — média aritmética dos conjuntos arredondados utilizando base 10. Isso significa um aumento de 6,878% para (aml-all-completo-base2) e de 9,524% para (aml-

all-completo-base10) do tamanho da árvore. Como pode ser observado, no geral o tamanho da árvore aumentou de forma não significativa para ambas as bases.

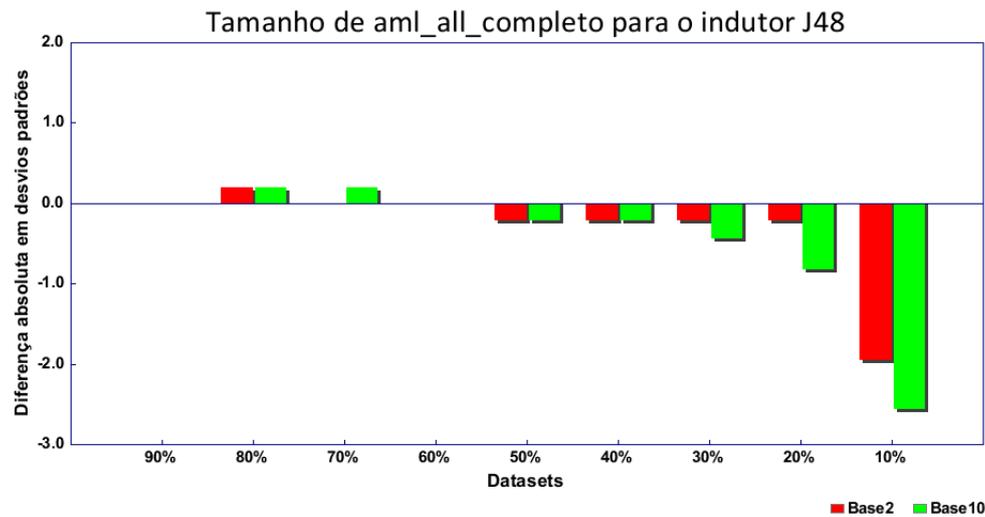


Figura 28: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) aml-all-completo

## 6.2.2 Resultados eucalyptus

Na Tabela 15 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos eucalyptus original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
eucalyptus	1,000 $\pm$ 0,032	1,000 $\pm$ 0,032	38,476 $\pm$ 1,660	38,476 $\pm$ 1,660	126,200 $\pm$ 29,705	126,200 $\pm$ 29,705
eucalyptus-90%	0,879 $\pm$ 0,044	0,786 $\pm$ 0,029	46,211 $\pm$ 1,544	45,676 $\pm$ 1,847	146,900 $\pm$ 31,028	125,300 $\pm$ 53,460
eucalyptus-80%	0,888 $\pm$ 0,022	0,814 $\pm$ 0,024	47,151 $\pm$ 1,215	45,676 $\pm$ 1,847	142,400 $\pm$ 45,571	125,300 $\pm$ 53,460
eucalyptus-70%	0,866 $\pm$ 0,023	0,804 $\pm$ 0,027	47,290 $\pm$ 1,700	45,676 $\pm$ 1,847	162,900 $\pm$ 56,363	125,300 $\pm$ 53,460
eucalyptus-60%	0,848 $\pm$ 0,036	0,773 $\pm$ 0,037	46,477 $\pm$ 1,204	45,542 $\pm$ 1,962	150,900 $\pm$ 41,616	128,300 $\pm$ 60,150
eucalyptus-50%	0,848 $\pm$ 0,021	0,798 $\pm$ 0,021	48,784 $\pm$ 1,494	45,813 $\pm$ 1,914	147,400 $\pm$ 30,486	130,400 $\pm$ 56,010
eucalyptus-40%	0,786 $\pm$ 0,042	0,776 $\pm$ 0,021	49,459 $\pm$ 1,633	46,625 $\pm$ 1,981	176,100 $\pm$ 47,290	140,100 $\pm$ 43,600
eucalyptus-30%	0,798 $\pm$ 0,036	0,801 $\pm$ 0,043	46,633 $\pm$ 2,481	47,701 $\pm$ 2,755	141,400 $\pm$ 32,143	132,800 $\pm$ 50,644
eucalyptus-20%	0,748 $\pm$ 0,045	0,702 $\pm$ 0,053	49,998 $\pm$ 2,470	53,793 $\pm$ 2,017	192,800 $\pm$ 34,867	168,800 $\pm$ 38,183
eucalyptus-10%	0,674 $\pm$ 0,036	0,693 $\pm$ 0,044	45,927 $\pm$ 1,655	53,386 $\pm$ 2,390	279,400 $\pm$ 54,052	270,200 $\pm$ 38,989

Tabela 15: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 eucalyptus

Na Figura 29 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre eucalyptus e eucalyptus-90%, entre eucalyptus e eucalyptus-80% e assim por diante, utilizando base 2 e base 10.

Como pode ser visto, o tempo de indução reduziu para todos os conjuntos de forma significativa utilizando arredondamento com as bases 2 e 10.

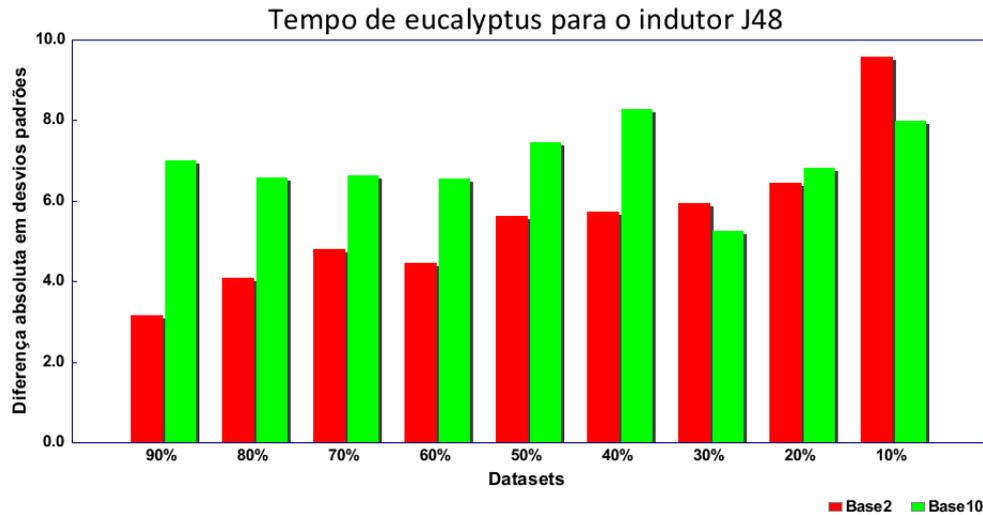


Figura 29: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) eucalyptus

Analogamente, na Figura 30 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 38,476% (eucalyptus), para 47,548% utilizando o arredondamento com a base 2, o que representa um aumento de 23,578% da taxa de erro. E aumentou para 47,765% utilizando o arredondamento com a base 10, representando assim um aumento de 24,143%.

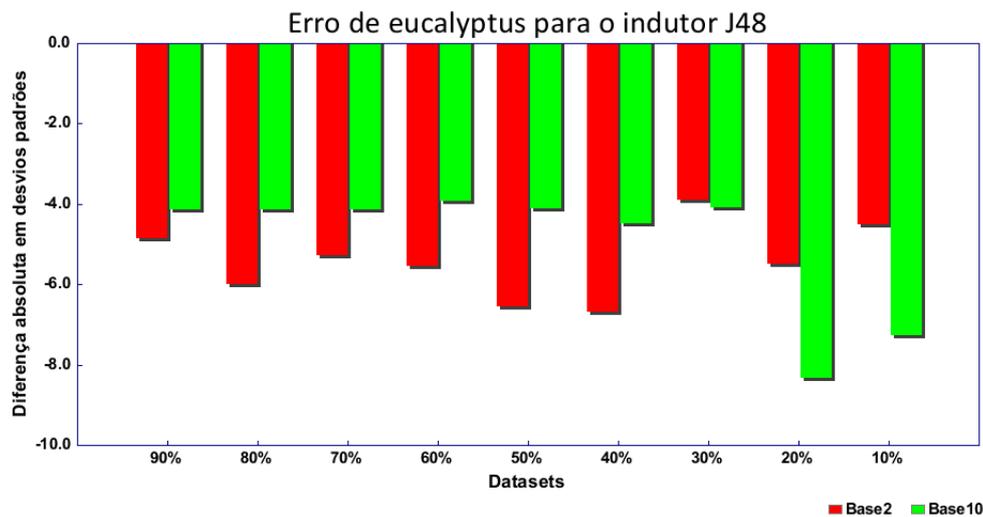


Figura 30: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) eucalyptus

A taxa de erro aumentou para todos os conjuntos de forma significativa utilizando arredondamento com ambas as bases.

Na Figura 31 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 126,200 (eucalyptus) para 171,133 (eucalyptus-base2), e 149,611 para (eucalyptus-base10). Isso significa um aumento de 35,605% para (eucalyptus-base2) e de 18,551% para (eucalyptus-base10) do tamanho da árvore. Como pode ser notado o tamanho da árvore aumentou para todos os conjuntos, embora de forma

não significativa para ambas as bases, exceto para os conjuntos eucalyptus-20%, que aumentou significativamente com base 2 e eucalyptus-10% com ambas as bases.

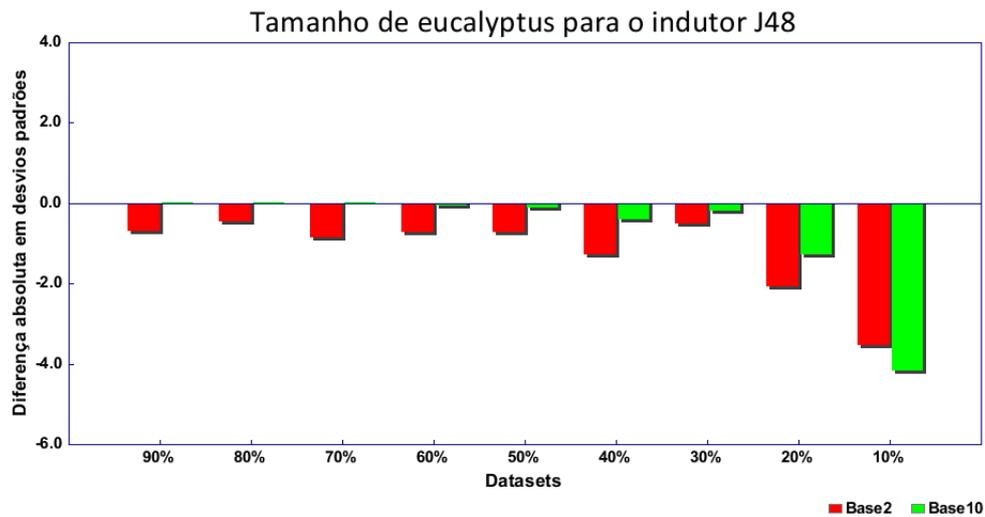


Figura 31: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) eucalyptus

### 6.2.3 Resultados hepatitis

Na Tabela 16 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos hepatitis original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
hepatitis	1,000 $\pm$ 0,167	1,000 $\pm$ 0,167	19,333 $\pm$ 3,539	19,333 $\pm$ 3,539	17,400 $\pm$ 4,195	17,400 $\pm$ 4,195
hepatitis-90%	0,886 $\pm$ 0,088	1,061 $\pm$ 0,434	20,583 $\pm$ 3,862	23,833 $\pm$ 2,724	15,000 $\pm$ 3,771	17,400 $\pm$ 5,481
hepatitis-80%	0,818 $\pm$ 0,086	0,841 $\pm$ 0,083	21,917 $\pm$ 3,910	23,833 $\pm$ 2,724	16,200 $\pm$ 3,293	17,400 $\pm$ 5,481
hepatitis-70%	0,826 $\pm$ 0,075	0,856 $\pm$ 0,037	21,917 $\pm$ 3,910	23,833 $\pm$ 2,724	16,600 $\pm$ 3,098	17,400 $\pm$ 5,481
hepatitis-60%	0,818 $\pm$ 0,060	0,795 $\pm$ 0,054	21,917 $\pm$ 3,910	23,833 $\pm$ 2,724	17,000 $\pm$ 3,127	17,400 $\pm$ 5,481
hepatitis-50%	0,795 $\pm$ 0,040	0,841 $\pm$ 0,056	22,000 $\pm$ 5,168	23,833 $\pm$ 2,724	17,200 $\pm$ 4,264	17,400 $\pm$ 5,481
hepatitis-40%	0,826 $\pm$ 0,043	0,803 $\pm$ 0,039	22,708 $\pm$ 4,716	23,833 $\pm$ 2,724	20,200 $\pm$ 2,700	17,400 $\pm$ 5,481
hepatitis-30%	0,826 $\pm$ 0,056	0,795 $\pm$ 0,054	22,667 $\pm$ 3,417	23,833 $\pm$ 2,724	18,600 $\pm$ 4,881	17,400 $\pm$ 5,641
hepatitis-20%	0,833 $\pm$ 0,071	0,818 $\pm$ 0,060	19,333 $\pm$ 3,843	23,167 $\pm$ 2,564	17,200 $\pm$ 5,371	13,600 $\pm$ 4,427
hepatitis-10%	0,848 $\pm$ 0,117	0,811 $\pm$ 0,080	17,417 $\pm$ 1,945	23,125 $\pm$ 3,035	10,600 $\pm$ 5,232	12,400 $\pm$ 5,168

Tabela 16: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 hepatitis

Na Figura 32 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre hepatitis e hepatitis-90%, entre hepatitis e hepatitis-80% e assim por diante, utilizando base 2 e base 10.

O tempo de indução reduziu para todos os conjuntos, sendo de forma não significativa, tanto para base 2 como para a base 10, exceto para o conjunto hepatitis-90%, que teve um tempo de indução maior com a base 10, porém de forma não significativa.

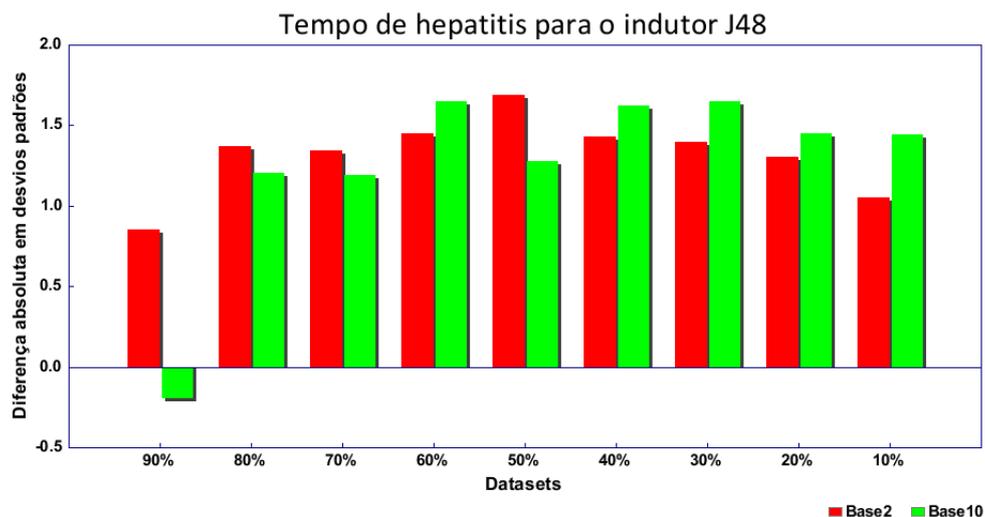


Figura 32: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

Analogamente, na Figura 33 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 19,333% (hepatitis), para 21,162% utilizando o arredondamento com a base 2, o que representa um aumento de 9,461% da taxa de erro. E aumentou para 23,680% utilizando o arredondamento com a base 10, representando assim um aumento de 22,487%.

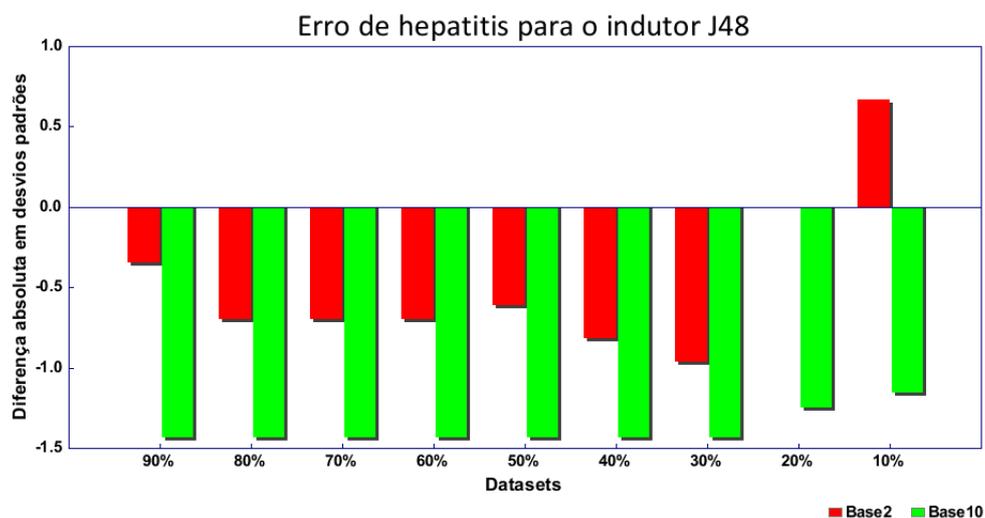


Figura 33: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

Como pode ser observado, a taxa de erro aumentou para todos os conjuntos de forma não significativa, utilizando arredondamento com a bases 2 e 10, exceto hepatitis-10% que teve uma menor taxa de erro com a base 2, embora de forma não significativa.

Na Figura 34 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore diminuiu de 17,400 (hepatitis) para 16,511 (hepatitis-base2), e 16,422 para (hepatitis-base10). Isso significa uma redução de 5,109% para (hepatitis-base2) e de 5,619% para (hepatitis-base10) do tamanho da árvore. Como pode ser

notado, no geral, o tamanho da árvore reduziu, embora de forma não significativa para ambas as bases.

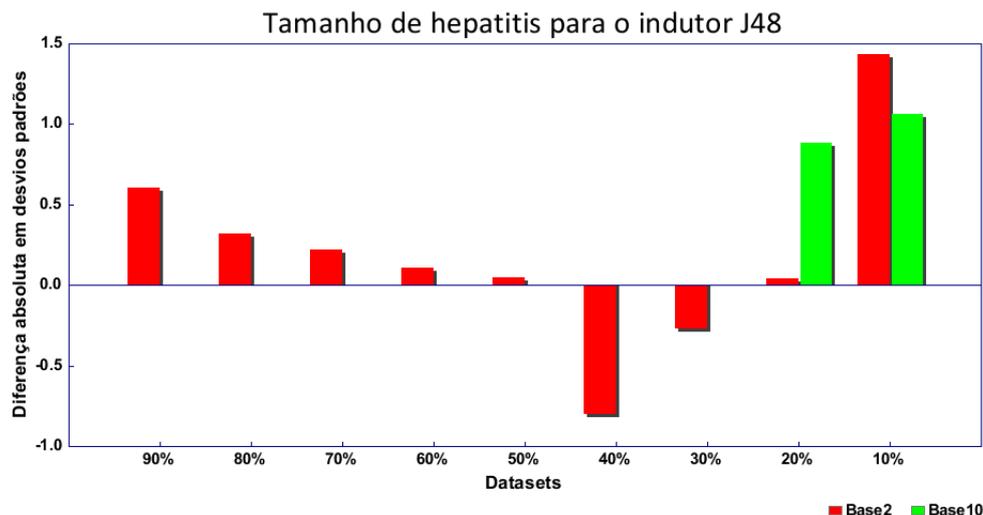


Figura 34: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) hepatitis

## 6.2.4 Resultados ionosphere

Na Tabela 17 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *ionosphere* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
<i>ionosphere</i>	1,000 $\pm$ 0,069	1,000 $\pm$ 0,069	10,540 $\pm$ 1,348	10,540 $\pm$ 1,348	17,400 $\pm$ 4,195	17,400 $\pm$ 4,195
<i>ionosphere-90%</i>	0,927 $\pm$ 0,055	0,818 $\pm$ 0,049	10,841 $\pm$ 1,588	9,127 $\pm$ 1,195	23,600 $\pm$ 4,993	24,000 $\pm$ 4,546
<i>ionosphere-80%</i>	0,895 $\pm$ 0,043	0,818 $\pm$ 0,039	11,397 $\pm$ 1,278	9,127 $\pm$ 1,195	22,200 $\pm$ 4,733	24,000 $\pm$ 4,546
<i>ionosphere-70%</i>	0,862 $\pm$ 0,025	0,836 $\pm$ 0,038	11,127 $\pm$ 1,317	9,127 $\pm$ 1,195	24,400 $\pm$ 4,222	24,000 $\pm$ 4,546
<i>ionosphere-60%</i>	0,847 $\pm$ 0,034	0,822 $\pm$ 0,043	9,413 $\pm$ 1,421	9,127 $\pm$ 1,195	25,000 $\pm$ 4,320	24,000 $\pm$ 4,546
<i>ionosphere-50%</i>	0,789 $\pm$ 0,025	0,760 $\pm$ 0,044	10,540 $\pm$ 1,348	9,127 $\pm$ 1,404	24,600 $\pm$ 4,088	23,600 $\pm$ 4,904
<i>ionosphere-40%</i>	0,767 $\pm$ 0,040	0,680 $\pm$ 0,018	10,556 $\pm$ 0,867	7,984 $\pm$ 1,466	24,800 $\pm$ 5,371	24,800 $\pm$ 4,849
<i>ionosphere-30%</i>	0,727 $\pm$ 0,024	0,691 $\pm$ 0,017	10,556 $\pm$ 0,867	6,841 $\pm$ 1,145	23,200 $\pm$ 3,190	25,000 $\pm$ 4,522
<i>ionosphere-20%</i>	0,705 $\pm$ 0,019	0,687 $\pm$ 0,021	11,405 $\pm$ 0,961	6,841 $\pm$ 1,145	24,000 $\pm$ 4,137	25,000 $\pm$ 4,522
<i>ionosphere-10%</i>	0,676 $\pm$ 0,019	0,684 $\pm$ 0,023	10,833 $\pm$ 1,116	6,841 $\pm$ 1,145	21,800 $\pm$ 6,546	25,400 $\pm$ 4,695

Tabela 17: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 *ionosphere*

Na Figura 35 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *ionosphere* e *ionosphere-90%*, entre *ionosphere* e *ionosphere-80%* e assim por diante, utilizando base 2 e base 10.

O tempo de indução reduziu para todos os conjuntos utilizando arredondamento, de forma significativa, tanto para base 2 como para a base 10, exceto para os conjuntos *ionosphere-90%* e *ionosphere-80%*, que o seu tempo de indução reduziu de forma não significativa para o

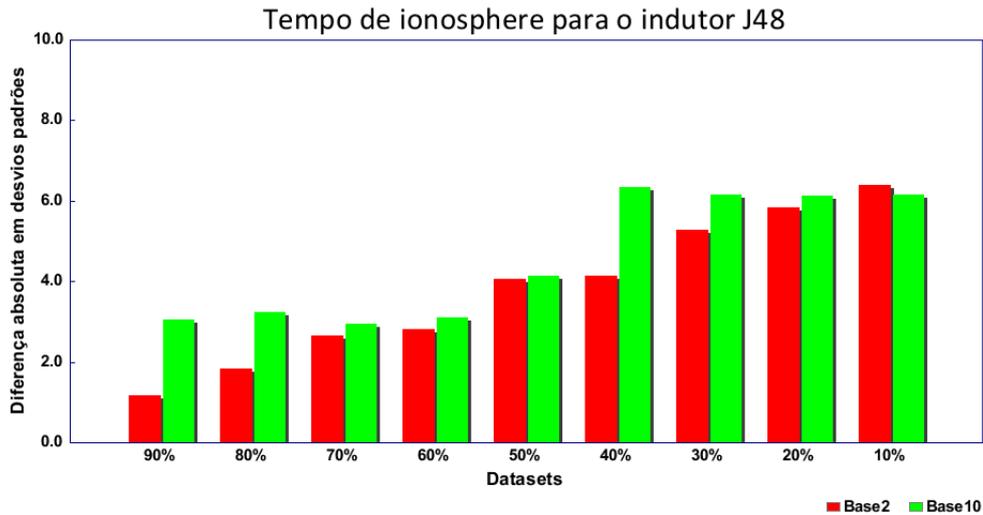


Figura 35: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) ionosphere

arredondamento utilizando base 2.

Analogamente, na Figura 36 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 10,540% (ionosphere), para 10,741% utilizando o arredondamento com a base 2, o que representa um aumento de 1,906% da taxa de erro. E reduziu para 8,238% utilizando o arredondamento com a base 10, representando assim uma redução de 21,841%.

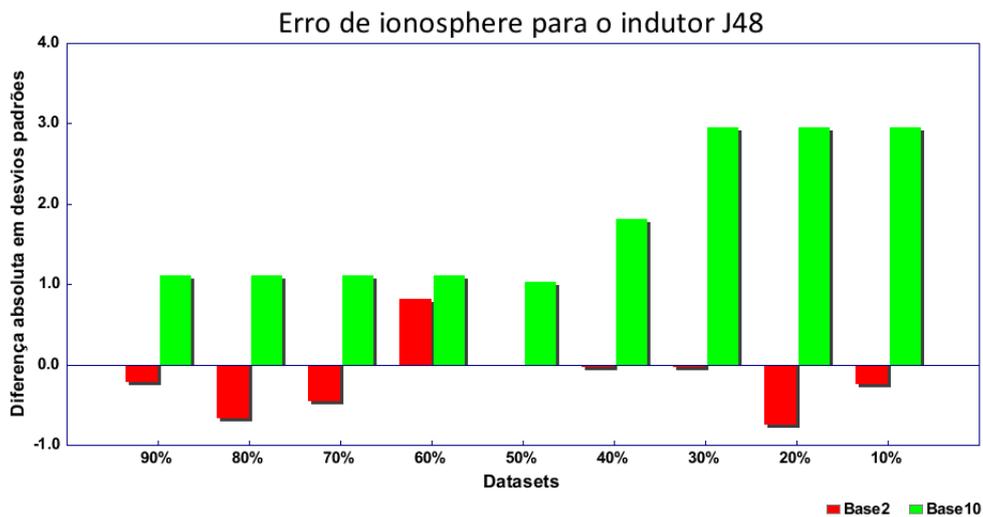


Figura 36: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) ionosphere

Como pode ser visto, para a maioria dos conjuntos que realizou-se o arredondamento com a base 2, a taxa de erro aumentou de forma não significativa. Já para a base 10, todos os conjuntos tiveram uma menor taxa de erro, sendo uma redução significativa para os conjuntos ionosphere-30%, ionosphere-20% e ionosphere-10%.

Na Figura 37 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 17,400 (ionosphere)

para 23,733 (ionosphere-base2), e para 24,422 (ionosphere-base10). Isso significa um aumento de 36,398% para (ionosphere-base2) e de 40,358% para (ionosphere-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para todos os conjuntos, embora de forma não significativa para ambas as bases.

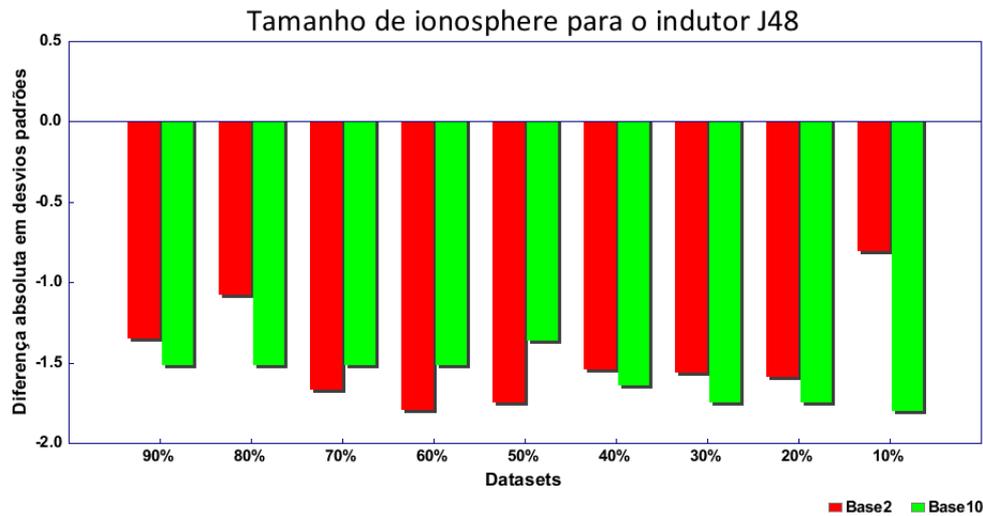


Figura 37: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) ionosphere

### 6.2.5 Resultados lymph

Na Tabela 18 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *lymph* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
lymph	1,000 $\pm$ 0,078	1,000 $\pm$ 0,078	19,619 $\pm$ 3,068	19,619 $\pm$ 3,068	17,400 $\pm$ 4,195	17,400 $\pm$ 4,195
lymph-90%	0,989 $\pm$ 0,080	1,011 $\pm$ 0,073	21,714 $\pm$ 3,647	24,286 $\pm$ 4,659	27,300 $\pm$ 5,250	21,600 $\pm$ 6,736
lymph-80%	0,977 $\pm$ 0,061	0,989 $\pm$ 0,080	21,714 $\pm$ 3,647	24,286 $\pm$ 4,659	27,300 $\pm$ 5,250	21,600 $\pm$ 6,736
lymph-70%	0,977 $\pm$ 0,081	1,034 $\pm$ 0,094	21,714 $\pm$ 3,647	24,286 $\pm$ 4,659	27,300 $\pm$ 5,250	21,600 $\pm$ 6,736
lymph-60%	0,989 $\pm$ 0,097	1,000 $\pm$ 0,078	21,048 $\pm$ 3,742	24,286 $\pm$ 4,659	27,300 $\pm$ 5,250	21,600 $\pm$ 6,736
lymph-50%	0,966 $\pm$ 0,059	1,000 $\pm$ 0,078	21,048 $\pm$ 3,742	24,286 $\pm$ 4,659	27,300 $\pm$ 5,250	21,600 $\pm$ 6,736
lymph-40%	1,000 $\pm$ 0,095	1,000 $\pm$ 0,078	22,286 $\pm$ 3,290	24,286 $\pm$ 4,659	21,600 $\pm$ 6,736	21,600 $\pm$ 6,736
lymph-30%	0,989 $\pm$ 0,059	0,989 $\pm$ 0,080	22,286 $\pm$ 3,290	24,286 $\pm$ 4,659	21,600 $\pm$ 6,736	21,600 $\pm$ 6,736
lymph-20%	0,885 $\pm$ 0,095	0,920 $\pm$ 0,077	19,619 $\pm$ 2,539	19,619 $\pm$ 2,539	23,000 $\pm$ 6,815	23,000 $\pm$ 6,815
lymph-10%	0,862 $\pm$ 0,061	0,839 $\pm$ 0,056	19,619 $\pm$ 2,539	19,619 $\pm$ 2,539	23,000 $\pm$ 6,815	23,000 $\pm$ 6,815

Tabela 18: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 *lymph*

Na Figura 38 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *lymph* e *lymph-90%*, entre *lymph* e *lymph-80%* e assim por diante, utilizando base 2 e base 10.

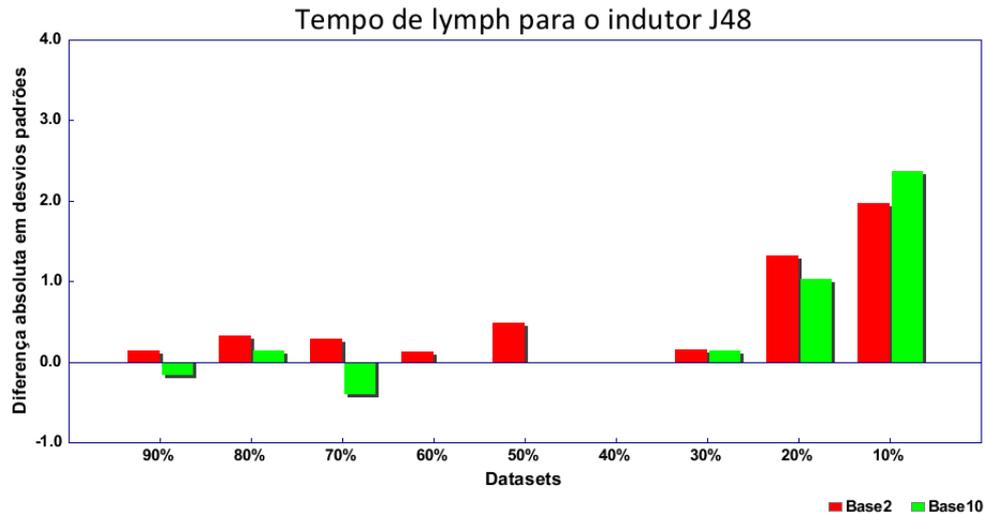


Figura 38: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

No geral, o tempo de indução reduziu para a maioria dos conjuntos utilizando arredondamento com ambas as bases, sendo de forma significativa apenas para o conjunto lymph-10%, para o arredondamento com a base 10.

Analogamente, na Figura 39 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 19,619% (lymph), para 21,228% utilizando o arredondamento com a base 2, o que representa um aumento de 8,199% da taxa de erro. E aumentou para 23,249% utilizando o arredondamento com a base 10, representando assim um aumento de 18,502%.

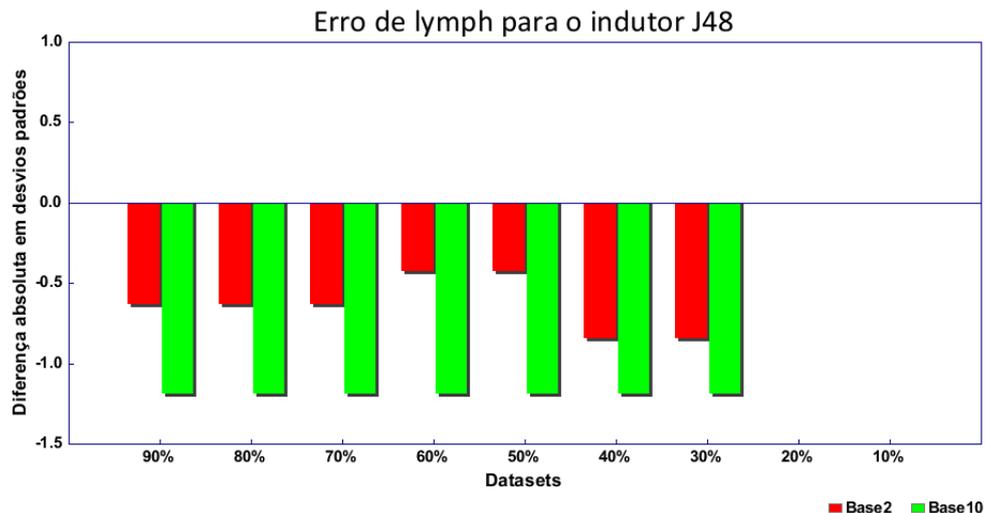


Figura 39: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

Como pode ser notado, a taxa de erro aumentou para todos os conjuntos de forma não significativa, utilizando arredondamento com as bases 2 e 10.

Na Figura 40 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 17,400 (lymph) para

25,078 (lymph-base2), e para 21,911 (lymph-base10). Isso significa um aumento de 44,125% para (lymph-base2) e de 25,926% para (lymph-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para todos os conjuntos, sendo de forma significativa para a maioria dos conjuntos utilizando o arredondamento com a base 2, e aumentou de forma não significativa para todos os conjuntos arredondados com a base 10.

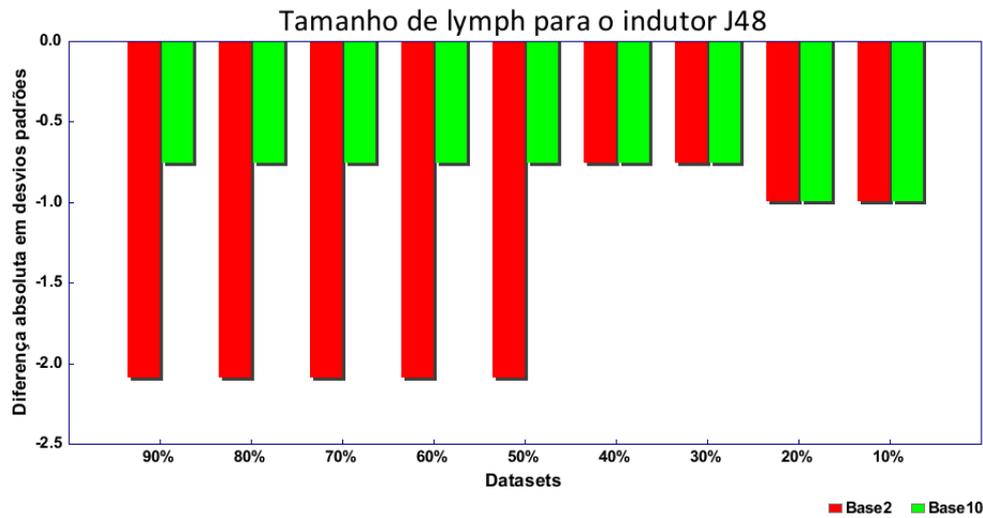


Figura 40: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) lymph

### 6.2.6 Resultados sick

Na Tabela 19 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sick original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
sick	1,000 $\pm$ 0,151	1,000 $\pm$ 0,151	1,272 $\pm$ 0,141	1,272 $\pm$ 0,141	48,300 $\pm$ 5,618	48,300 $\pm$ 5,618
sick-90%	0,651 $\pm$ 0,036	0,887 $\pm$ 0,048	1,140 $\pm$ 0,143	4,427 $\pm$ 0,220	43,200 $\pm$ 3,155	98,400 $\pm$ 16,133
sick-80%	0,697 $\pm$ 0,038	0,885 $\pm$ 0,046	2,015 $\pm$ 0,194	4,427 $\pm$ 0,220	52,700 $\pm$ 6,734	98,400 $\pm$ 16,133
sick-70%	0,694 $\pm$ 0,035	0,889 $\pm$ 0,044	2,121 $\pm$ 0,172	4,427 $\pm$ 0,220	53,000 $\pm$ 5,228	98,400 $\pm$ 16,133
sick-60%	0,695 $\pm$ 0,040	0,885 $\pm$ 0,040	2,200 $\pm$ 0,194	4,507 $\pm$ 0,323	53,000 $\pm$ 5,754	98,600 $\pm$ 14,546
sick-50%	0,682 $\pm$ 0,036	0,887 $\pm$ 0,043	2,174 $\pm$ 0,184	4,507 $\pm$ 0,323	52,500 $\pm$ 6,060	98,600 $\pm$ 14,546
sick-40%	0,727 $\pm$ 0,025	0,882 $\pm$ 0,042	2,094 $\pm$ 0,257	4,507 $\pm$ 0,323	45,500 $\pm$ 3,689	98,600 $\pm$ 14,546
sick-30%	0,730 $\pm$ 0,022	0,876 $\pm$ 0,034	2,068 $\pm$ 0,303	4,401 $\pm$ 0,296	44,200 $\pm$ 5,692	76,600 $\pm$ 11,027
sick-20%	0,938 $\pm$ 0,065	0,887 $\pm$ 0,035	2,333 $\pm$ 0,293	4,401 $\pm$ 0,296	40,800 $\pm$ 10,675	76,600 $\pm$ 11,027
sick-10%	1,177 $\pm$ 0,048	1,124 $\pm$ 0,050	5,753 $\pm$ 0,209	6,124 $\pm$ 0,025	34,600 $\pm$ 6,995	1,000 $\pm$ 0,000

Tabela 19: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 sick

Na Figura 41 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre sick e sick-90%, entre sick e sick-80% e assim por diante, utilizando base 2 e base 10.

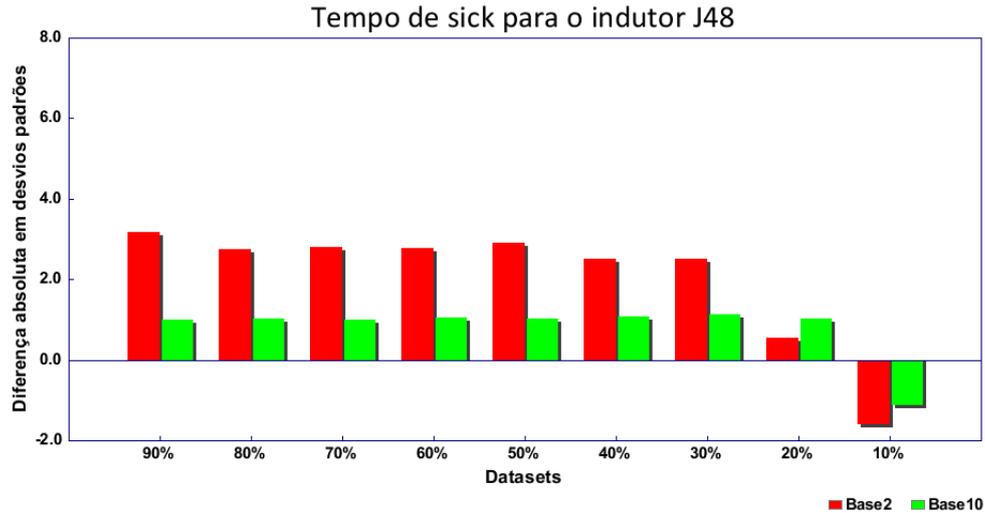


Figura 41: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

Como pode ser visto, o tempo de indução reduziu para os conjuntos utilizando arredondamento com base 2, sendo de forma significativa para a maioria dos conjuntos. Já para a base 10, também a maioria dos conjuntos teve uma redução no tempo de indução, porém de forma não significativa.

Analogamente, na Figura 42 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 1,272% (sick), para 2,433% utilizando o arredondamento com a base 2, o que representa um aumento de 91,282% da taxa de erro. E aumentou para 4,636% utilizando o arredondamento com a base 10, representando assim um aumento de 264,500%.

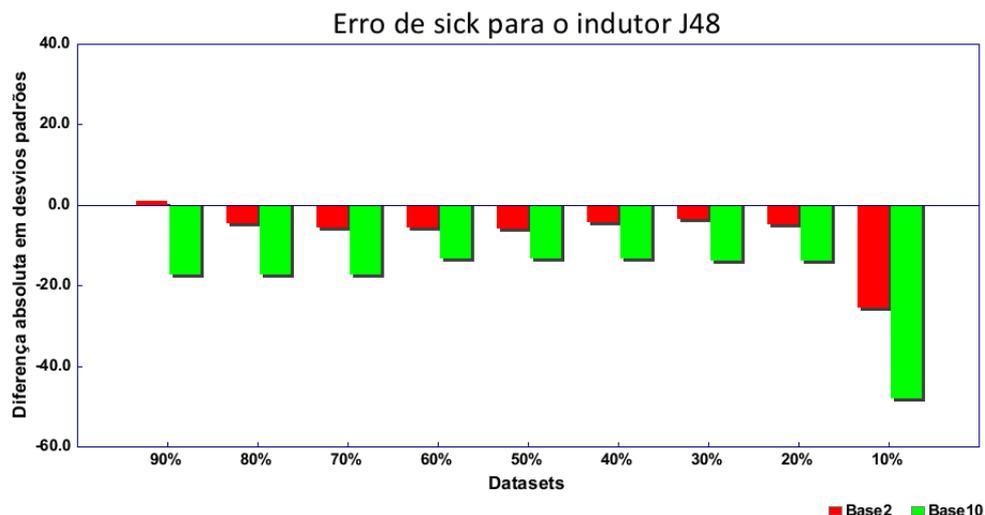


Figura 42: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

A taxa de erro aumentou de forma significativa para todos os conjuntos utilizando arredondamento com ambas as bases, exceto para (sick-90%) que teve uma menor taxa de erro para a base 2, embora de forma não significativa.

Na Figura 43 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore reduziu de 48,300 (sick) para 46,611 (sick-base2), e aumentou para 82,800 (sick-base10). Isso significa uma redução de 3,497% para (sick-base2) e um aumento de 71,429% para (sick-base10) do tamanho da árvore. Para o arredondamento utilizando a base 2, aumentou para 4 conjuntos de forma não significativa e reduziu para 5 conjuntos, sendo o conjunto (sick-10%) o único com uma redução significativa. Já para o arredondamento utilizando a base 10, o tamanho da árvore aumentou para todos os conjuntos de forma não significativa, exceto para (sick-10%) que teve uma redução significativa.

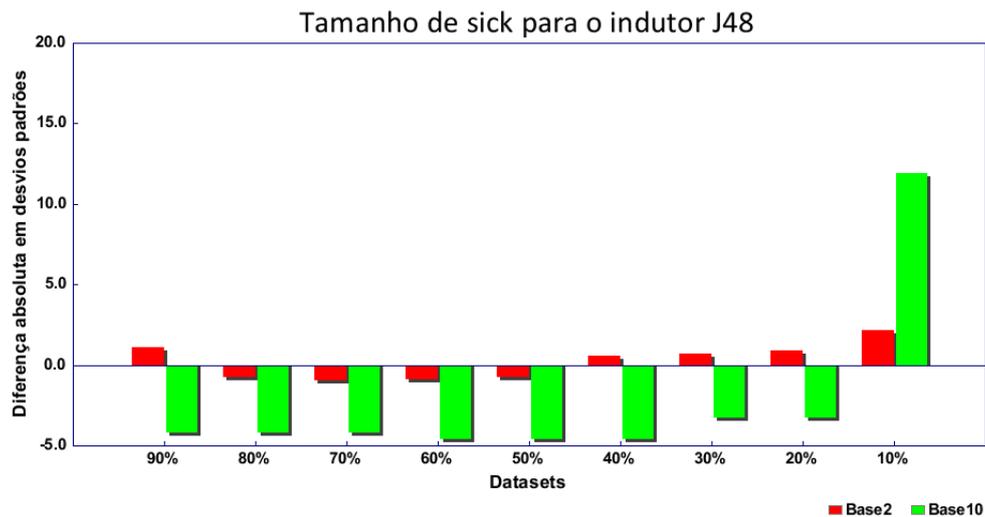


Figura 43: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) sick

## 6.2.7 Resultados sonar

Na Tabela 20 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sonar original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
sonar	1,000 $\pm$ 0,075	1,000 $\pm$ 0,075	27,405 $\pm$ 2,770	27,405 $\pm$ 2,770	27,800 $\pm$ 3,155	27,800 $\pm$ 3,155
sonar-90%	0,895 $\pm$ 0,043	0,853 $\pm$ 0,041	27,929 $\pm$ 3,887	28,810 $\pm$ 2,556	26,200 $\pm$ 2,530	32,200 $\pm$ 2,530
sonar-80%	0,891 $\pm$ 0,032	0,845 $\pm$ 0,040	26,976 $\pm$ 4,109	28,810 $\pm$ 2,556	26,200 $\pm$ 2,530	32,200 $\pm$ 2,530
sonar-70%	0,872 $\pm$ 0,027	0,857 $\pm$ 0,050	26,976 $\pm$ 4,109	28,810 $\pm$ 2,556	26,200 $\pm$ 2,530	32,200 $\pm$ 2,530
sonar-60%	0,845 $\pm$ 0,025	0,857 $\pm$ 0,043	25,476 $\pm$ 3,623	28,810 $\pm$ 2,556	26,600 $\pm$ 3,373	32,200 $\pm$ 2,530
sonar-50%	0,829 $\pm$ 0,033	0,853 $\pm$ 0,048	25,952 $\pm$ 3,187	28,810 $\pm$ 2,556	26,400 $\pm$ 3,534	32,200 $\pm$ 2,530
sonar-40%	0,795 $\pm$ 0,020	0,775 $\pm$ 0,041	26,405 $\pm$ 3,100	22,595 $\pm$ 2,422	27,000 $\pm$ 3,127	30,400 $\pm$ 3,406
sonar-30%	0,791 $\pm$ 0,020	0,740 $\pm$ 0,043	24,071 $\pm$ 3,123	23,571 $\pm$ 2,792	28,200 $\pm$ 3,795	31,800 $\pm$ 4,442
sonar-20%	0,775 $\pm$ 0,026	0,829 $\pm$ 0,133	23,571 $\pm$ 3,733	29,810 $\pm$ 1,199	31,200 $\pm$ 3,938	35,400 $\pm$ 3,627
sonar-10%	0,756 $\pm$ 0,027	0,760 $\pm$ 0,027	20,286 $\pm$ 3,179	34,095 $\pm$ 2,774	34,400 $\pm$ 4,115	34,800 $\pm$ 2,394

Tabela 20: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 sonar

Na Figura 44 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre sonar e sonar-90%, entre sonar e sonar-80% e assim por diante, utilizando base 2 e base 10.

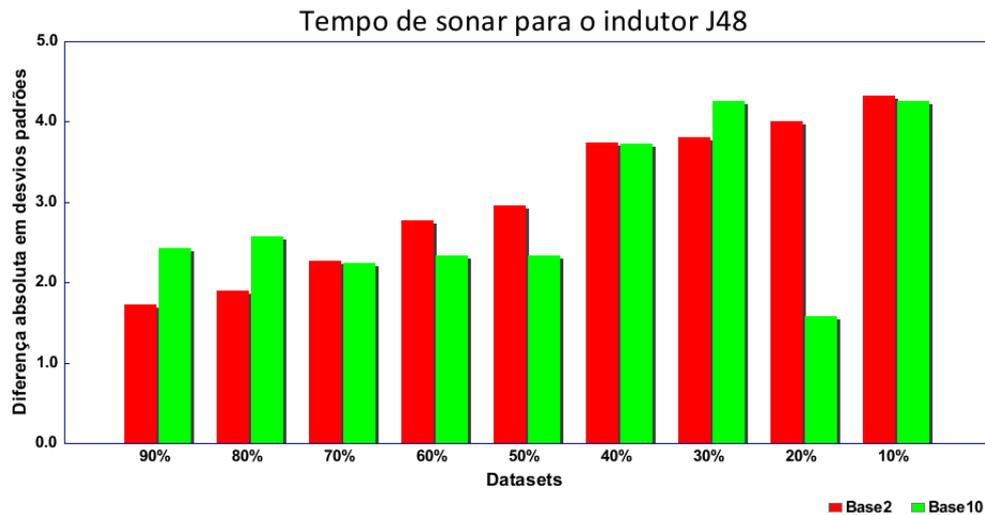


Figura 44: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

O tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo de forma significativa tanto para base 2 como para a base 10, exceto para o conjunto sonar-20%, que teve uma redução não significativa para o arredondamento utilizando a base 10.

Analogamente, na Figura 45 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 27,405% (sonar), para 25,294% utilizando o arredondamento com a base 2, o que representa uma redução de 7,705% da taxa de erro. E aumentou para 28,236% utilizando o arredondamento com a base 10, representando assim um aumento de 3,031%.

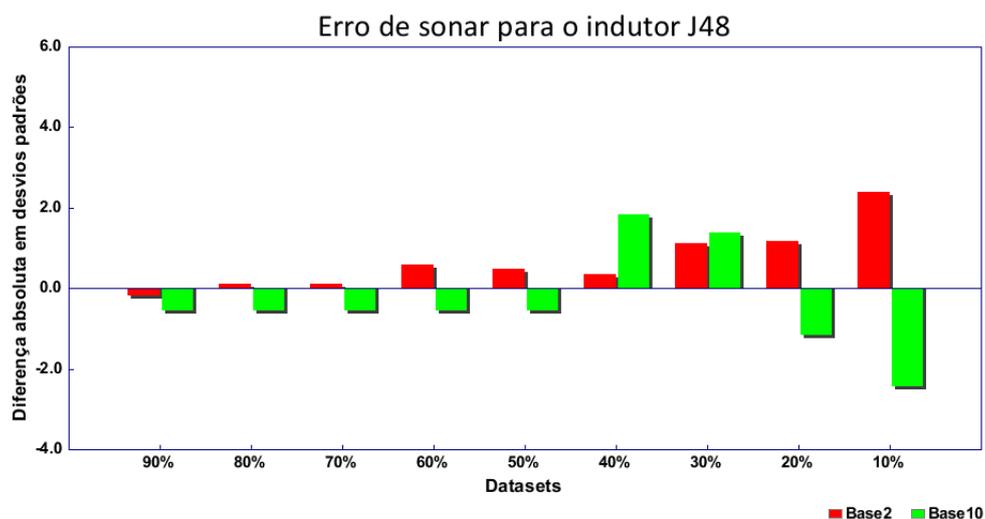


Figura 45: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

Como pode ser notado, no geral, a taxa de erro reduziu para os conjuntos utilizando

arredondamento com base 2, de forma não significativa, exceto para sonar-10% que a taxa de erro reduziu significativamente. Para o arredondamento utilizando a base 10, a maioria dos conjuntos teve a taxa de erro aumentada, embora de forma não significativa, exceto para sonar-10%, que teve uma aumento significativo.

Na Figura 46 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 27,800 (sonar) para 28,044 (sonar-base2), e para 32,600 (sonar-base10). Isso significa um aumento de 0,879% para (sonar-base2) e de 17,266% para (sonar-base10) do tamanho da árvore. O tamanho da árvore aumentou para a maioria dos conjuntos, que foram arredondados com a base 2, embora de forma não significativa. E aumentou para todos os conjuntos arredondados com a base 10, sendo de forma significativa para os conjuntos sonar-20% e sonar-10%.

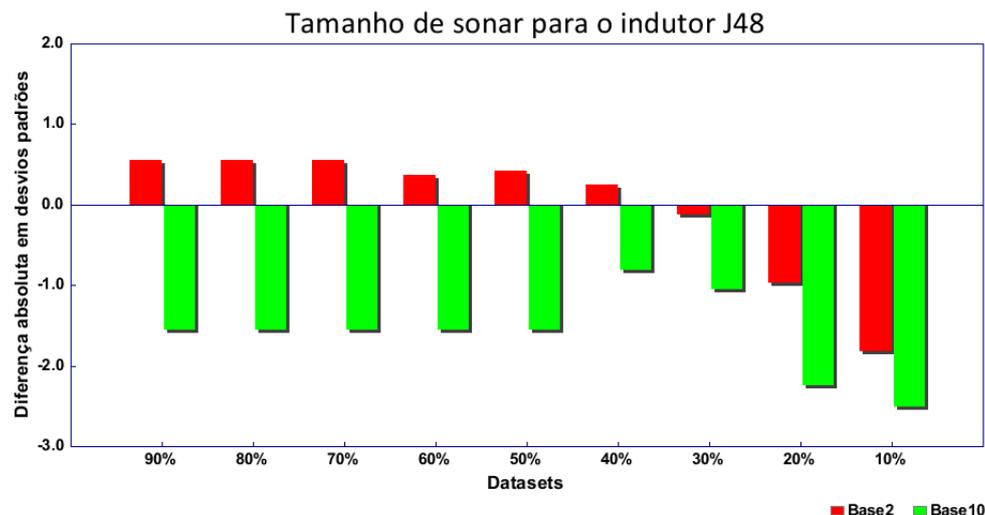


Figura 46: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) sonar

### 6.2.8 Resultados vehicle

Na Tabela 21 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos vehicle original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
vehicle	1,000 $\pm$ 0,062	1,000 $\pm$ 0,062	27,535 $\pm$ 1,361	27,535 $\pm$ 1,361	140,200 $\pm$ 23,818	140,200 $\pm$ 23,818
vehicle-90%	1,099 $\pm$ 0,043	1,055 $\pm$ 0,024	29,550 $\pm$ 2,043	41,485 $\pm$ 1,559	173,000 $\pm$ 9,475	173,200 $\pm$ 7,800
vehicle-80%	1,102 $\pm$ 0,035	1,055 $\pm$ 0,024	29,550 $\pm$ 2,043	41,485 $\pm$ 1,559	173,000 $\pm$ 9,475	173,200 $\pm$ 7,800
vehicle-70%	1,081 $\pm$ 0,074	1,055 $\pm$ 0,034	28,606 $\pm$ 1,803	41,485 $\pm$ 1,559	172,400 $\pm$ 17,964	173,200 $\pm$ 7,800
vehicle-60%	1,049 $\pm$ 0,048	1,049 $\pm$ 0,029	29,560 $\pm$ 2,043	41,485 $\pm$ 1,559	178,200 $\pm$ 12,336	173,200 $\pm$ 7,800
vehicle-50%	1,058 $\pm$ 0,063	1,058 $\pm$ 0,028	31,328 $\pm$ 1,531	41,485 $\pm$ 1,559	181,600 $\pm$ 10,543	173,200 $\pm$ 7,800
vehicle-40%	1,055 $\pm$ 0,063	1,055 $\pm$ 0,034	31,566 $\pm$ 1,556	41,485 $\pm$ 1,559	178,400 $\pm$ 11,815	173,200 $\pm$ 7,800
vehicle-30%	1,026 $\pm$ 0,066	1,055 $\pm$ 0,024	29,076 $\pm$ 1,159	41,604 $\pm$ 1,472	172,000 $\pm$ 11,441	172,200 $\pm$ 10,119
vehicle-20%	0,994 $\pm$ 0,018	1,035 $\pm$ 0,031	29,060 $\pm$ 1,695	41,845 $\pm$ 1,584	167,200 $\pm$ 17,650	170,800 $\pm$ 13,612
vehicle-10%	0,965 $\pm$ 0,036	0,637 $\pm$ 0,029	31,434 $\pm$ 0,920	66,671 $\pm$ 0,839	124,200 $\pm$ 15,754	23,200 $\pm$ 1,135

Tabela 21: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 vehicle

Na Figura 47 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre

o conjunto original e os conjuntos derivados, ou seja, entre *vehicle* e *vehicle-90%*, entre *vehicle* e *vehicle-80%* e assim por diante, utilizando base 2 e base 10.

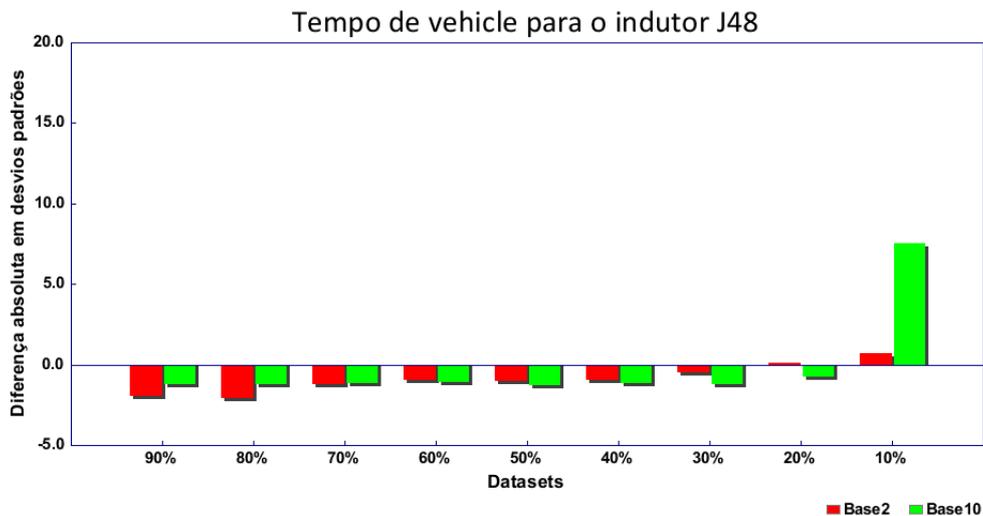


Figura 47: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *vehicle*

O tempo de indução aumentou não significativamente para a maioria dos conjuntos utilizando arredondamento com a base 2, exceto para *vehicle-80%* que teve um aumento significativo. Para o arredondamento com a base 10, o tempo de indução aumentou para todos os conjuntos de forma não significativa, exceto para o conjunto *vehicle-10%*, que o teve uma redução significativa.

Analogamente, na Figura 48 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 27,535% (*vehicle*), para 29,970% utilizando o arredondamento com a base 2, o que representa um aumento de 8,843% da taxa de erro. E aumentou para 44,337% utilizando o arredondamento com a base 10, representando assim um aumento de 61,019%.

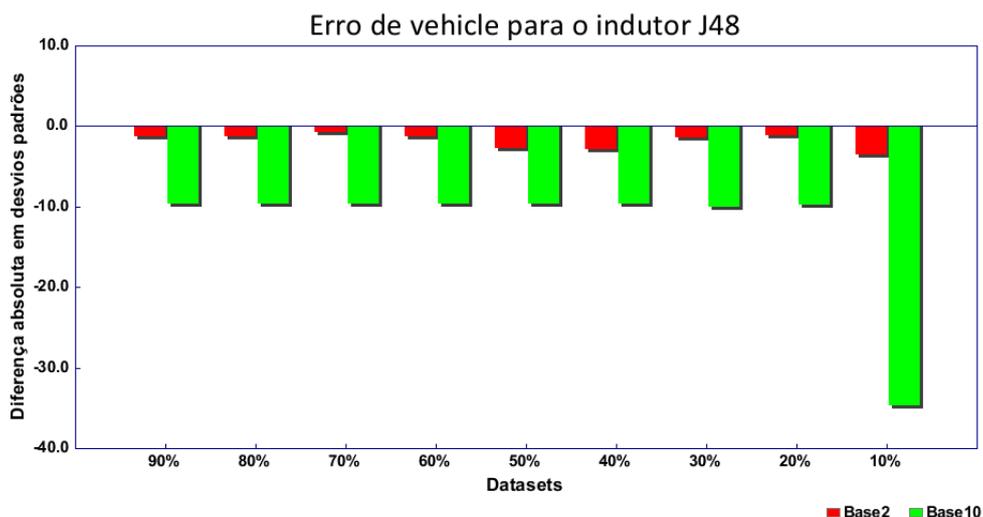


Figura 48: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *vehicle*

A taxa de erro aumentou para todos os conjuntos de forma não significativa com arredondamento utilizando a base 2, exceto para *vehicle-50%*, *vehicle-40%* e *vehicle-10%* que teve um aumento significativo. Para os conjuntos que foram arredondados utilizando a base 10, todos tiveram um aumento significativo.

Na Figura 49 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 140,200 (*vehicle*) para 168,889 (*vehicle-base2*), e para 156,156 (*vehicle-base10*). Isso significa um aumento de 20,463% para (*vehicle-base2*) e de 11,381% para (*vehicle-base10*) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para a maioria dos conjuntos, tendo um aumento significativo para *vehicle-60%*, *vehicle-50%* e *vehicle-40%*, com o arredondamento utilizando base 2. E teve uma redução significativa para *vehicle-10%* com arredondamento utilizando a base 10.

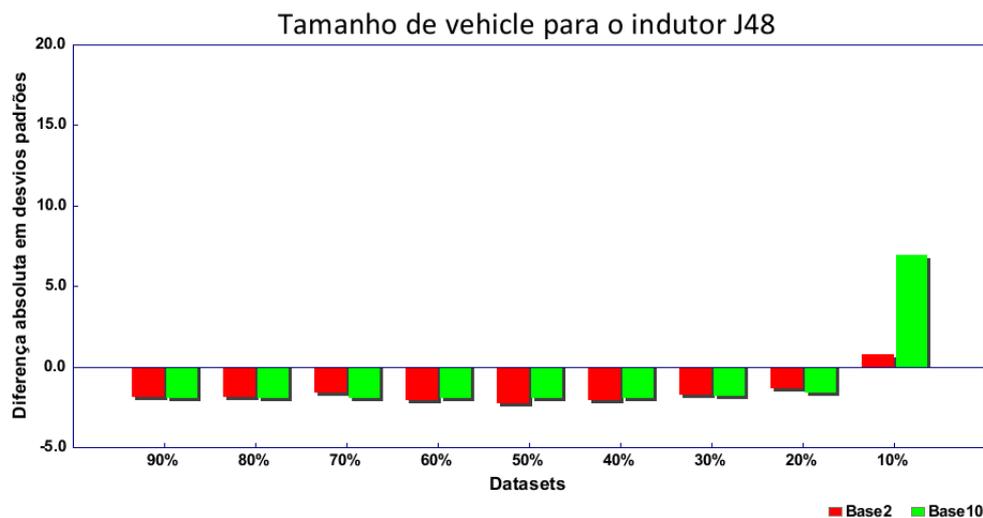


Figura 49: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) *vehicle*

### 6.2.9 Resultados vowel

Na Tabela 22 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *vowel* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Na Figura 50 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *vowel* e *vowel-90%*, entre *vowel* e *vowel-80%* e assim por diante, utilizando base 2 e base 10.

Como pode ser visto, o tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo de forma significativa, tanto para base 2 como para a base 10, exceto para o conjunto *vowel-90%*, que teve um aumento não significativo para o arredondamento

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
vowel	1,000 ± 0,042	1,000 ± 0,042	19,596 ± 0,978	19,596 ± 0,978	221,200 ± 20,422	221,200 ± 20,422
vowel-90%	0,940 ± 0,013	0,779 ± 0,015	19,394 ± 0,875	18,889 ± 0,852	224,000 ± 22,151	223,400 ± 21,046
vowel-80%	0,884 ± 0,012	0,786 ± 0,013	19,697 ± 0,894	18,889 ± 0,852	214,700 ± 22,391	223,400 ± 21,046
vowel-70%	0,878 ± 0,014	0,777 ± 0,009	19,293 ± 1,059	18,889 ± 0,852	218,200 ± 23,706	223,400 ± 21,046
vowel-60%	0,824 ± 0,025	0,781 ± 0,012	19,293 ± 0,697	18,889 ± 0,852	216,400 ± 23,234	223,400 ± 21,046
vowel-50%	0,800 ± 0,014	0,776 ± 0,018	19,596 ± 0,709	19,091 ± 0,909	218,800 ± 20,842	224,000 ± 20,412
vowel-40%	0,750 ± 0,021	0,724 ± 0,016	18,586 ± 1,253	19,091 ± 1,448	210,700 ± 17,531	224,800 ± 19,332
vowel-30%	0,709 ± 0,027	0,548 ± 0,021	19,495 ± 1,322	18,485 ± 0,917	216,100 ± 13,220	224,100 ± 28,742
vowel-20%	0,640 ± 0,020	0,545 ± 0,016	18,990 ± 1,466	18,687 ± 0,931	218,500 ± 14,136	225,600 ± 29,125
vowel-10%	0,579 ± 0,014	0,551 ± 0,021	20,505 ± 1,224	18,687 ± 0,931	217,900 ± 23,302	225,600 ± 29,125

Tabela 22: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 vowel

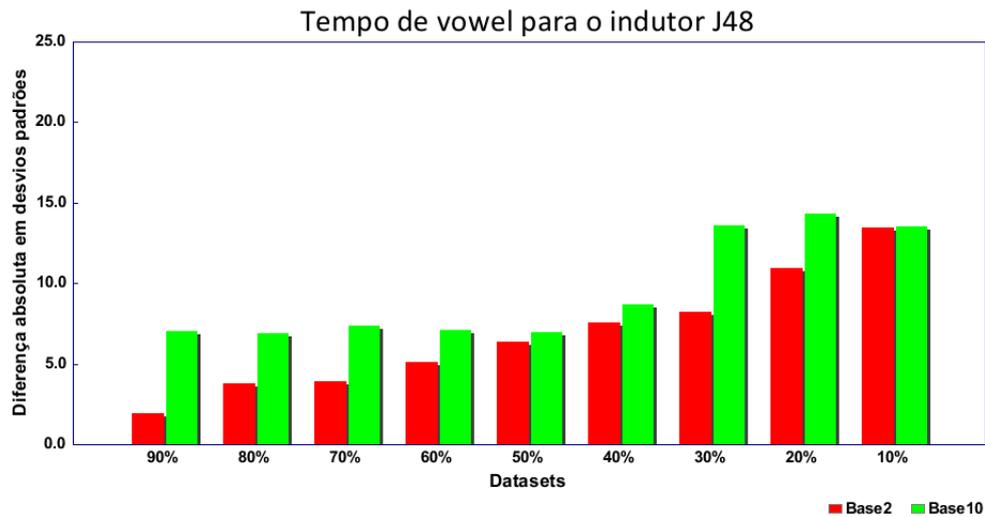


Figura 50: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

utilizando a base 2.

Analogamente, na Figura 51 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 19,596% (vehicle), para 19,428% utilizando o arredondamento com a base 2, o que representa um aumento de 0,859% da taxa de erro. E reduziu para 18,844% utilizando o arredondamento com a base 10, representando assim um aumento de 3,837%.

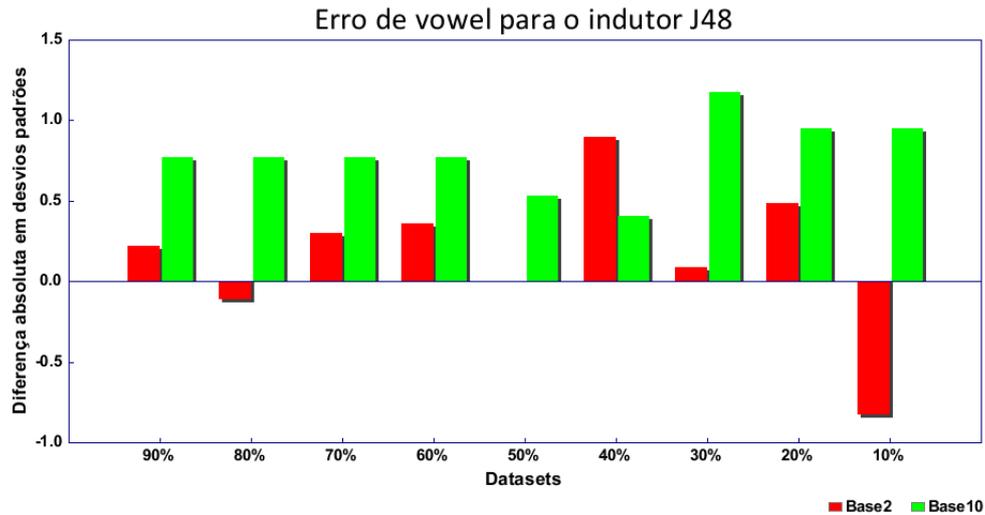


Figura 51: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

Para a maioria dos conjuntos, a taxa de erro reduziu, embora de forma não significativa.

Na Figura 52 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore reduziu de 221,200 (vowel) para 217,256 (vowel-base2), e aumentou para 224,189 (vowel-base10). Isso significa uma redução de 1,783% para (vowel-base2) e um aumento de 1,351% para (vowel-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore reduziu para a maioria dos conjuntos arredondados com a base 2, e aumentou para a maioria dos conjuntos arredondados com a base 10.

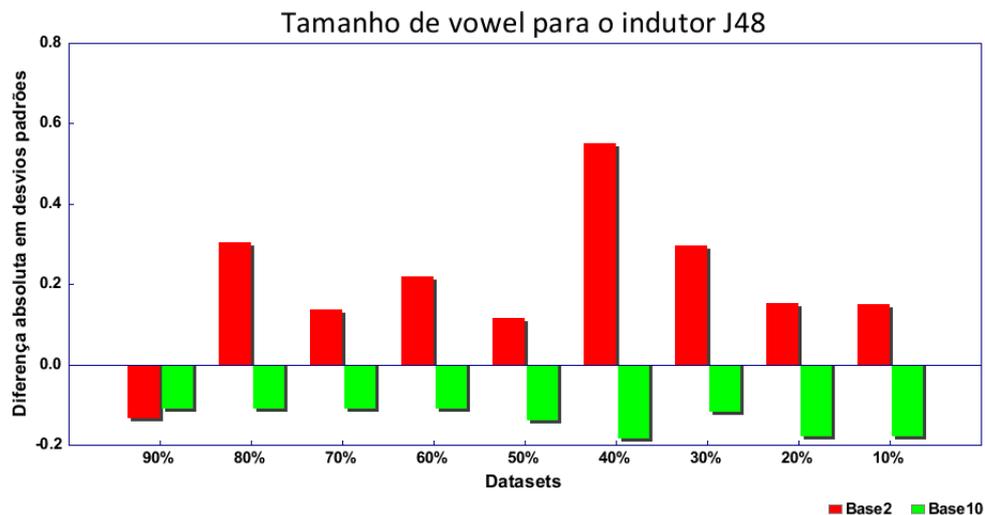


Figura 52: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) vowel

## 6.2.10 Resultados wine

Na Tabela 23 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos wine original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
wine	1,000 $\pm$ 0,052	1,000 $\pm$ 0,052	5,098 $\pm$ 2,000	5,098 $\pm$ 2,000	9,800 $\pm$ 1,398	9,800 $\pm$ 1,398
wine-90%	0,978 $\pm$ 0,047	0,978 $\pm$ 0,047	5,654 $\pm$ 2,248	5,000 $\pm$ 1,747	9,800 $\pm$ 1,398	11,400 $\pm$ 3,373
wine-80%	0,978 $\pm$ 0,047	0,956 $\pm$ 0,057	6,765 $\pm$ 2,051	5,000 $\pm$ 1,747	10,200 $\pm$ 1,398	11,400 $\pm$ 3,373
wine-70%	0,967 $\pm$ 0,054	0,967 $\pm$ 0,054	6,765 $\pm$ 2,051	6,699 $\pm$ 1,813	10,600 $\pm$ 2,066	10,600 $\pm$ 2,797
wine-60%	0,978 $\pm$ 0,047	0,956 $\pm$ 0,057	6,176 $\pm$ 1,953	11,242 $\pm$ 2,030	9,800 $\pm$ 1,398	11,200 $\pm$ 1,751
wine-50%	0,967 $\pm$ 0,075	0,967 $\pm$ 0,075	7,288 $\pm$ 1,877	10,686 $\pm$ 2,109	10,600 $\pm$ 2,066	11,000 $\pm$ 1,633
wine-40%	0,922 $\pm$ 0,075	0,956 $\pm$ 0,078	5,065 $\pm$ 1,568	10,686 $\pm$ 2,109	10,200 $\pm$ 2,150	11,000 $\pm$ 1,633
wine-30%	0,956 $\pm$ 0,057	0,956 $\pm$ 0,078	7,288 $\pm$ 2,194	12,876 $\pm$ 2,740	10,200 $\pm$ 1,398	11,200 $\pm$ 1,751
wine-20%	0,944 $\pm$ 0,059	0,878 $\pm$ 0,035	9,510 $\pm$ 2,497	11,732 $\pm$ 2,676	12,600 $\pm$ 2,271	11,000 $\pm$ 1,333
wine-10%	0,900 $\pm$ 0,035	1,056 $\pm$ 0,094	6,209 $\pm$ 2,160	28,693 $\pm$ 3,520	12,800 $\pm$ 2,898	16,200 $\pm$ 3,293

Tabela 23: Tempo de indução e taxa de erro do classificador para o indutor J48 utilizando arredondamento com bases 2 e 10 wine

Na Figura 53 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre wine e wine-90%, entre wine e wine-80% e assim por diante, utilizando base 2 e base 10.

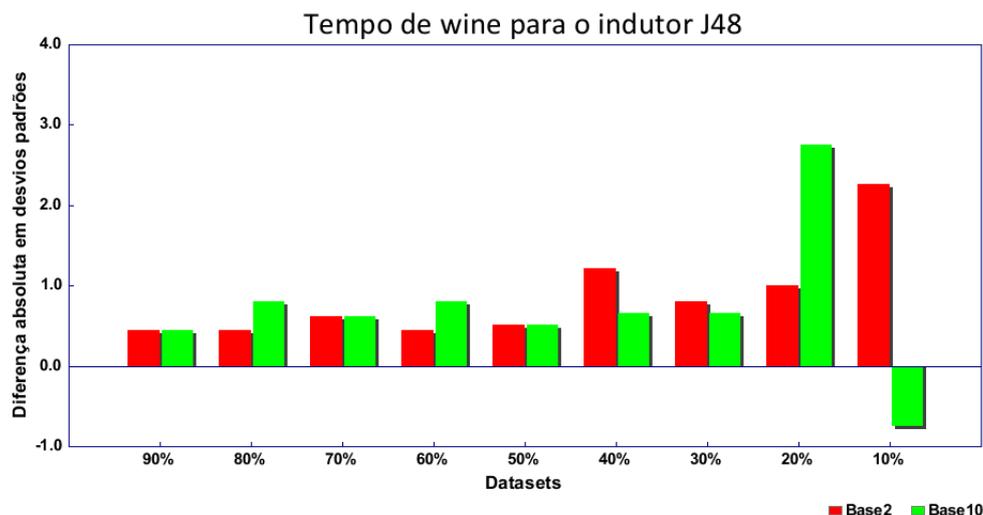


Figura 53: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

O tempo de indução reduziu para a maioria dos conjuntos utilizando arredondamento, sendo de forma não significativa, tanto para base 2 como para a base 10, exceto para os conjuntos wine-20% com arredondamento utilizando base 10, e wine-10%, com arredondamento utilizando base 2, que tiveram uma redução significativa.

Analogamente, na Figura 54 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 5,098% (wine), para 6,747% utilizando o arredondamento com a base 2, o que representa uma redução de 32,339% da taxa de erro. E aumentou para 11,402% utilizando o arredondamento com a base 10, representando assim um aumento de 123,648%.

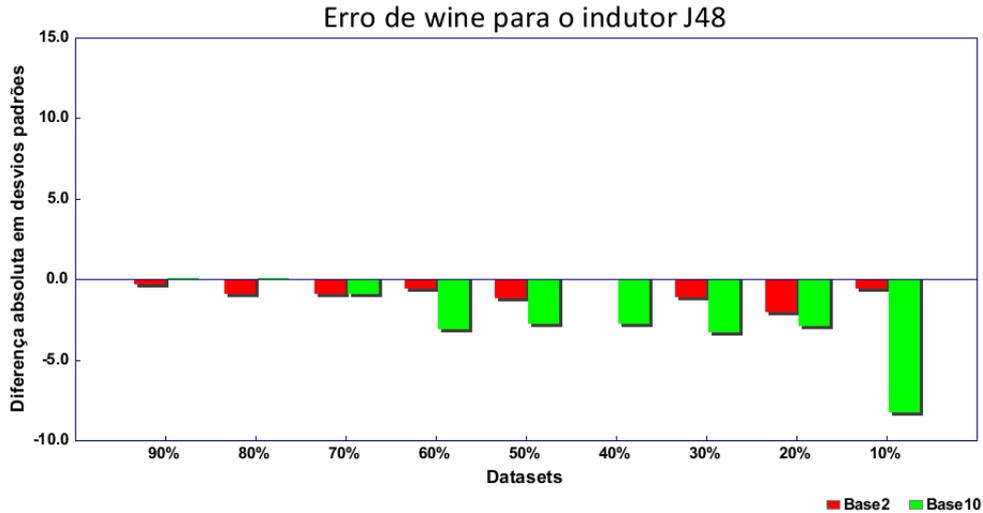


Figura 54: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

A taxa de erro aumentou para todos os conjuntos de forma não significativa para o arredondamento com base 2. Para o arredondamento com base 10, a maioria dos conjuntos tiveram um aumento significativo, exceto para os conjuntos wine-90%, wine-80% e wine-70%.

Na Figura 55 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 9,800 (wine) para 10,756 (wine-base2), e para 11,667 (wine-base10). Isso significa um aumento de 9,751% para (wine-base2) e de 19,048% para (wine-base10) do tamanho da árvore. Como pode ser notado o tamanho da árvore aumentou para todos os conjuntos, embora de forma não significativa para ambas as bases, exceto para o conjunto wine-70%, que teve um aumento significativo.

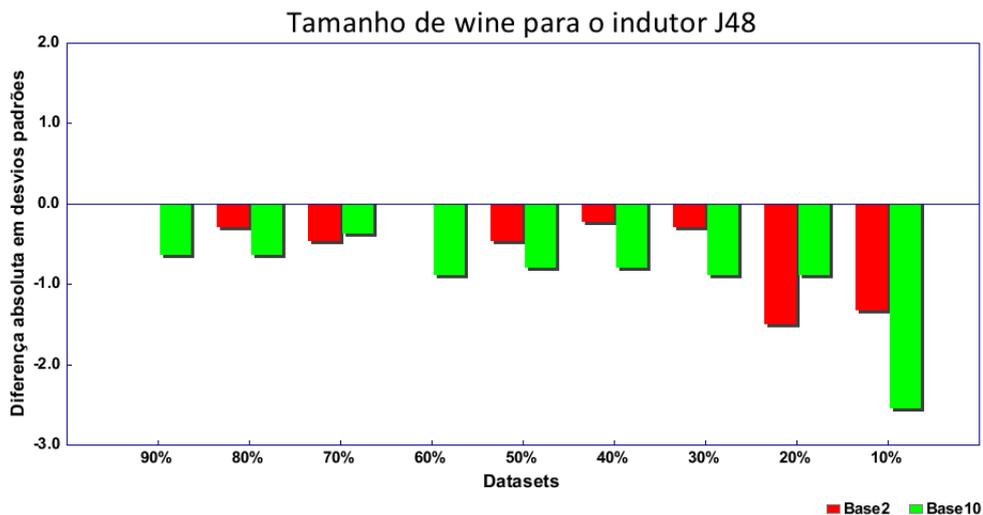


Figura 55: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) wine

### 6.2.11 Discussão

Nas Figuras 56, 57, 58, 59, 60 e 61, é mostrado um resumo dos resultados obtidos por meio do indutor J48 do tempo de indução utilizando base 2, tempo de indução utilizando base 10, taxa de erro utilizando base 2, taxa de erro utilizando base 10, tamanho do classificador utilizando base 2, tamanho do classificador utilizando base 10, respectivamente. No total são 90 conjuntos (9 conjuntos arredondados  $\times$  10 conjuntos de exemplos). No eixo  $y$  do gráfico é mostrado o número conjuntos que ficaram em cada uma das categorias mostradas no eixo  $x$  do gráfico, que são:

- aqueles que tiveram o tempo significativamente maior para os conjuntos derivados em relação ao conjunto original (barras maior que -2);
- os que tiveram o tempo maior para os conjuntos derivados (barras variando entre -2 e 0);
- os que tiveram o tempo igual para os conjuntos derivados e o conjunto original (barras igual 0);
- os que tiveram o tempo menor para os conjuntos derivados (barras variando entre 0 e 2);
- os que tiveram o tempo significativamente menor para os conjuntos derivados em relação ao conjunto original (barras maior que 2);

Analogamente para a taxa de erro e tamanho do classificador.

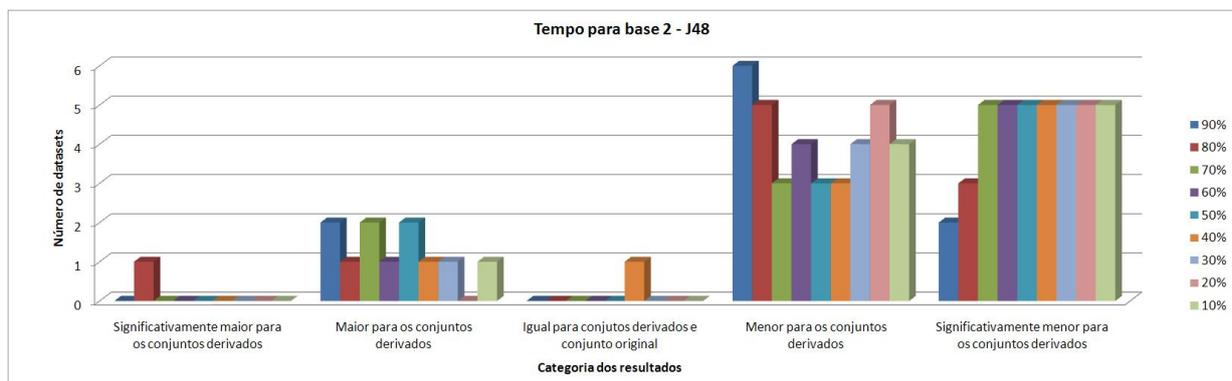


Figura 56: Resumo dos resultados do tempo de indução utilizando base 2 de J48

Como pode ser visto, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma significativa.

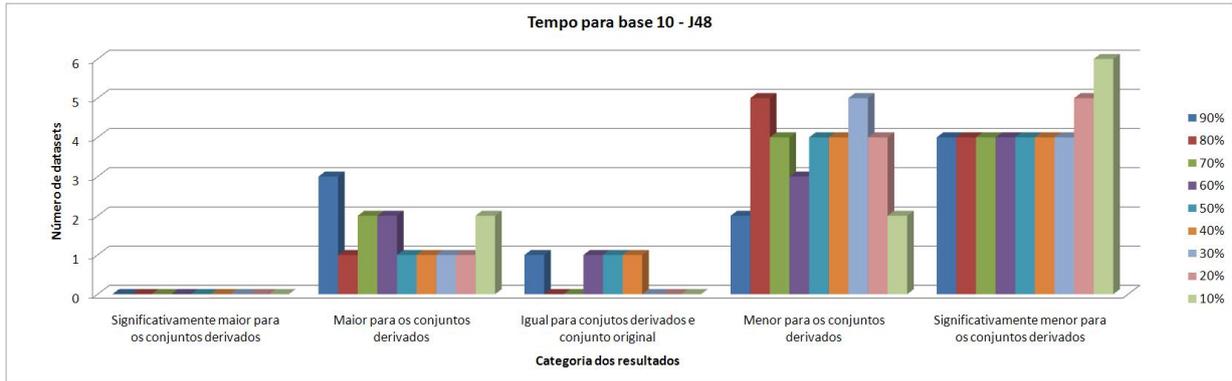


Figura 57: Resumo dos resultados do tempo de indução utilizando base 10 de J48

Como pode ser observado, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma significativa.

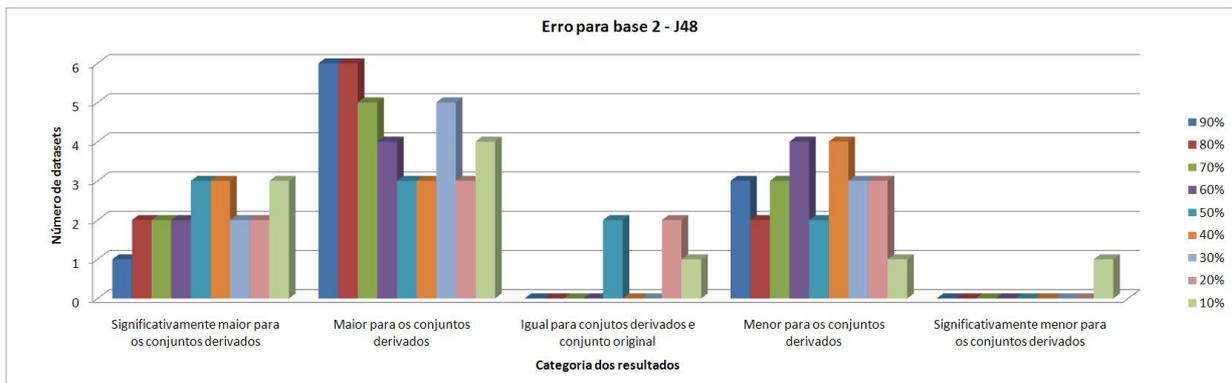


Figura 58: Resumo dos resultados da taxa de erro utilizando base 2 de J48

Como pode ser notado, houve uma tendência da taxa de erro aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa.

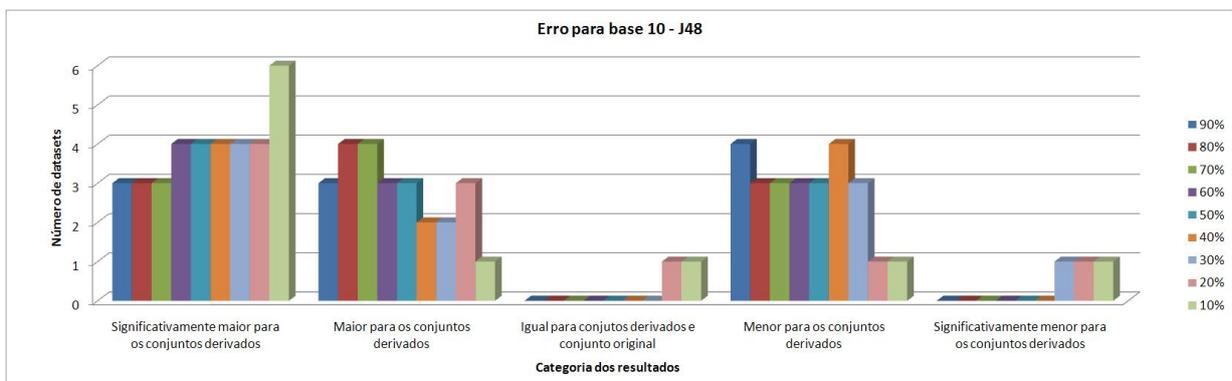


Figura 59: Resumo dos resultados da taxa de erro utilizando base 10 de J48

Como pode ser visto, houve uma tendência da taxa de erro aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma significativa.

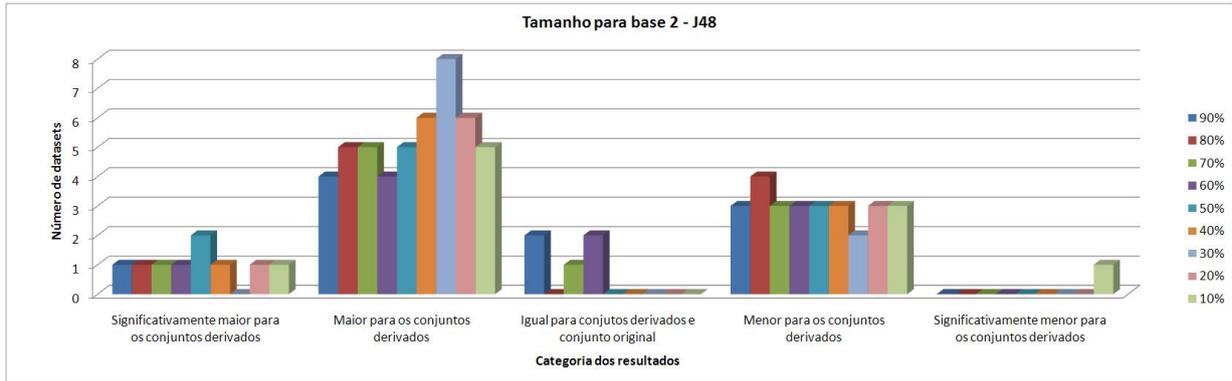


Figura 60: Resumo dos resultados do tamanho do classificador utilizando base 2 de J48

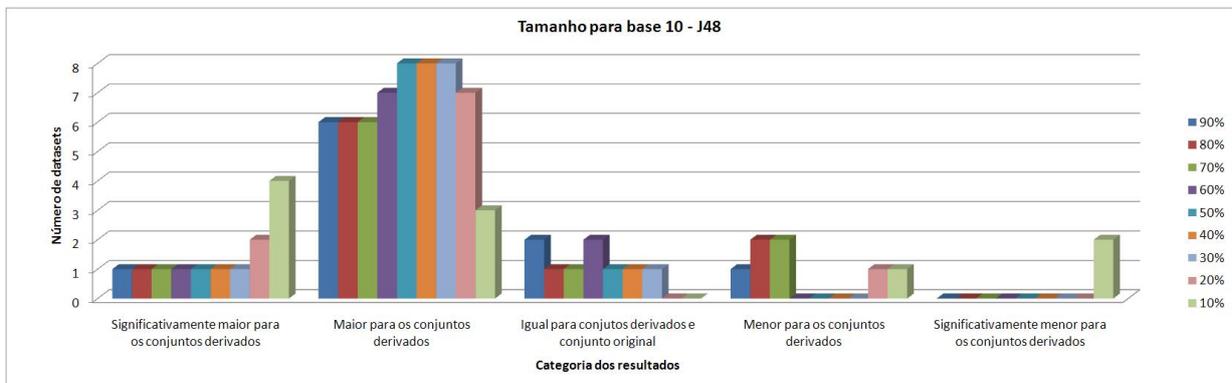


Figura 61: Resumo dos resultados do tamanho do classificador utilizando base 10 de J48

Como pode ser observado, houve uma tendência do tamanho do classificador aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa, tanto para os conjuntos arredondados com a base 2 quanto os arredondados com a base 10.

### 6.3 Resultados do indutor JRip

#### 6.3.1 Resultados aml-all-completo

Na Tabela 24 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos aml-all-completo original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
aml-all-completo	1,000 ± 0,078	1,000 ± 0,078	14,107 ± 3,693	14,107 ± 3,693	2,100 ± 0,316	2,100 ± 0,316
aml-all-completo-90%	0,965 ± 0,073	0,926 ± 0,109	12,679 ± 4,490	8,393 ± 3,128	2,200 ± 0,422	2,100 ± 0,316
aml-all-completo-80%	0,944 ± 0,111	0,855 ± 0,093	15,179 ± 3,851	14,107 ± 4,263	2,200 ± 0,422	2,100 ± 0,316
aml-all-completo-70%	0,920 ± 0,086	0,898 ± 0,082	11,250 ± 2,843	15,179 ± 3,851	2,100 ± 0,316	2,100 ± 0,316
aml-all-completo-60%	0,923 ± 0,119	0,903 ± 0,091	12,679 ± 4,490	9,643 ± 4,643	2,100 ± 0,316	2,200 ± 0,422
aml-all-completo-50%	0,898 ± 0,174	0,906 ± 0,095	18,036 ± 5,192	15,179 ± 3,851	2,200 ± 0,422	2,400 ± 0,516
aml-all-completo-40%	0,920 ± 0,089	0,779 ± 0,076	14,286 ± 5,216	16,607 ± 3,471	2,100 ± 0,316	2,400 ± 0,516
aml-all-completo-30%	0,858 ± 0,052	0,775 ± 0,094	15,714 ± 4,972	10,000 ± 4,286	2,000 ± 0,000	2,500 ± 0,527
aml-all-completo-20%	0,815 ± 0,119	0,822 ± 0,121	11,429 ± 4,151	14,107 ± 4,263	2,200 ± 0,422	2,400 ± 0,516
aml-all-completo-10%	0,837 ± 0,086	0,863 ± 0,089	22,143 ± 4,160	15,536 ± 3,347	2,400 ± 0,516	2,600 ± 0,516

Tabela 24: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 aml-all-completo

Na Figura 62 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre aml-all-completo e aml-all-completo-90%, entre aml-all-completo e aml-all-completo-80% e assim por diante, utilizando base 2 e base 10.

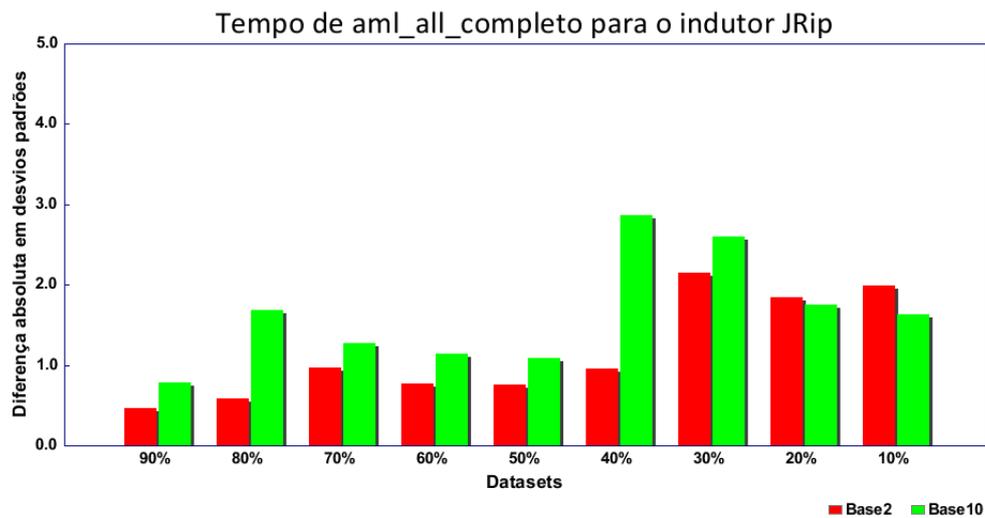


Figura 62: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) aml-all-completo

O tempo de indução aumentou para todos os conjuntos utilizando arredondamento com bases 2 e 10, sendo que um aumento significativo para aml-all-completo-30% para o arredondamento utilizando ambas as bases e aml-all-completo-40% utilizando a base 10.

Analogamente, na Figura 63 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 14,107% (aml-all-completo), para 14,822% utilizando o arredondamento com a base 2, o que representa um aumento de 5,066% da taxa de erro. E reduziu para 13,195% utilizando o arredondamento com a base 10, representando assim uma redução de 6,468%.

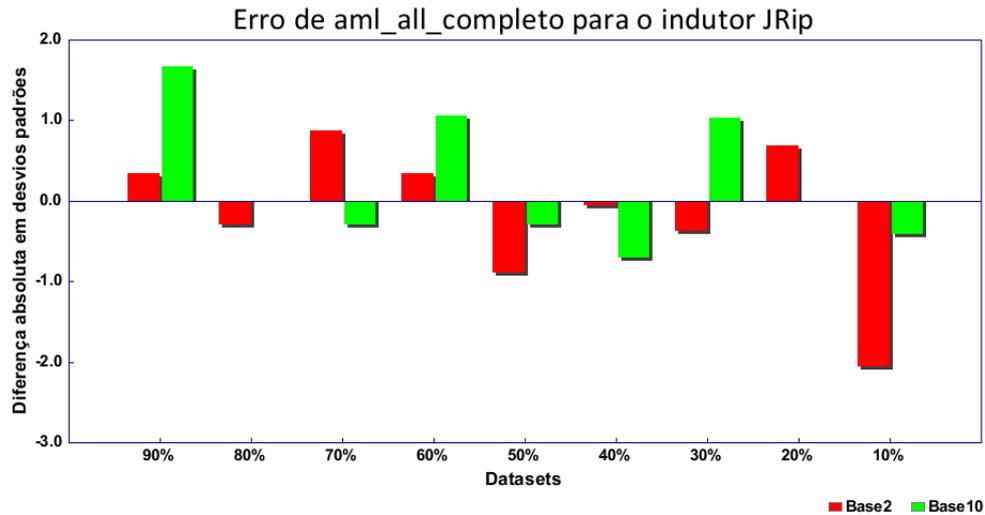


Figura 63: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) aml-all-completo

Como pode ser visto, a taxa de erro não apresentou nenhum padrão, ou seja, para alguns conjuntos a taxa de erro aumentou e para outras reduziu. O único conjunto que teve um aumento significativo foi aml-all-completo-10%, utilizando arredondamento para a base 2.

Na Figura 28 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 2,100 (aml-all-completo) para 2,167 (aml-all-completo-base2), e para 2,311 (aml-all-completo-base10). Isso significa um aumento de 3,175% para (aml-all-completo-base2) e de 10,053% para (aml-all-completo-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para a maioria dos conjuntos, embora de forma não significativa para ambas as bases.

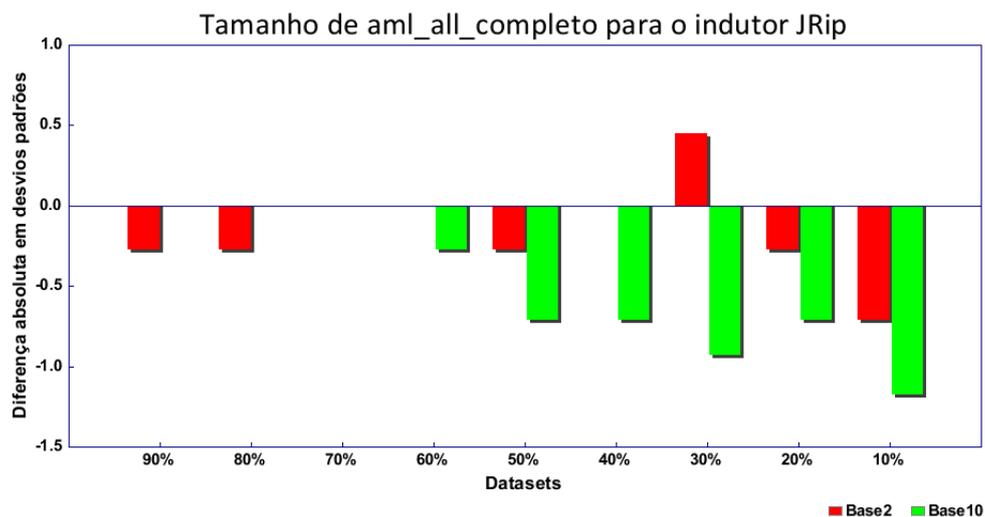


Figura 64: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) aml-all-completo

### 6.3.2 Resultados eucalyptus

Na Tabela 25 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *eucalyptus* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
<i>eucalyptus</i>	1,000 $\pm$ 0,132	1,000 $\pm$ 0,132	38,989 $\pm$ 1,618	38,989 $\pm$ 1,618	10,800 $\pm$ 2,150	10,800 $\pm$ 2,150
<i>eucalyptus-90%</i>	0,971 $\pm$ 0,227	0,960 $\pm$ 0,179	49,339 $\pm$ 2,338	55,017 $\pm$ 1,651	9,400 $\pm$ 0,966	10,900 $\pm$ 1,729
<i>eucalyptus-80%</i>	0,954 $\pm$ 0,129	0,946 $\pm$ 0,157	53,267 $\pm$ 1,377	55,017 $\pm$ 1,651	10,400 $\pm$ 1,955	10,900 $\pm$ 1,729
<i>eucalyptus-70%</i>	0,929 $\pm$ 0,163	0,971 $\pm$ 0,168	51,779 $\pm$ 1,980	55,017 $\pm$ 1,651	10,200 $\pm$ 1,932	10,900 $\pm$ 1,729
<i>eucalyptus-60%</i>	0,962 $\pm$ 0,196	1,005 $\pm$ 0,204	51,361 $\pm$ 1,709	53,384 $\pm$ 1,639	10,400 $\pm$ 1,265	11,100 $\pm$ 1,287
<i>eucalyptus-50%</i>	1,056 $\pm$ 0,144	0,949 $\pm$ 0,150	54,087 $\pm$ 1,954	53,669 $\pm$ 1,608	11,800 $\pm$ 1,317	11,800 $\pm$ 1,989
<i>eucalyptus-40%</i>	1,054 $\pm$ 0,247	0,886 $\pm$ 0,125	55,592 $\pm$ 1,551	54,487 $\pm$ 1,274	10,600 $\pm$ 0,966	11,000 $\pm$ 1,491
<i>eucalyptus-30%</i>	1,028 $\pm$ 0,131	0,988 $\pm$ 0,200	58,014 $\pm$ 1,084	53,941 $\pm$ 1,245	11,900 $\pm$ 1,792	12,600 $\pm$ 2,319
<i>eucalyptus-20%</i>	0,942 $\pm$ 0,140	0,927 $\pm$ 0,158	58,691 $\pm$ 1,161	61,288 $\pm$ 1,195	11,900 $\pm$ 2,378	11,100 $\pm$ 1,729
<i>eucalyptus-10%</i>	0,981 $\pm$ 0,112	1,120 $\pm$ 0,256	62,906 $\pm$ 1,196	58,286 $\pm$ 0,796	11,000 $\pm$ 1,491	12,300 $\pm$ 2,214

Tabela 25: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 *eucalyptus*

Na Figura 65 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *eucalyptus* e *eucalyptus-90%*, entre *eucalyptus* e *eucalyptus-80%* e assim por diante, utilizando base 2 e base 10.

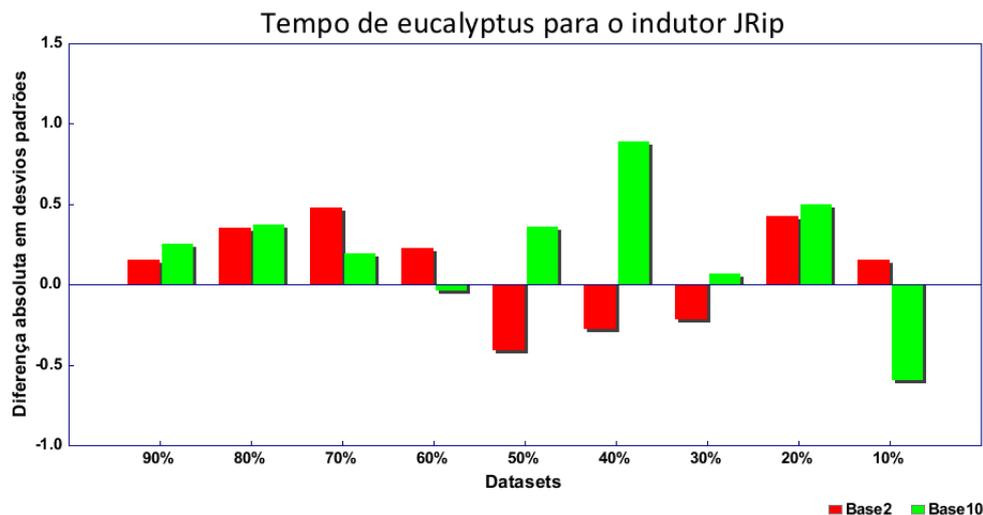


Figura 65: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *eucalyptus*

Como pode ser notado, o tempo de indução aumentou para maioria dos conjuntos de forma não significativa utilizando arredondamento com as bases 2 e 10.

Analogamente, na Figura 66 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 38,989% (*eucalyptus*), para 55,004% utilizando o arredondamento com a base 2, o que representa um aumento de 41,076% da taxa de erro. E aumentou para 55,567% utilizando o arredondamento com a base 10, representando assim um aumento de 42,521%.

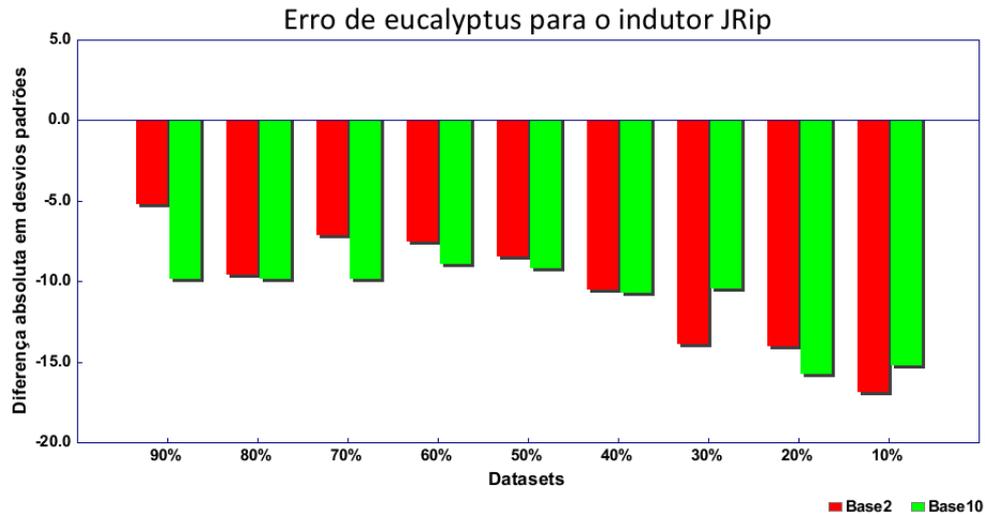


Figura 66: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) eucalyptus

A taxa de erro aumentou para todos os conjuntos de forma significativa utilizando arredondamento com ambas as bases.

Na Figura 67 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 10,800 (eucalyptus) para 10,844 (eucalyptus-base2), e 11,400 para (eucalyptus-base10). Isso significa um aumento de 0,412% para (eucalyptus-base2) e de 5,556% para (eucalyptus-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para alguns conjuntos e reduziu para outros, não mantendo um padrão, sendo de forma não significativa para ambas as bases.

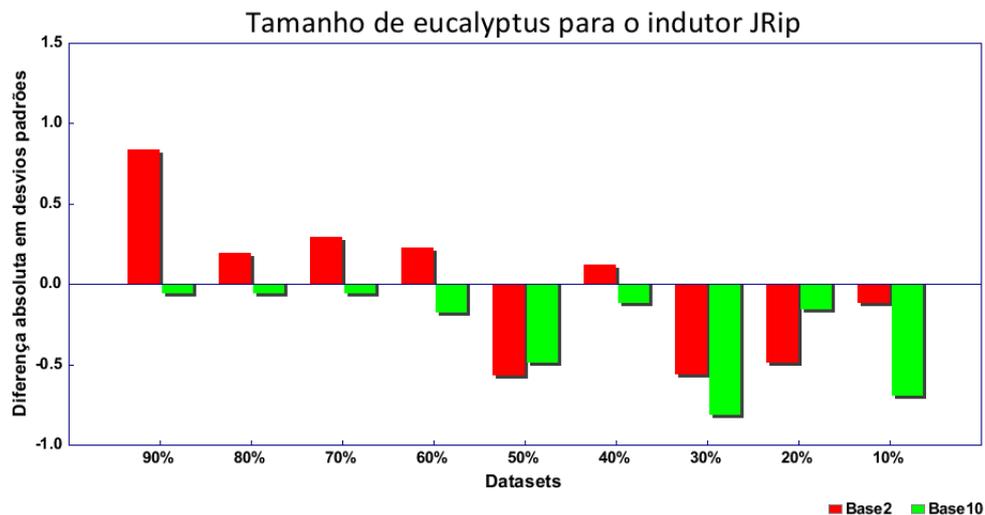


Figura 67: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) eucalyptus

### 6.3.3 Resultados hepatitis

Na Tabela 26 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos hepatitis original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
hepatitis	1,000 $\pm$ 0,081	1,000 $\pm$ 0,081	26,583 $\pm$ 4,462	26,583 $\pm$ 4,462	2,700 $\pm$ 0,675	2,700 $\pm$ 0,675
hepatitis-90%	1,218 $\pm$ 0,238	1,103 $\pm$ 0,112	23,333 $\pm$ 3,566	19,875 $\pm$ 2,466	2,600 $\pm$ 0,699	3,300 $\pm$ 1,160
hepatitis-80%	1,006 $\pm$ 0,086	1,256 $\pm$ 0,547	21,333 $\pm$ 3,537	19,875 $\pm$ 2,466	2,700 $\pm$ 0,675	3,300 $\pm$ 1,160
hepatitis-70%	1,045 $\pm$ 0,125	1,135 $\pm$ 0,290	23,208 $\pm$ 3,388	19,875 $\pm$ 2,466	2,800 $\pm$ 0,789	3,300 $\pm$ 1,160
hepatitis-60%	1,038 $\pm$ 0,108	1,013 $\pm$ 0,104	23,833 $\pm$ 3,963	19,875 $\pm$ 2,466	2,800 $\pm$ 0,919	3,300 $\pm$ 1,160
hepatitis-50%	1,109 $\pm$ 0,132	1,122 $\pm$ 0,218	17,375 $\pm$ 1,909	19,875 $\pm$ 2,466	2,500 $\pm$ 0,707	3,300 $\pm$ 1,160
hepatitis-40%	0,910 $\pm$ 0,217	1,109 $\pm$ 0,445	18,083 $\pm$ 2,519	19,875 $\pm$ 2,466	2,500 $\pm$ 0,850	3,300 $\pm$ 1,160
hepatitis-30%	0,917 $\pm$ 0,061	1,205 $\pm$ 0,443	22,458 $\pm$ 2,653	17,375 $\pm$ 2,690	2,700 $\pm$ 0,483	3,300 $\pm$ 0,675
hepatitis-20%	0,994 $\pm$ 0,248	0,923 $\pm$ 0,092	19,917 $\pm$ 1,926	20,000 $\pm$ 2,220	2,800 $\pm$ 0,919	3,100 $\pm$ 0,876
hepatitis-10%	0,859 $\pm$ 0,101	0,942 $\pm$ 0,157	15,458 $\pm$ 1,957	17,375 $\pm$ 1,878	2,700 $\pm$ 0,675	2,900 $\pm$ 0,316

Tabela 26: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 hepatitis

Na Figura 68 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre hepatitis e hepatitis-90%, entre hepatitis e hepatitis-80% e assim por diante, utilizando base 2 e base 10.

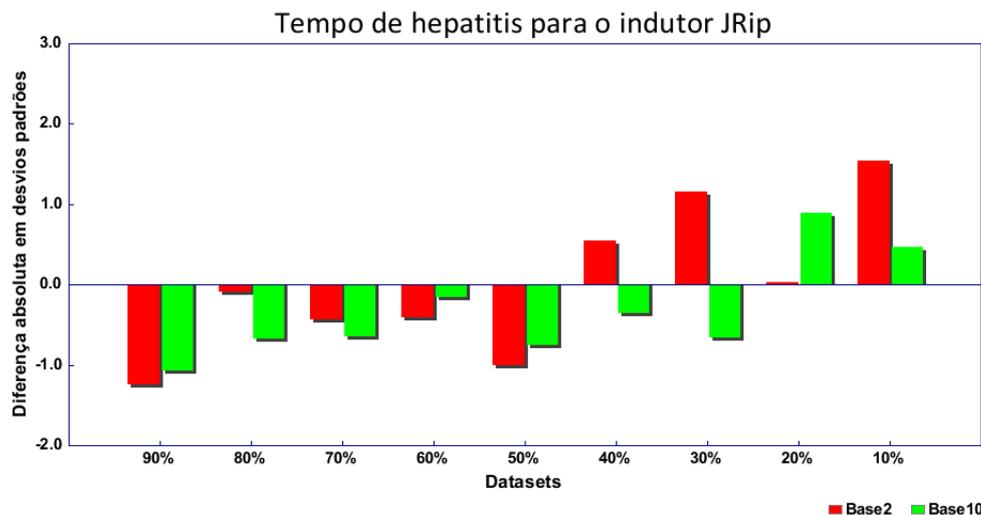


Figura 68: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

O tempo de indução aumentou para a maioria dos conjuntos, sendo de forma não significativa, tanto para base 2 como para a base 10, porém para os conjuntos que tiveram seus valores arredondados para uma menor porcentagem de distintos, alguns deles tiveram uma redução no tempo de indução, também de forma não significativa.

Analogamente, na Figura 69 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 26,583% (hepatitis), para 20,555% utilizando o arredondamento com a base 2, o que representa uma redução de 22,675% da taxa de erro. E reduziu para 19,333% utilizando o arredondamento com a base 10, representando assim uma redução de 27,272%.

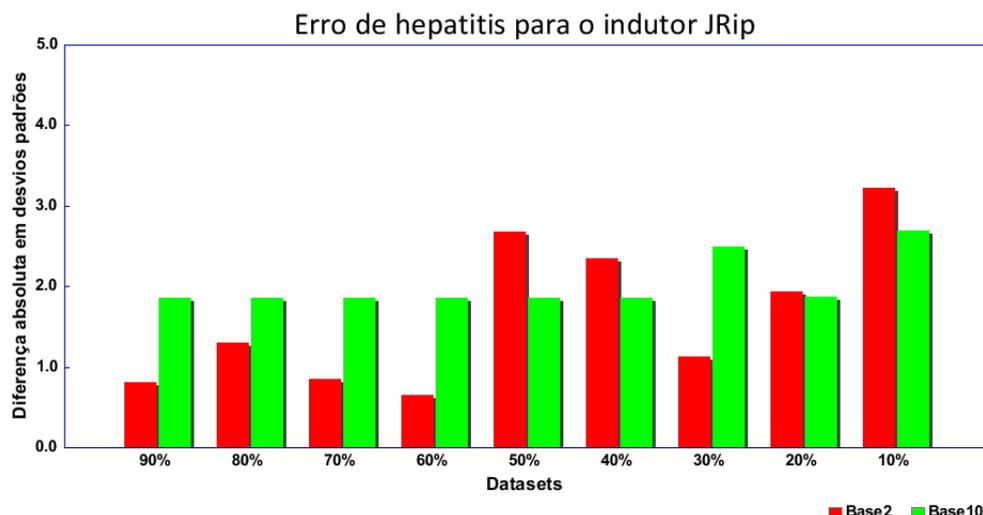


Figura 69: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

Como pode ser visto, a taxa de erro reduziu para os conjuntos de forma não significativa, utilizando arredondamento com a bases 2 e 10, exceto para hepatitis-50%, hepatitis-40% e hepatitis-10% que tiveram uma menor taxa de erro com a base 2, e para hepatitis-30% e hepatitis-10% que também tiveram uma menor taxa de erro com a base 10, ambos de forma não significativa.

Na Figura 70 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore diminuiu de 2,700 (hepatitis) para 2,678 (hepatitis-base2), e aumentou para 3,233 (hepatitis-base10). Isso significa uma redução de 0,823% para (hepatitis-base2) e de 19,753% para (hepatitis-base10) do tamanho da árvore. Como pode ser notado, no geral, o tamanho da árvore se manteve próximo ao do conjunto original para a base 2, e para a base 10 o tamanho aumentou, embora de forma não significativa para ambas as bases.

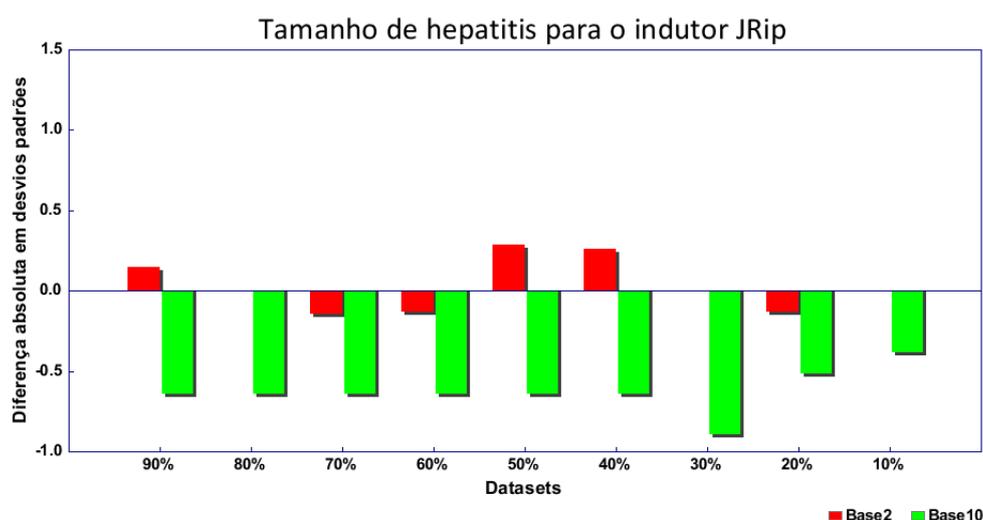


Figura 70: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) hepatitis

### 6.3.4 Resultados ionosphere

Na Tabela 27 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *ionosphere* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
<i>ionosphere</i>	1,000 $\pm$ 0,159	1,000 $\pm$ 0,159	9,698 $\pm$ 1,492	9,698 $\pm$ 1,492	5,400 $\pm$ 1,897	5,400 $\pm$ 1,897
<i>ionosphere-90%</i>	0,975 $\pm$ 0,214	0,891 $\pm$ 0,150	11,127 $\pm$ 1,384	9,405 $\pm$ 1,050	5,400 $\pm$ 1,506	6,200 $\pm$ 1,687
<i>ionosphere-80%</i>	0,921 $\pm$ 0,134	0,903 $\pm$ 0,152	11,127 $\pm$ 1,625	9,405 $\pm$ 1,050	5,400 $\pm$ 1,838	6,200 $\pm$ 1,687
<i>ionosphere-70%</i>	0,928 $\pm$ 0,145	0,922 $\pm$ 0,166	11,984 $\pm$ 1,531	9,405 $\pm$ 1,050	5,800 $\pm$ 1,814	6,200 $\pm$ 1,687
<i>ionosphere-60%</i>	0,890 $\pm$ 0,139	0,903 $\pm$ 0,154	11,976 $\pm$ 1,900	9,405 $\pm$ 1,050	5,300 $\pm$ 2,058	6,200 $\pm$ 1,687
<i>ionosphere-50%</i>	1,068 $\pm$ 0,133	0,994 $\pm$ 0,124	12,524 $\pm$ 1,804	11,127 $\pm$ 1,246	5,900 $\pm$ 1,370	6,000 $\pm$ 1,700
<i>ionosphere-40%</i>	0,994 $\pm$ 0,149	0,986 $\pm$ 0,080	16,238 $\pm$ 1,650	9,976 $\pm$ 1,150	5,000 $\pm$ 1,826	7,400 $\pm$ 0,966
<i>ionosphere-30%</i>	1,023 $\pm$ 0,215	0,983 $\pm$ 0,127	12,802 $\pm$ 1,523	9,690 $\pm$ 1,716	7,000 $\pm$ 2,055	7,700 $\pm$ 0,949
<i>ionosphere-20%</i>	0,981 $\pm$ 0,138	0,994 $\pm$ 0,109	12,548 $\pm$ 1,368	9,690 $\pm$ 1,716	7,100 $\pm$ 0,994	7,700 $\pm$ 0,949
<i>ionosphere-10%</i>	0,975 $\pm$ 0,098	1,021 $\pm$ 0,092	10,548 $\pm$ 1,957	9,690 $\pm$ 1,716	7,700 $\pm$ 1,059	7,700 $\pm$ 0,949

Tabela 27: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 *ionosphere*

Na Figura 71 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *ionosphere* e *ionosphere-90%*, entre *ionosphere* e *ionosphere-80%* e assim por diante, utilizando base 2 e base 10.

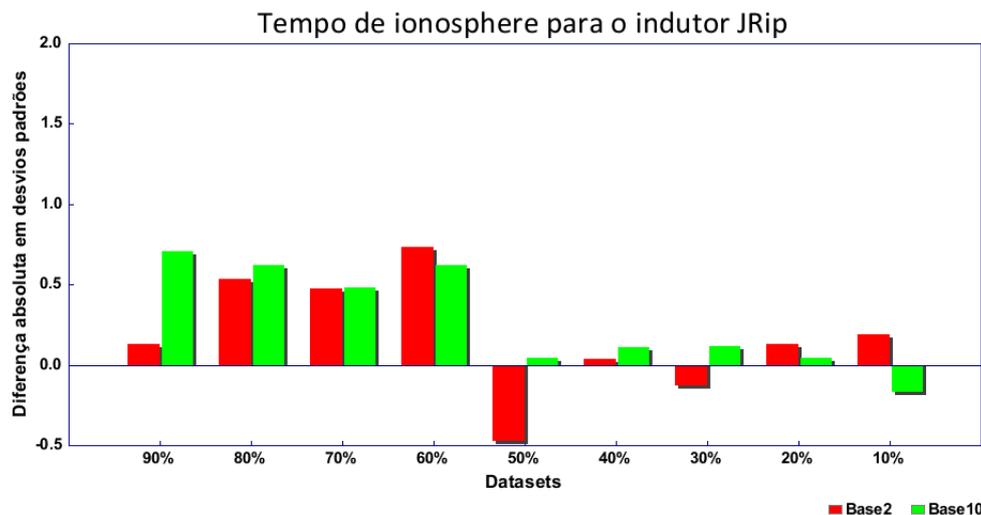


Figura 71: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *ionosphere*

O tempo de indução reduziu para a maioria dos conjuntos utilizando arredondamento, de forma significativa, tanto para base 2 como para a base 10.

Analogamente, na Figura 72 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 9,698% (*ionosphere*), para 12,319% utilizando o arredondamento com a base 2, o que representa um aumento de 27,030% da taxa de erro. E aumentou para 9,755% utilizando o arredondamento com a base 10, representando assim um aumento de 0,585%.

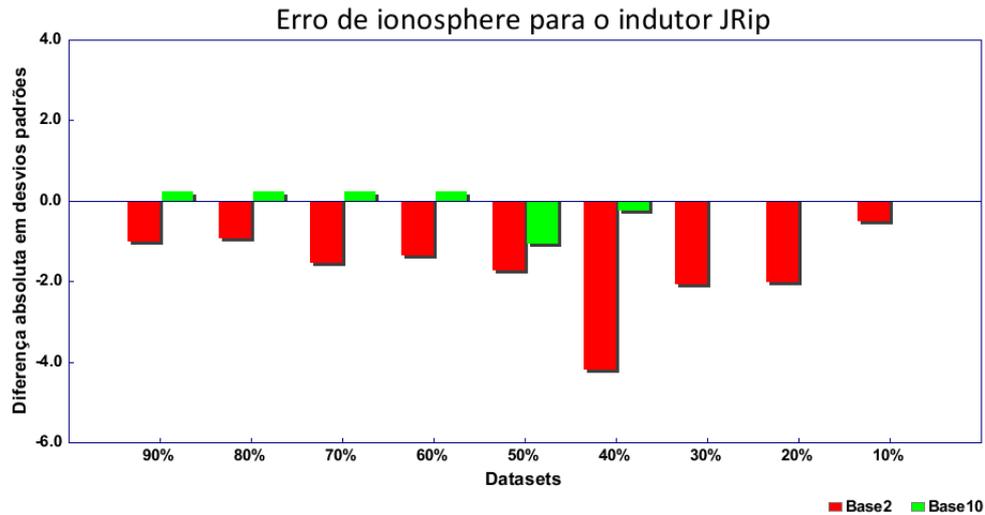


Figura 72: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) ionosphere

Como pode ser observado, todos os conjuntos que realizou-se o arredondamento com a base 2, a taxa de erro aumentou, sendo de forma significativa apenas para ionosphere-40% e ionosphere-30%. Já para a base 10, os conjuntos não mantiveram um padrão, apresentando redução para alguns e aumento para outros, sendo uma redução não significativa.

Na Figura 73 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 5,400 (ionosphere) para 6,067 (ionosphere-base2), e para 6,811 (ionosphere-base10). Isso significa um aumento de 12,346% para (ionosphere-base2) e de 26,132% para (ionosphere-base10) do tamanho da árvore. Como pode ser notado o tamanho da árvore aumentou para a maioria dos conjuntos, embora de forma não significativa para ambas as bases.

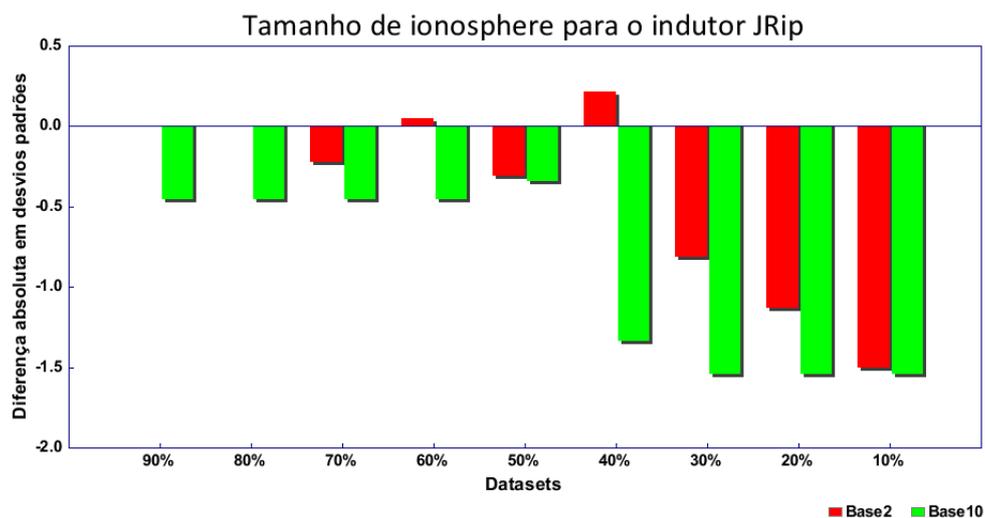


Figura 73: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) ionosphere

### 6.3.5 Resultados lymph

Na Tabela 28 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos lymph original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
lymph	1,000 $\pm$ 0,089	1,000 $\pm$ 0,089	23,667 $\pm$ 2,870	23,667 $\pm$ 2,870	6,100 $\pm$ 0,568	6,100 $\pm$ 0,568
lymph-90%	1,088 $\pm$ 0,098	1,088 $\pm$ 0,093	24,952 $\pm$ 2,421	22,143 $\pm$ 2,933	6,300 $\pm$ 1,160	6,900 $\pm$ 1,370
lymph-80%	1,054 $\pm$ 0,091	1,101 $\pm$ 0,101	24,952 $\pm$ 2,421	22,143 $\pm$ 2,933	6,300 $\pm$ 1,160	6,900 $\pm$ 1,370
lymph-70%	1,054 $\pm$ 0,091	1,074 $\pm$ 0,081	24,952 $\pm$ 2,421	22,143 $\pm$ 2,933	6,300 $\pm$ 1,160	6,900 $\pm$ 1,370
lymph-60%	1,081 $\pm$ 0,096	1,081 $\pm$ 0,096	22,952 $\pm$ 2,750	22,143 $\pm$ 2,933	6,800 $\pm$ 1,229	6,900 $\pm$ 1,370
lymph-50%	1,074 $\pm$ 0,098	1,088 $\pm$ 0,093	22,952 $\pm$ 2,750	22,143 $\pm$ 2,933	6,800 $\pm$ 1,229	6,900 $\pm$ 1,370
lymph-40%	1,068 $\pm$ 0,118	1,122 $\pm$ 0,116	20,857 $\pm$ 2,494	22,143 $\pm$ 2,933	6,700 $\pm$ 0,949	6,900 $\pm$ 1,370
lymph-30%	1,074 $\pm$ 0,103	1,027 $\pm$ 0,105	20,190 $\pm$ 1,920	22,810 $\pm$ 3,117	6,600 $\pm$ 1,265	6,400 $\pm$ 1,174
lymph-20%	1,142 $\pm$ 0,192	1,020 $\pm$ 0,112	20,810 $\pm$ 2,658	20,952 $\pm$ 2,523	6,400 $\pm$ 1,430	6,700 $\pm$ 0,949
lymph-10%	1,014 $\pm$ 0,169	0,939 $\pm$ 0,112	22,952 $\pm$ 3,017	22,952 $\pm$ 3,017	6,400 $\pm$ 0,843	6,400 $\pm$ 0,843

Tabela 28: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 lymph

Na Figura 74 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre lymph e lymph-90%, entre lymph e lymph-80% e assim por diante, utilizando base 2 e base 10.

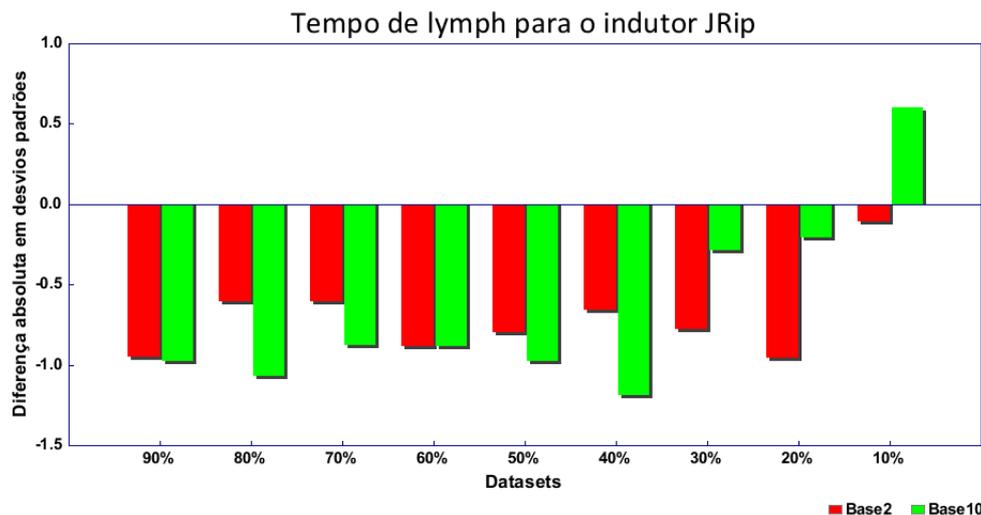


Figura 74: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

No geral, o tempo de indução aumentou para a maioria dos conjuntos utilizando arredondamento com ambas as bases, sendo de forma não significativa.

Analogamente, na Figura 75 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 23,667% (lymph), para 22,841% utilizando o arredondamento com a base 2, o que representa uma redução de 3,490% da taxa de erro. E aumentou para 22,175% utilizando o arredondamento com a base 10, representando assim uma redução de 6,306%.

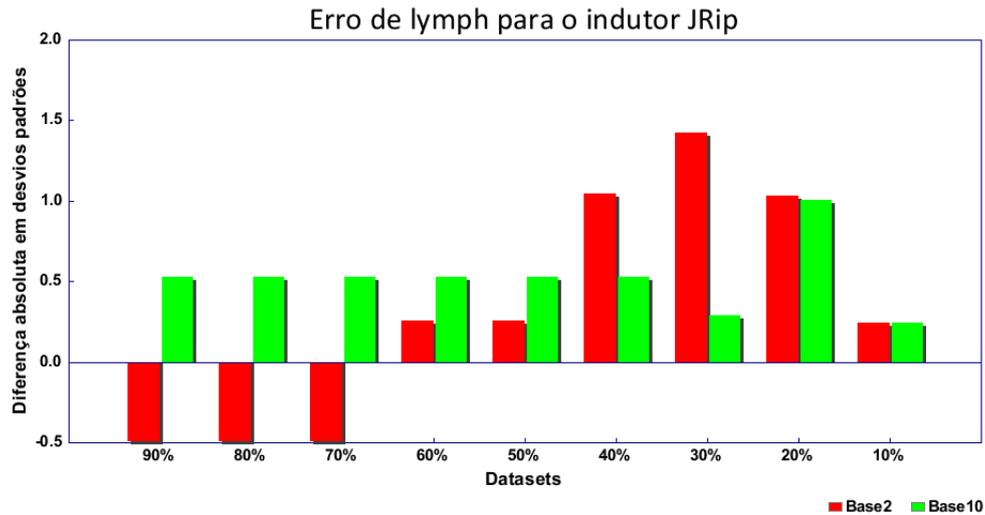


Figura 75: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

Como pode ser visto, a taxa de erro aumentou para a maioria dos conjuntos de forma não significativa, utilizando arredondamento com as bases 2 e 10.

Na Figura 76 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 6,100 (lymph) para 6,511 (lymph-base2), e para 6,767 (lymph-base10). Isso significa um aumento de 6,740% para (lymph-base2) e de 10,929% para (lymph-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para todos os conjuntos, de forma não significativa para ambas as bases.

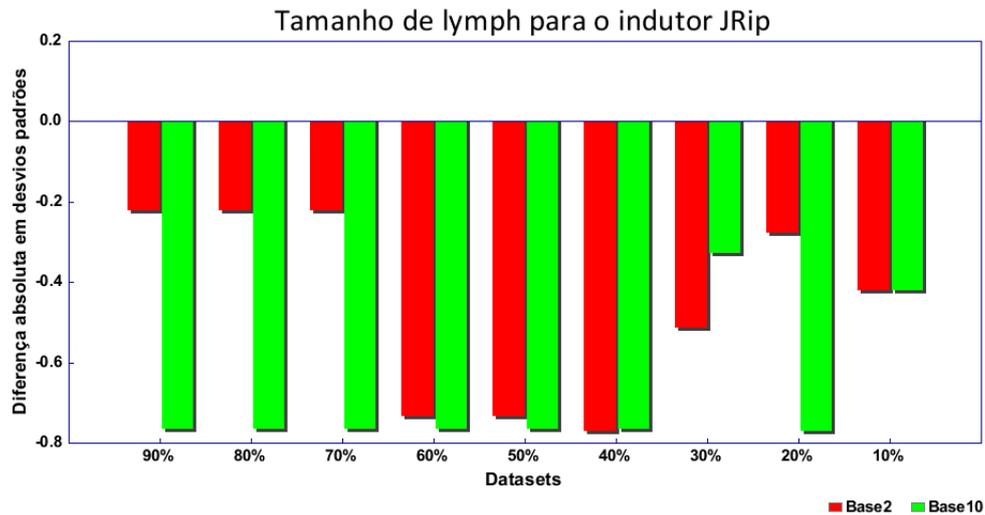


Figura 76: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) lymph

### 6.3.6 Resultados sick

Na Tabela 29 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sick original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
sick	1,000 $\pm$ 0,299	1,000 $\pm$ 0,299	1,670 $\pm$ 0,194	1,670 $\pm$ 0,194	7,700 $\pm$ 1,889	7,700 $\pm$ 1,889
sick-90%	0,996 $\pm$ 0,219	1,045 $\pm$ 0,433	1,405 $\pm$ 0,153	4,560 $\pm$ 0,184	7,500 $\pm$ 1,080	5,500 $\pm$ 1,958
sick-80%	0,970 $\pm$ 0,476	1,051 $\pm$ 0,436	2,784 $\pm$ 0,210	4,560 $\pm$ 0,184	6,700 $\pm$ 2,452	5,500 $\pm$ 1,958
sick-70%	0,804 $\pm$ 0,331	1,045 $\pm$ 0,429	1,670 $\pm$ 0,476	3,527 $\pm$ 0,509	6,300 $\pm$ 1,767	5,500 $\pm$ 1,958
sick-60%	0,927 $\pm$ 0,270	0,852 $\pm$ 0,309	2,545 $\pm$ 0,484	4,374 $\pm$ 0,253	6,500 $\pm$ 1,581	5,200 $\pm$ 1,619
sick-50%	1,075 $\pm$ 0,632	0,847 $\pm$ 0,310	2,624 $\pm$ 0,480	4,003 $\pm$ 0,304	6,900 $\pm$ 1,853	5,200 $\pm$ 1,619
sick-40%	1,032 $\pm$ 0,340	0,857 $\pm$ 0,312	3,261 $\pm$ 0,358	4,401 $\pm$ 0,231	6,100 $\pm$ 0,994	5,200 $\pm$ 1,619
sick-30%	1,064 $\pm$ 0,290	1,108 $\pm$ 0,486	1,882 $\pm$ 0,598	3,976 $\pm$ 0,533	6,100 $\pm$ 1,197	5,600 $\pm$ 1,430
sick-20%	0,684 $\pm$ 0,218	1,102 $\pm$ 0,493	2,677 $\pm$ 0,373	4,497 $\pm$ 0,261	5,400 $\pm$ 1,174	5,600 $\pm$ 1,430
sick-10%	0,543 $\pm$ 0,165	0,312 $\pm$ 0,075	6,018 $\pm$ 0,153	6,151 $\pm$ 0,035	4,000 $\pm$ 0,943	1,700 $\pm$ 0,483

Tabela 29: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 sick

Na Figura 77 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre sick e sick-90%, entre sick e sick-80% e assim por diante, utilizando base 2 e base 10.

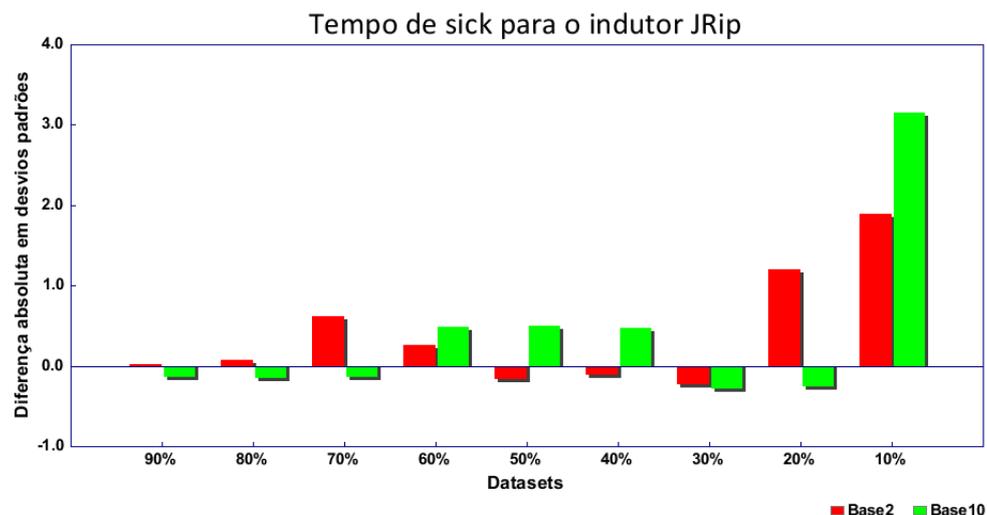


Figura 77: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

Como pode ser notado, o tempo de indução não manteve um padrão para os conjuntos utilizando arredondamento com ambas as bases, apresentando aumentos e reduções não significativos, exceto para sick-10% que teve uma redução significativa do tempo de indução, para o arredondamento utilizando a base 10.

Analogamente, na Figura 78 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 1,670% (sick), para 2,763% utilizando o arredondamento com a base 2, o que representa um aumento de 65,442% da taxa de erro. E aumentou para 4,450% utilizando o arredondamento com a base 10, representando assim um aumento de 166,460%.

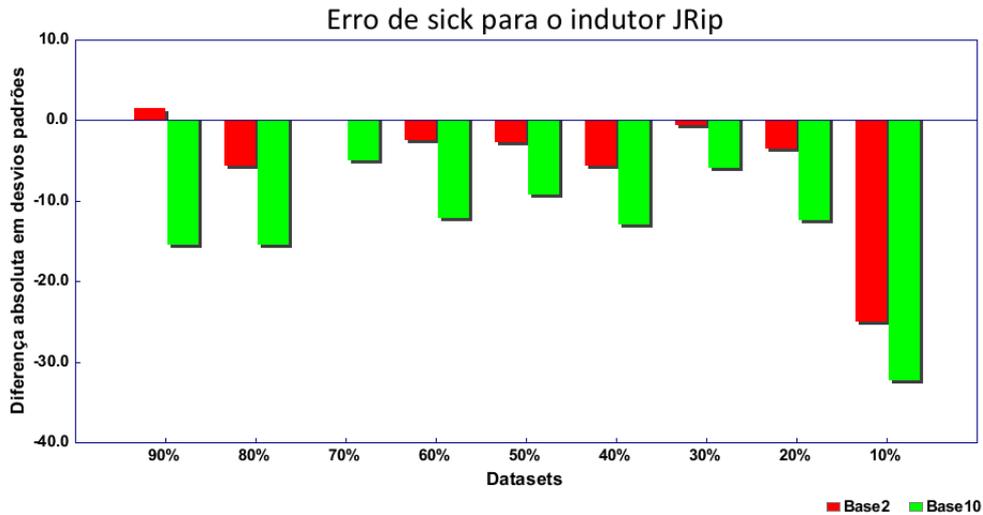


Figura 78: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

A taxa de erro aumentou de forma significativa para a maioria dos conjuntos utilizando arredondamento com ambas as bases.

Na Figura 79 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore reduziu de 7,700 (sick) para 6,167 (sick-base2), e para 5,000 (sick-base10). Isso significa uma redução de 19,913% para (sick-base2) e um aumento de 35,065% para (sick-base10) do tamanho da árvore. O tamanho da árvore reduziu para a todos os conjuntos, para ambas as bases, sendo de forma significativa apenas para sick-10%.

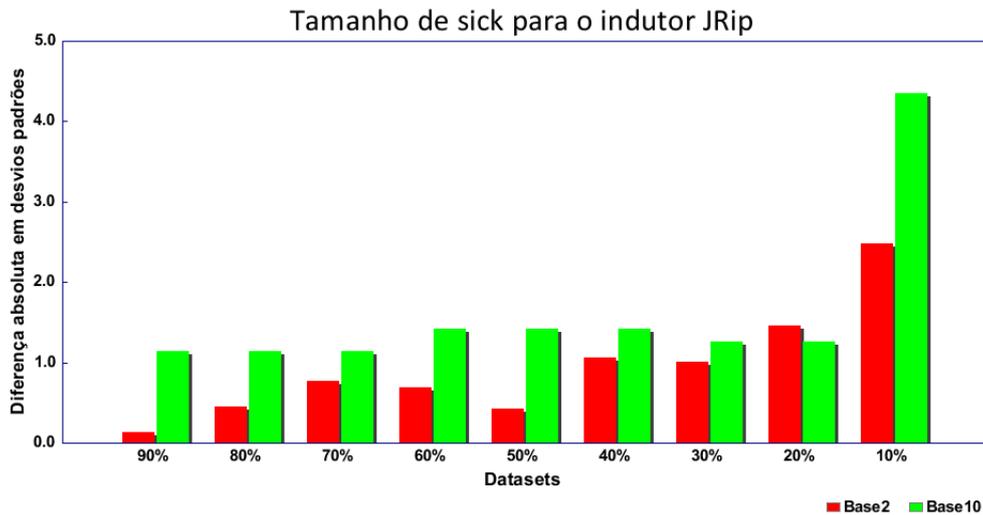


Figura 79: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) sick

### 6.3.7 Resultados sonar

Na Tabela 30 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sonar original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
sonar	1,000 $\pm$ 0,085	1,000 $\pm$ 0,085	26,905 $\pm$ 2,360	26,905 $\pm$ 2,360	4,600 $\pm$ 0,843	4,600 $\pm$ 0,843
sonar-90%	1,106 $\pm$ 0,117	0,958 $\pm$ 0,121	10,143 $\pm$ 4,057	12,048 $\pm$ 4,509	5,200 $\pm$ 1,229	5,500 $\pm$ 0,850
sonar-80%	0,949 $\pm$ 0,156	0,958 $\pm$ 0,110	15,833 $\pm$ 3,921	20,015 $\pm$ 3,483	4,700 $\pm$ 1,252	5,500 $\pm$ 0,850
sonar-70%	1,212 $\pm$ 0,222	0,978 $\pm$ 0,119	14,015 $\pm$ 3,512	21,817 $\pm$ 2,693	5,000 $\pm$ 1,155	5,500 $\pm$ 0,850
sonar-60%	1,029 $\pm$ 0,051	1,091 $\pm$ 0,188	13,000 $\pm$ 4,713	11,571 $\pm$ 3,973	4,900 $\pm$ 0,738	5,500 $\pm$ 0,850
sonar-50%	1,146 $\pm$ 0,216	0,996 $\pm$ 0,143	18,429 $\pm$ 3,832	19,576 $\pm$ 3,208	5,300 $\pm$ 1,252	5,500 $\pm$ 0,850
sonar-40%	1,015 $\pm$ 0,145	0,832 $\pm$ 0,149	26,381 $\pm$ 2,811	23,048 $\pm$ 2,236	4,700 $\pm$ 1,160	4,100 $\pm$ 0,994
sonar-30%	0,949 $\pm$ 0,055	0,837 $\pm$ 0,103	13,952 $\pm$ 4,731	20,690 $\pm$ 4,103	4,900 $\pm$ 0,738	5,200 $\pm$ 1,476
sonar-20%	0,949 $\pm$ 0,140	1,009 $\pm$ 0,254	17,452 $\pm$ 4,908	21,396 $\pm$ 3,669	5,500 $\pm$ 1,080	4,700 $\pm$ 0,949
sonar-10%	0,872 $\pm$ 0,090	0,843 $\pm$ 0,056	24,524 $\pm$ 2,063	31,595 $\pm$ 3,781	5,600 $\pm$ 1,955	6,100 $\pm$ 0,876

Tabela 30: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 sonar

Na Figura 80 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre sonar e sonar-90%, entre sonar e sonar-80% e assim por diante, utilizando base 2 e base 10.

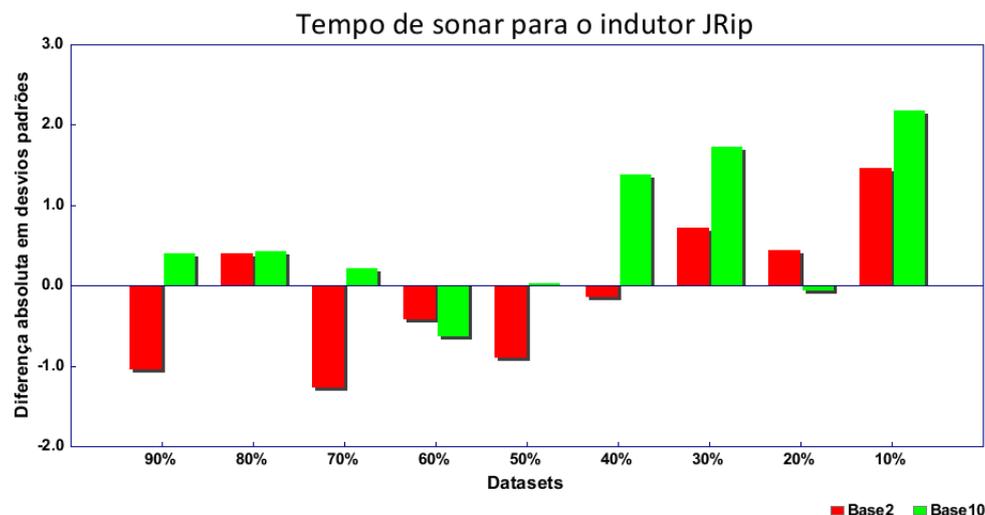


Figura 80: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

O tempo de indução não manteve um padrão para os conjuntos utilizando arredondamento com ambas as bases, apresentando aumentos e reduções não significativos, exceto para sick-10% que teve uma redução significativa do tempo de indução, para o arredondamento utilizando a base 10.

Analogamente, na Figura 81 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 26,905% (sonar), para 17,081% utilizando o arredondamento com a base 2, o que representa uma redução de 36,514% da taxa de erro. E reduziu para 20,195% utilizando o arredondamento com a base 10, representando assim uma redução de 24,939%.

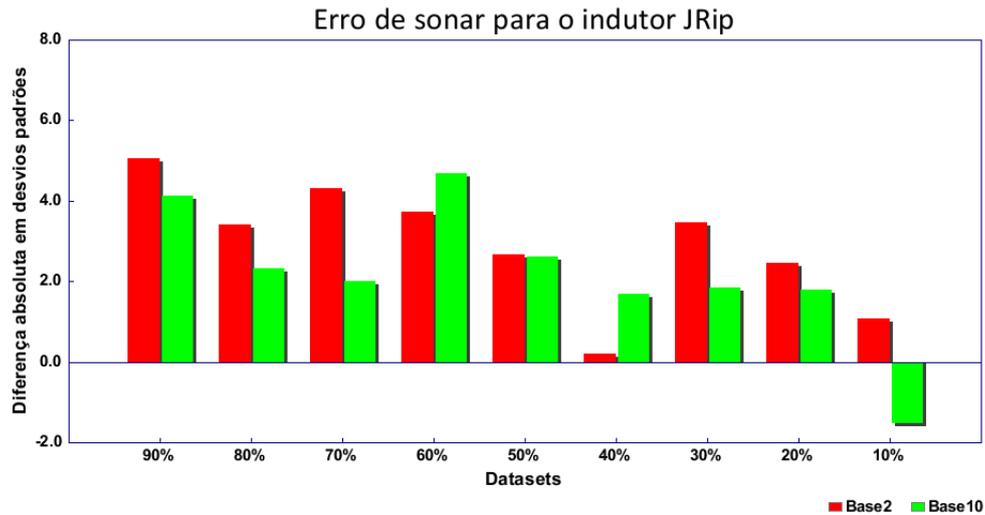


Figura 81: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

Como pode ser notado, a taxa de erro reduziu de forma significativa para a maioria dos conjuntos utilizando arredondamento com ambas as bases.

Na Figura 82 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 4,600 (sonar) para 5,089 (sonar-base2), e para 5,289 (sonar-base10). Isso significa um aumento de 10,628% para (sonar-base2) e de 14,976% para (sonar-base10) do tamanho da árvore. O tamanho da árvore aumentou para a maioria dos conjuntos, que foram arredondados com ambas as bases, embora de forma não significativa.

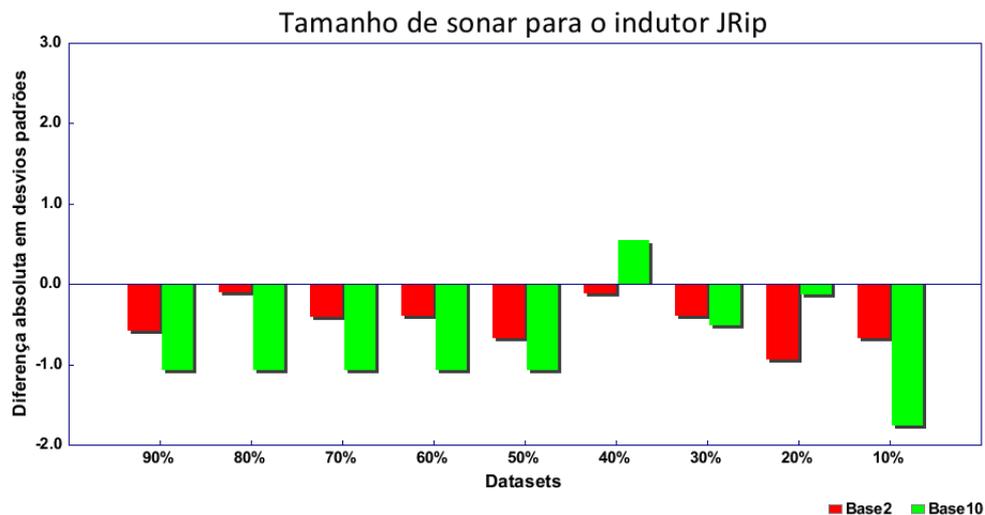


Figura 82: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) sonar

### 6.3.8 Resultados vehicle

Na Tabela 31 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *vehicle* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
vehicle	1,000 $\pm$ 0,213	1,000 $\pm$ 0,213	30,022 $\pm$ 1,704	30,022 $\pm$ 1,704	13,600 $\pm$ 1,776	13,600 $\pm$ 1,776
vehicle-90%	0,920 $\pm$ 0,144	1,005 $\pm$ 0,181	45,167 $\pm$ 1,397	57,095 $\pm$ 0,553	14,600 $\pm$ 1,350	13,100 $\pm$ 1,912
vehicle-80%	0,921 $\pm$ 0,145	1,005 $\pm$ 0,187	45,167 $\pm$ 1,397	57,095 $\pm$ 0,553	14,600 $\pm$ 1,350	13,100 $\pm$ 1,912
vehicle-70%	0,842 $\pm$ 0,135	1,005 $\pm$ 0,181	51,195 $\pm$ 3,024	57,095 $\pm$ 0,553	14,600 $\pm$ 1,174	13,100 $\pm$ 1,912
vehicle-60%	0,948 $\pm$ 0,103	1,005 $\pm$ 0,181	57,218 $\pm$ 1,341	57,095 $\pm$ 0,553	14,000 $\pm$ 1,633	13,100 $\pm$ 1,912
vehicle-50%	0,889 $\pm$ 0,126	1,005 $\pm$ 0,179	57,584 $\pm$ 1,757	57,095 $\pm$ 0,553	15,000 $\pm$ 2,000	13,100 $\pm$ 1,912
vehicle-40%	0,916 $\pm$ 0,143	1,007 $\pm$ 0,182	58,398 $\pm$ 1,418	57,095 $\pm$ 0,553	14,100 $\pm$ 2,079	13,100 $\pm$ 1,912
vehicle-30%	0,978 $\pm$ 0,309	1,037 $\pm$ 0,126	57,332 $\pm$ 2,011	57,444 $\pm$ 1,205	14,200 $\pm$ 4,022	13,100 $\pm$ 1,595
vehicle-20%	0,976 $\pm$ 0,209	0,955 $\pm$ 0,180	57,585 $\pm$ 1,590	54,493 $\pm$ 1,096	13,100 $\pm$ 2,558	12,600 $\pm$ 1,955
vehicle-10%	0,942 $\pm$ 0,092	0,535 $\pm$ 0,075	69,619 $\pm$ 0,759	74,590 $\pm$ 0,335	13,600 $\pm$ 1,578	5,300 $\pm$ 1,567

Tabela 31: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 *vehicle*

Na Figura 83 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *vehicle* e *vehicle-90%*, entre *vehicle* e *vehicle-80%* e assim por diante, utilizando base 2 e base 10.

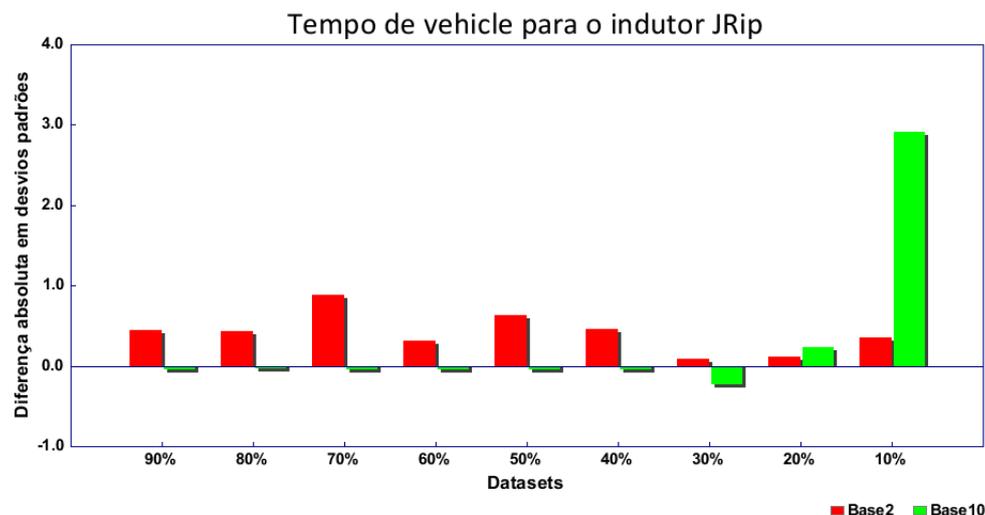


Figura 83: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *vehicle*

Para os conjuntos que tiveram seus valores arredondados utilizando a base 2, todos reduziram o tempo de indução de forma não significativa. Para o arredondamento com a base 10, o tempo de indução aumentou ligeiramente para a maioria dos conjuntos, porém o conjunto *vehicle-10%* teve uma redução significativa.

Analogamente, na Figura 84 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 30,022% (*vehicle*), para 55,474% utilizando o arredondamento com a base 2, o que representa um aumento de 84,777% da taxa de erro. E aumentou para 58,789% utilizando o arredondamento com a base 10, representando assim um aumento de 95,818%.

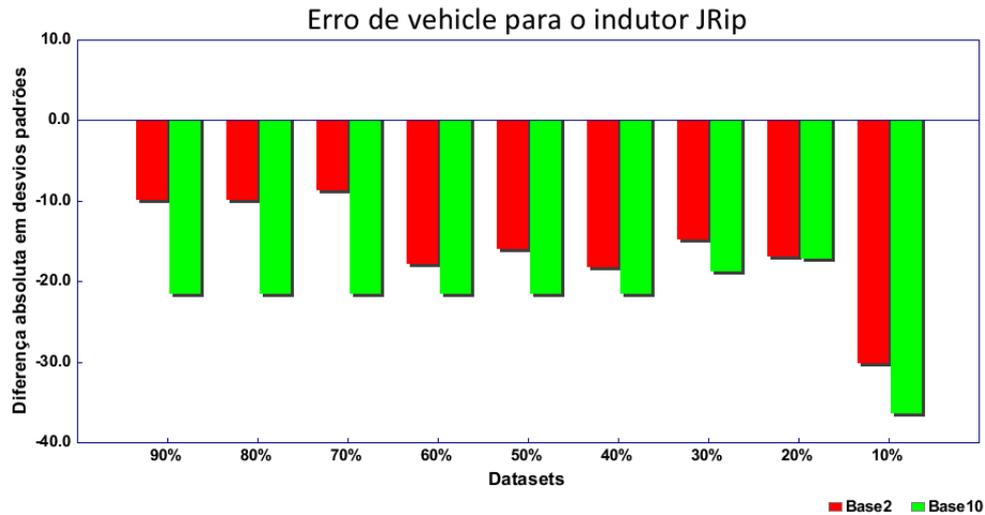


Figura 84: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vehicle

A taxa de erro aumentou para todos os conjuntos de forma significativa com arredondamento utilizando ambas as bases.

Na Figura 85 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 13,600 (vehicle) para 14,200 (vehicle-base2), e reduziu para 12,178 (vehicle-base10). Isso significa um aumento de 4,412% para (vehicle-base2) e uma redução de 10,458% para (vehicle-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para a maioria dos conjuntos com o arredondamento utilizando base 2. E reduziu para todos os conjuntos com o arredondamento utilizando a base 10, sendo o conjunto vehicle-10% o único a ter uma redução significativa.

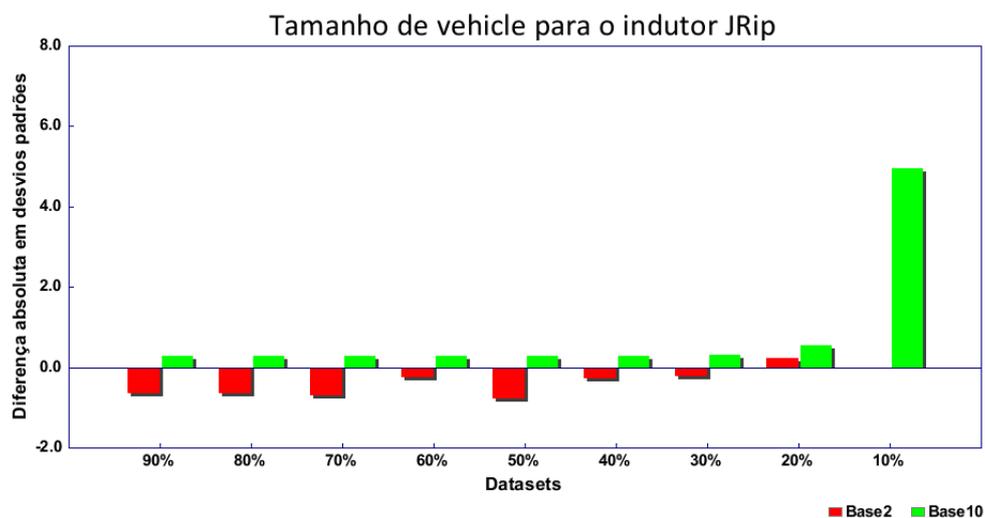


Figura 85: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) vehicle

### 6.3.9 Resultados vowel

Na Tabela 32 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos vowel original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
vowel	1,000 $\pm$ 0,101	1,000 $\pm$ 0,101	28,586 $\pm$ 1,521	28,586 $\pm$ 1,521	40,500 $\pm$ 2,321	40,500 $\pm$ 2,321
vowel-90%	1,189 $\pm$ 0,175	0,928 $\pm$ 0,055	32,424 $\pm$ 1,367	29,091 $\pm$ 1,748	41,600 $\pm$ 3,239	44,300 $\pm$ 3,164
vowel-80%	1,013 $\pm$ 0,092	0,916 $\pm$ 0,051	28,384 $\pm$ 1,131	29,091 $\pm$ 1,748	42,800 $\pm$ 2,658	44,300 $\pm$ 3,164
vowel-70%	1,027 $\pm$ 0,089	0,919 $\pm$ 0,047	31,919 $\pm$ 1,845	29,091 $\pm$ 1,748	43,600 $\pm$ 2,591	44,300 $\pm$ 3,164
vowel-60%	0,995 $\pm$ 0,063	0,921 $\pm$ 0,054	29,394 $\pm$ 0,884	29,091 $\pm$ 1,748	41,300 $\pm$ 2,669	44,300 $\pm$ 3,164
vowel-50%	0,971 $\pm$ 0,091	0,922 $\pm$ 0,054	27,980 $\pm$ 1,429	28,586 $\pm$ 1,750	43,300 $\pm$ 3,561	44,100 $\pm$ 3,315
vowel-40%	1,014 $\pm$ 0,069	0,981 $\pm$ 0,055	27,172 $\pm$ 0,713	29,798 $\pm$ 1,454	44,900 $\pm$ 2,331	42,900 $\pm$ 3,929
vowel-30%	0,929 $\pm$ 0,050	0,791 $\pm$ 0,049	29,192 $\pm$ 1,015	29,091 $\pm$ 1,702	43,200 $\pm$ 3,645	47,600 $\pm$ 4,115
vowel-20%	0,888 $\pm$ 0,087	0,776 $\pm$ 0,049	29,798 $\pm$ 1,035	28,485 $\pm$ 1,722	45,300 $\pm$ 3,433	48,500 $\pm$ 4,249
vowel-10%	0,813 $\pm$ 0,047	0,776 $\pm$ 0,051	31,010 $\pm$ 1,820	28,485 $\pm$ 1,722	47,200 $\pm$ 2,348	48,500 $\pm$ 4,249

Tabela 32: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 vowel

Na Figura 86 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre vowel e vowel-90%, entre vowel e vowel-80% e assim por diante, utilizando base 2 e base 10.

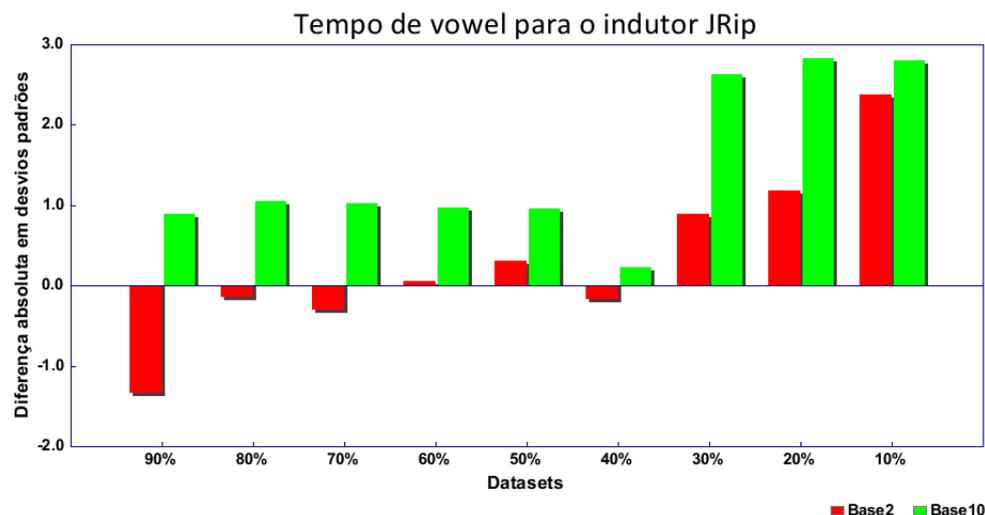


Figura 86: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

Como pode ser visto, o tempo de indução reduziu para a maioria dos conjuntos, sendo de forma significativa para o conjunto vowel-10%, utilizando arredondamento com base 2 e para os conjuntos vowel-30%, vowel-20% e vowel-10% para o arredondamento com a base 10.

Analogamente, na Figura 87 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 28,586% (vehicle), para 29,697% utilizando o arredondamento com a base 2, o que representa um aumento de 3,887% da taxa de erro. E aumentou para 28,979% utilizando o arredondamento com a base 10, representando assim um aumento de 1,374%.

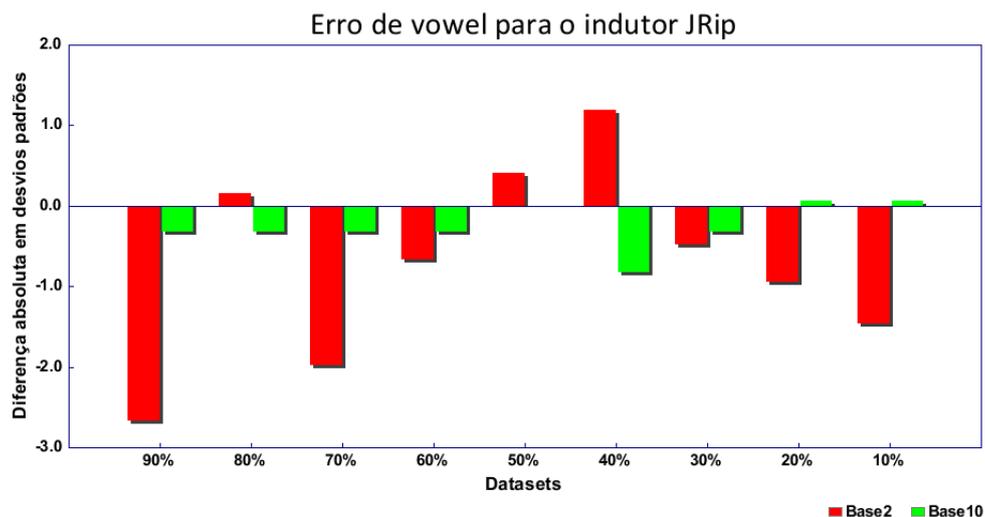


Figura 87: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

Para a maioria dos conjuntos, a taxa de erro aumentou, sendo de forma significativa para o conjunto vowel-90% utilizando arredondamento com a base 2.

Na Figura 88 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore aumentou de 40,500 (vowel) para 43,689 (vowel-base2), e para 45,422 (vowel-base10). Isso significa um aumento de 7,874% para (vowel-base2) e de 12,154% para (vowel-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para a maioria dos conjuntos arredondados com ambas as bases, sendo de forma significativa para o conjunto vowel-10%, utilizando arredondamento com base 2 e para os conjuntos vowel-30%, vowel-20% e vowel-10% para o arredondamento com a base 10.

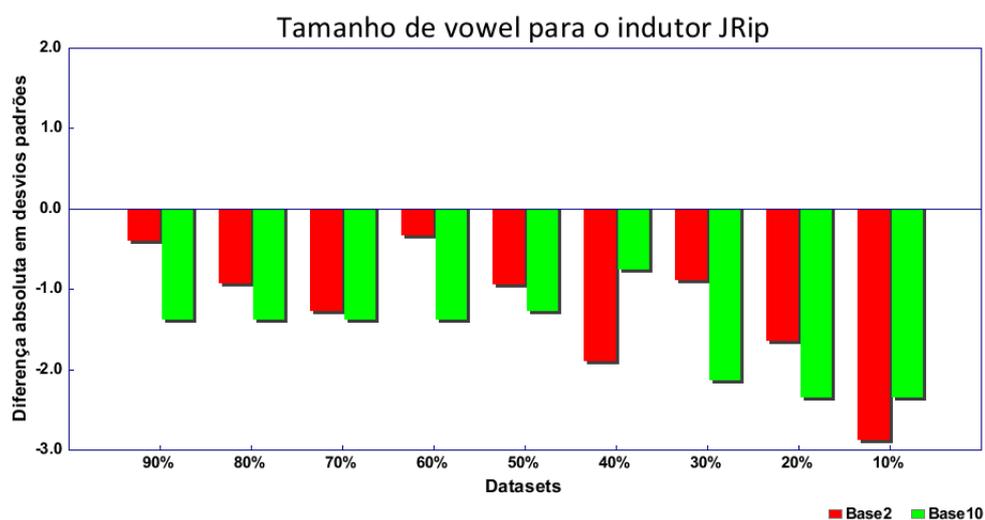


Figura 88: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) vowel

### 6.3.10 Resultados wine

Na Tabela 33 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos wine original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente. A sexta e sétima colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
wine	1,000 $\pm$ 0,035	1,000 $\pm$ 0,035	7,810 $\pm$ 1,693	7,810 $\pm$ 1,693	4,200 $\pm$ 0,632	4,200 $\pm$ 0,632
wine-90%	0,986 $\pm$ 0,048	0,925 $\pm$ 0,048	7,320 $\pm$ 2,089	5,588 $\pm$ 1,435	4,400 $\pm$ 0,843	4,400 $\pm$ 0,699
wine-80%	0,986 $\pm$ 0,048	0,925 $\pm$ 0,058	11,242 $\pm$ 1,876	5,588 $\pm$ 1,435	4,400 $\pm$ 0,699	4,400 $\pm$ 0,699
wine-70%	0,938 $\pm$ 0,033	0,932 $\pm$ 0,035	7,843 $\pm$ 2,075	10,065 $\pm$ 2,444	3,900 $\pm$ 0,994	4,500 $\pm$ 0,527
wine-60%	0,911 $\pm$ 0,046	0,932 $\pm$ 0,058	10,621 $\pm$ 1,915	11,209 $\pm$ 2,020	3,700 $\pm$ 0,823	3,900 $\pm$ 0,568
wine-50%	0,932 $\pm$ 0,058	0,911 $\pm$ 0,046	11,797 $\pm$ 3,506	11,209 $\pm$ 2,020	3,400 $\pm$ 0,699	3,800 $\pm$ 0,632
wine-40%	0,890 $\pm$ 0,032	0,918 $\pm$ 0,035	7,288 $\pm$ 1,660	11,209 $\pm$ 2,020	3,900 $\pm$ 0,994	3,800 $\pm$ 0,632
wine-30%	0,932 $\pm$ 0,058	0,911 $\pm$ 0,046	11,797 $\pm$ 1,931	13,464 $\pm$ 2,650	4,200 $\pm$ 0,632	4,100 $\pm$ 0,876
wine-20%	0,911 $\pm$ 0,033	0,897 $\pm$ 0,051	10,065 $\pm$ 2,597	13,431 $\pm$ 1,866	4,300 $\pm$ 0,483	3,800 $\pm$ 0,919
wine-10%	0,911 $\pm$ 0,046	0,973 $\pm$ 0,063	14,608 $\pm$ 2,374	34,804 $\pm$ 3,877	4,400 $\pm$ 0,516	4,900 $\pm$ 1,370

Tabela 33: Tempo de indução e taxa de erro do classificador para o indutor JRip utilizando arredondamento com bases 2 e 10 wine

Na Figura 89 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre wine e wine-90%, entre wine e wine-80% e assim por diante, utilizando base 2 e base 10.

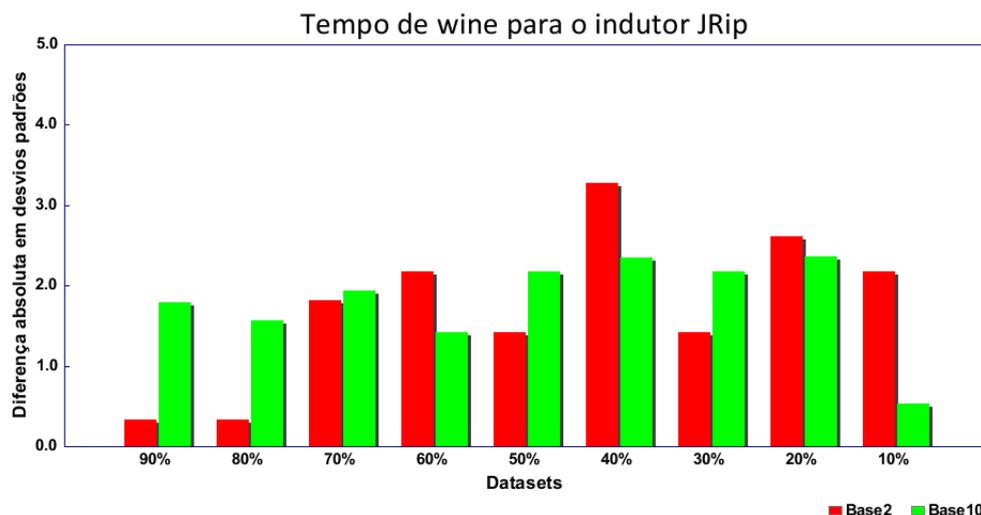


Figura 89: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 versus conjunto original) wine

O tempo de indução reduziu para a todos os conjuntos, sendo de forma significativa para metade destes conjuntos.

Analogamente, na Figura 90 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 7,810% (wine), para 10,287% utilizando o arredondamento com a base 2, o que representa um aumento de 31,713% da taxa de erro. E aumentou para 12,952% utilizando o arredondamento com a base 10, representando assim um aumento de 65,837%.

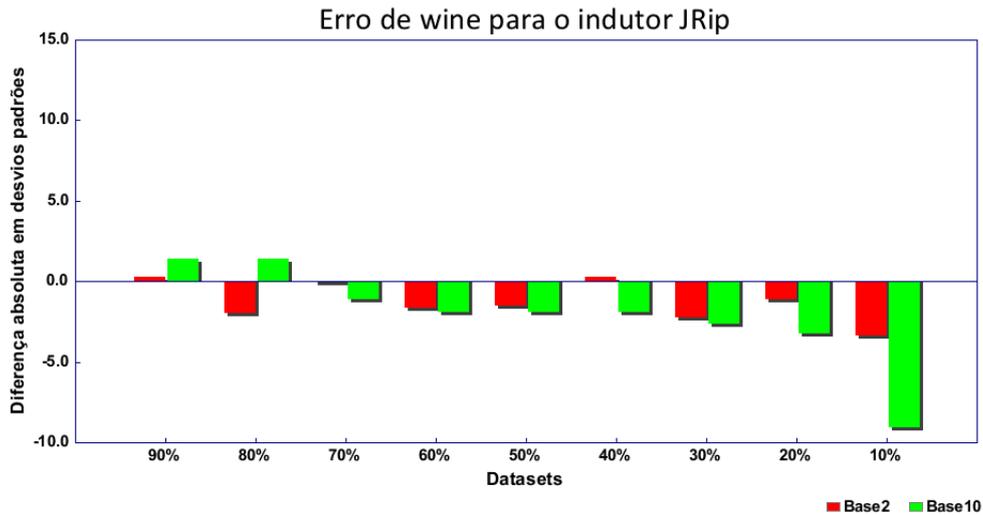


Figura 90: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

A taxa de erro aumentou para a maioria dos conjuntos arredondados com ambas as bases, sendo de forma significativa para os conjuntos wine-30% e wine-10% utilizando arredondamento com base 2 e para os conjuntos wine-30%, wine-20% e wine-10% para o arredondamento com a base 10.

Na Figura 91 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore no eixo vertical do gráfico. Em média, o tamanho da árvore reduziu de 4,200 (wine) para 4,067 (wine-base2), e para 4,178 (wine-base10). Isso significa uma redução de 3,175% para (wine-base2) e de 0,529% para (wine-base10) do tamanho da árvore. Como pode ser notado o tamanho da árvore não manteve um padrão, reduzindo para alguns conjuntos e aumentando para outros, sendo de forma não significativa para ambas as bases.

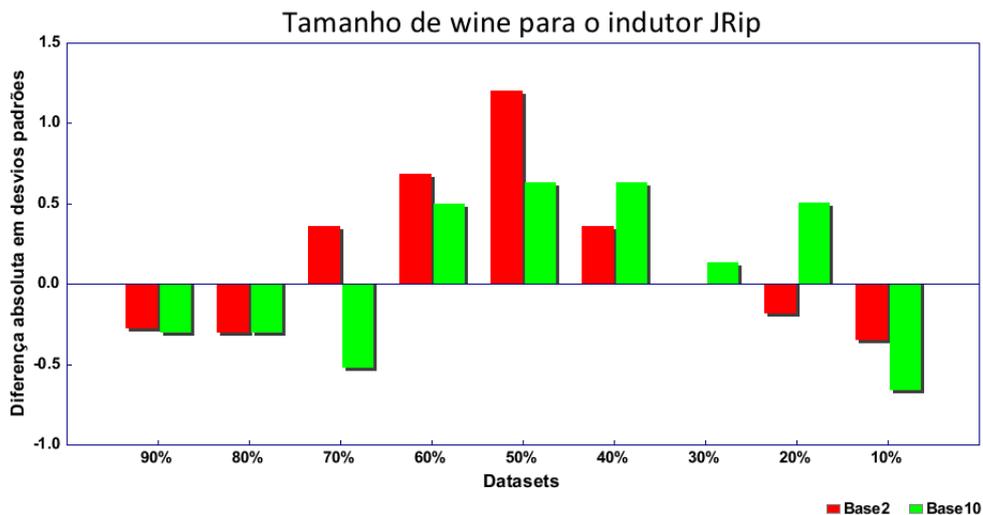


Figura 91: Diferença absoluta do tamanho do classificador (arredondamento utilizando bases 2 e 10 *versus* conjunto original) wine

### 6.3.11 Discussão

Nas Figuras 92, 93, 94, 95, 96 e 97, é mostrado um resumo dos resultados obtidos por meio do indutor JRip do tempo de indução utilizando base 2, tempo de indução utilizando base 10, taxa de erro utilizando base 2, taxa de erro utilizando base 10, tamanho do classificador utilizando base 2, tamanho do classificador utilizando base 10, respectivamente. No total são 90 conjuntos (9 conjuntos arredondados  $\times$  10 conjuntos de exemplos). No eixo  $y$  do gráfico é mostrado o número conjuntos que ficaram em cada uma das categorias mostradas no eixo  $x$  do gráfico, que são:

- aqueles que tiveram o tempo significativamente maior para os conjuntos derivados em relação ao conjunto original (barras maior que -2);
- os que tiveram o tempo maior para os conjuntos derivados (barras variando entre -2 e 0);
- os que tiveram o tempo igual para os conjuntos derivados e o conjunto original (barras igual 0);
- os que tiveram o tempo menor para os conjuntos derivados (barras variando entre 0 e 2);
- os que tiveram o tempo significativamente menor para os conjuntos derivados em relação ao conjunto original (barras maior que 2);

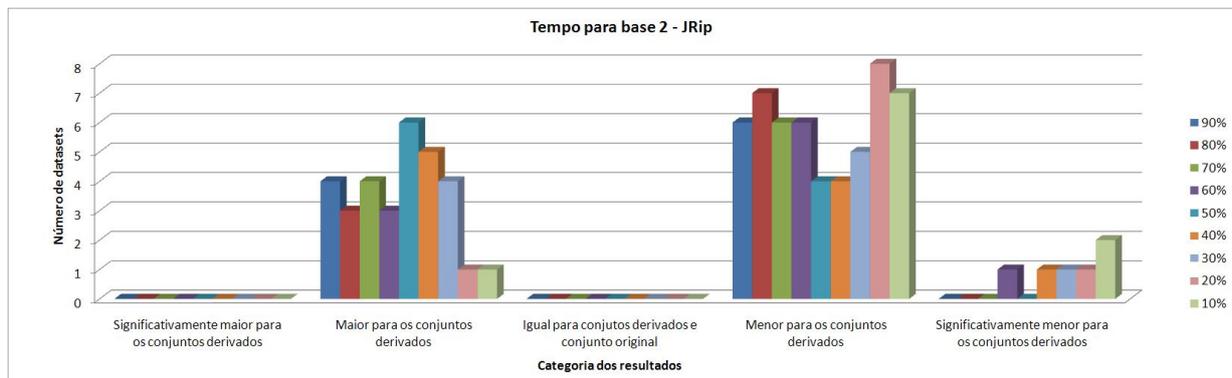


Figura 92: Resumo dos resultados do tempo de indução utilizando base 2 de JRip

Como pode ser visto, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma não significativa.

Analogamente para a taxa de erro e tamanho do classificador.

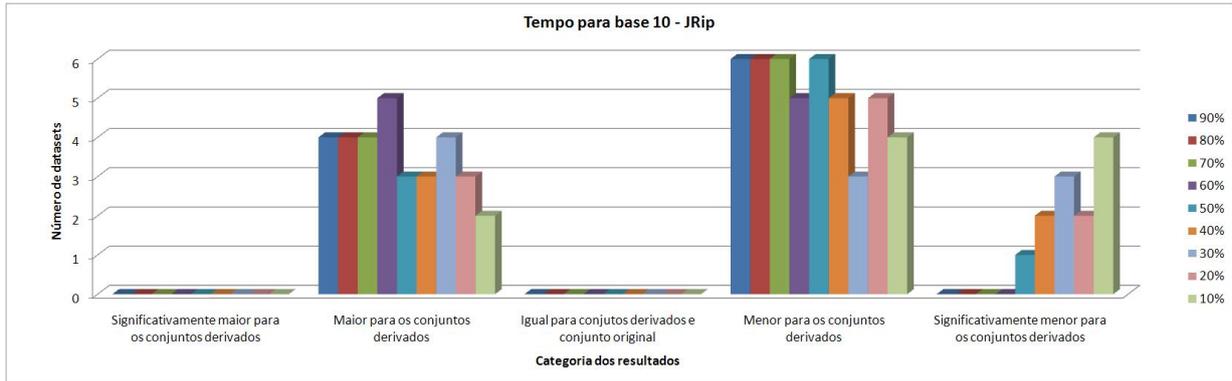


Figura 93: Resumo dos resultados do tempo de indução utilizando base 10 de JRip

Como pode ser notado, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma não significativa.

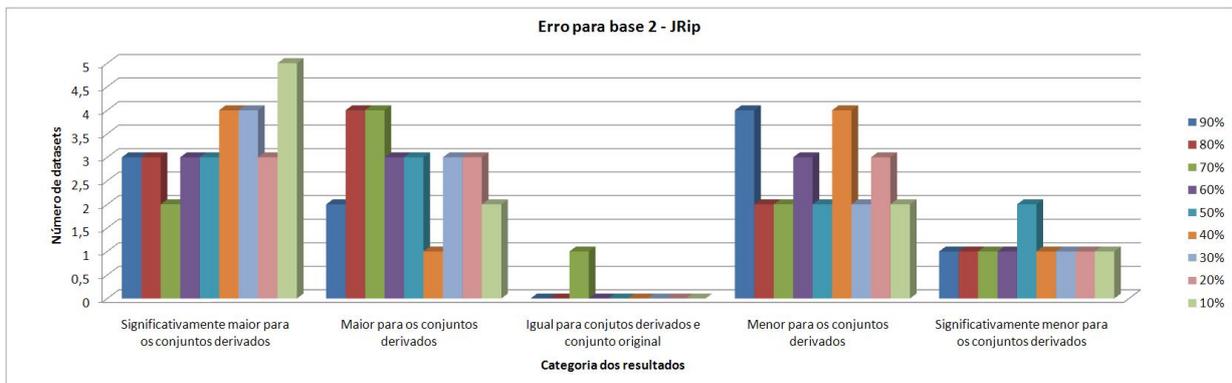


Figura 94: Resumo dos resultados da taxa de erro utilizando base 2 de JRip

Como pode ser observado, houve uma tendência da taxa de erro aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma significativa.

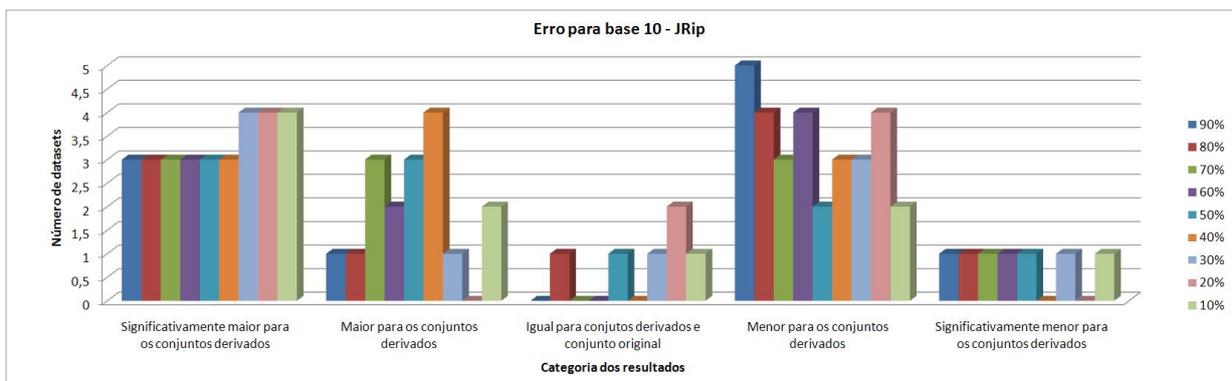


Figura 95: Resumo dos resultados da taxa de erro utilizando base 10 de JRip

O mesmo número de conjuntos teve um aumento significativo e uma redução não significativa na taxa de erro aumentar para os conjuntos derivados, porém no geral houve mais aumentos do

que reduções na taxa de erro.

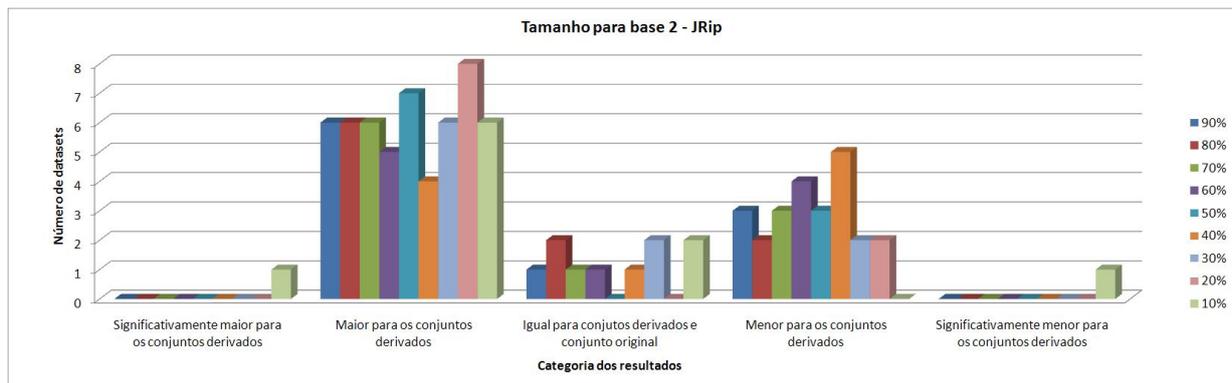


Figura 96: Resumo dos resultados do tamanho do classificador utilizando base 2 de JRip

Como pode ser visto, houve uma tendência do tamanho do classificador aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa.

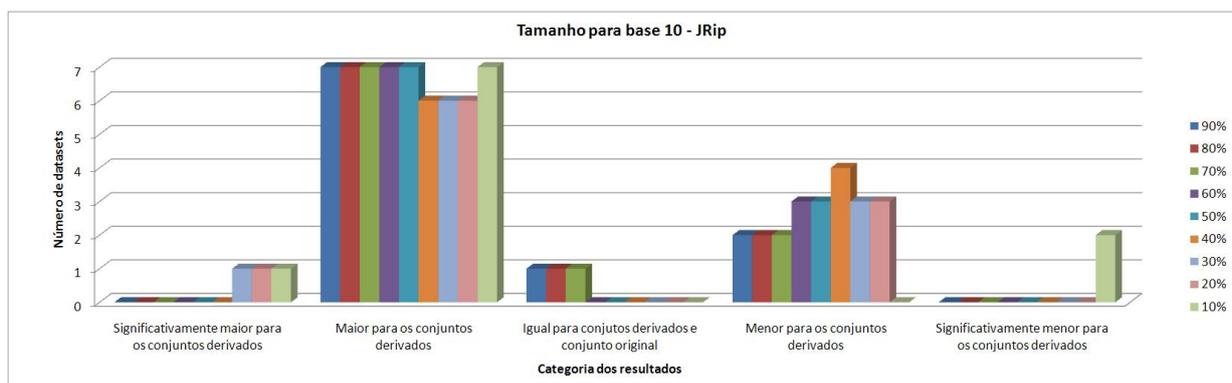


Figura 97: Resumo dos resultados do tamanho do classificador utilizando base 10 de JRip

Como pode ser notado, houve uma tendência do tamanho do classificador aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa.

## 6.4 Resultados do indutor MLP

### 6.4.1 Resultados aml-all-completo

Como após 3 dias de processamento os resultados dos experimentos não haviam sido processados pelo indutor, os experimentos para esse conjunto de exemplos foram cancelados.

### 6.4.2 Resultados eucalyptus

Na Tabela 34 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos eucalyptus original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente.

A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
eucalyptus	1,000 ± 0,009	1,000 ± 0,009	37,892 ± 2,018	37,892 ± 2,018
eucalyptus-90%	0,996 ± 0,004	0,994 ± 0,005	45,642 ± 1,468	47,549 ± 1,486
eucalyptus-80%	0,995 ± 0,004	0,994 ± 0,005	48,077 ± 1,748	47,549 ± 1,486
eucalyptus-70%	0,998 ± 0,010	0,995 ± 0,005	50,946 ± 1,866	47,549 ± 1,486
eucalyptus-60%	0,993 ± 0,003	0,995 ± 0,005	46,866 ± 1,541	49,315 ± 1,706
eucalyptus-50%	0,996 ± 0,009	0,997 ± 0,006	49,315 ± 1,508	49,852 ± 1,704
eucalyptus-40%	0,995 ± 0,003	0,993 ± 0,002	48,501 ± 1,855	50,396 ± 1,787
eucalyptus-30%	0,995 ± 0,003	0,993 ± 0,002	52,973 ± 1,598	53,662 ± 1,358
eucalyptus-20%	0,997 ± 0,003	0,996 ± 0,003	58,680 ± 0,964	55,170 ± 1,433
eucalyptus-10%	1,000 ± 0,008	1,005 ± 0,014	73,493 ± 0,882	75,144 ± 1,521

Tabela 34: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 eucalyptus

Na Figura 98 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre eucalyptus e eucalyptus-90%, entre eucalyptus e eucalyptus-80% e assim por diante, utilizando base 2 e base 10.

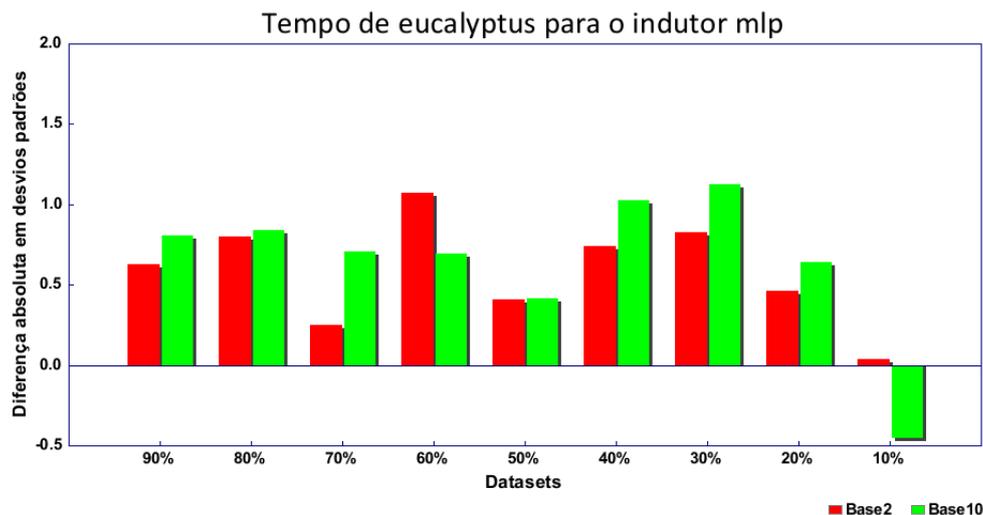


Figura 98: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) eucalyptus

No geral, o tempo de indução reduziu para a maioria dos conjuntos utilizando arredondamento com as bases 2 e 10, embora de forma não significativa.

Analogamente, na Figura 99 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 37,892% (eucalyptus), para 52,721% utilizando o arredondamento com a base 2, o que representa um aumento de 39,136% da taxa de erro. E aumentou para 52,910% utilizando o arredondamento com a base 10, representando assim um aumento de 39,633%.

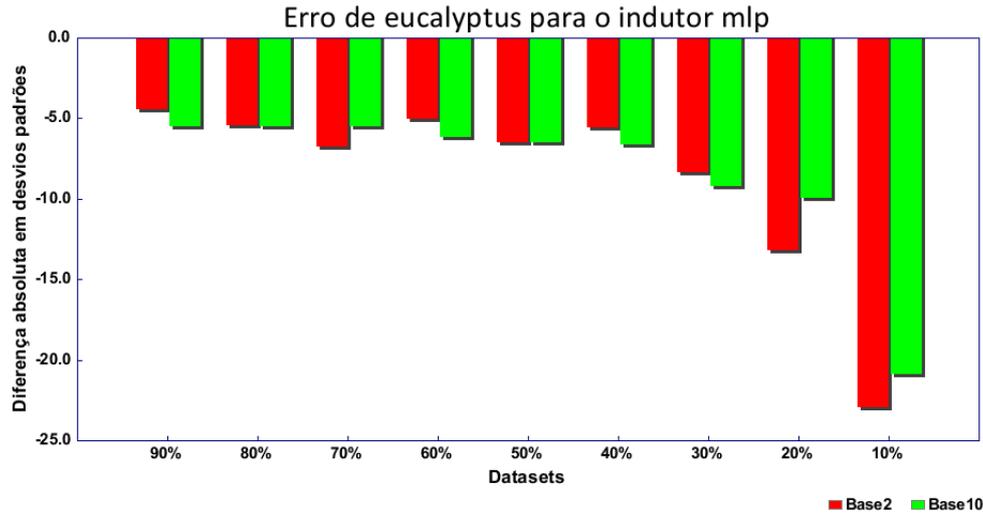


Figura 99: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) eucalyptus

Como pode ser visto, a taxa de erro aumentou para todos os conjuntos de forma significativa utilizando arredondamento com ambas as bases.

### 6.4.3 Resultados hepatitis

Na Tabela 35 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos hepatitis original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
hepatitis	1,000 $\pm$ 0,015	1,000 $\pm$ 0,015	18,042 $\pm$ 2,728	18,042 $\pm$ 2,728
hepatitis-90%	0,961 $\pm$ 0,008	0,967 $\pm$ 0,024	20,042 $\pm$ 2,298	20,083 $\pm$ 2,296
hepatitis-80%	0,960 $\pm$ 0,006	0,966 $\pm$ 0,023	20,708 $\pm$ 2,612	20,083 $\pm$ 2,296
hepatitis-70%	0,956 $\pm$ 0,005	0,966 $\pm$ 0,024	20,083 $\pm$ 2,736	20,083 $\pm$ 2,296
hepatitis-60%	0,959 $\pm$ 0,005	0,966 $\pm$ 0,023	20,750 $\pm$ 2,633	20,083 $\pm$ 2,296
hepatitis-50%	0,959 $\pm$ 0,005	0,965 $\pm$ 0,023	20,083 $\pm$ 2,127	20,083 $\pm$ 2,296
hepatitis-40%	0,958 $\pm$ 0,007	0,966 $\pm$ 0,023	20,042 $\pm$ 2,100	20,083 $\pm$ 2,296
hepatitis-30%	0,957 $\pm$ 0,006	0,954 $\pm$ 0,003	21,958 $\pm$ 3,575	20,625 $\pm$ 2,268
hepatitis-20%	0,960 $\pm$ 0,005	0,955 $\pm$ 0,004	20,583 $\pm$ 2,870	20,000 $\pm$ 2,481
hepatitis-10%	0,962 $\pm$ 0,013	0,956 $\pm$ 0,006	19,958 $\pm$ 2,458	22,708 $\pm$ 3,355

Tabela 35: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 hepatitis

Na Figura 100 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre hepatitis e hepatitis-90%, entre hepatitis e hepatitis-80% e assim por diante, utilizando base 2 e base 10.

O tempo de indução reduziu de forma significativa para todos os conjuntos que utiliza o

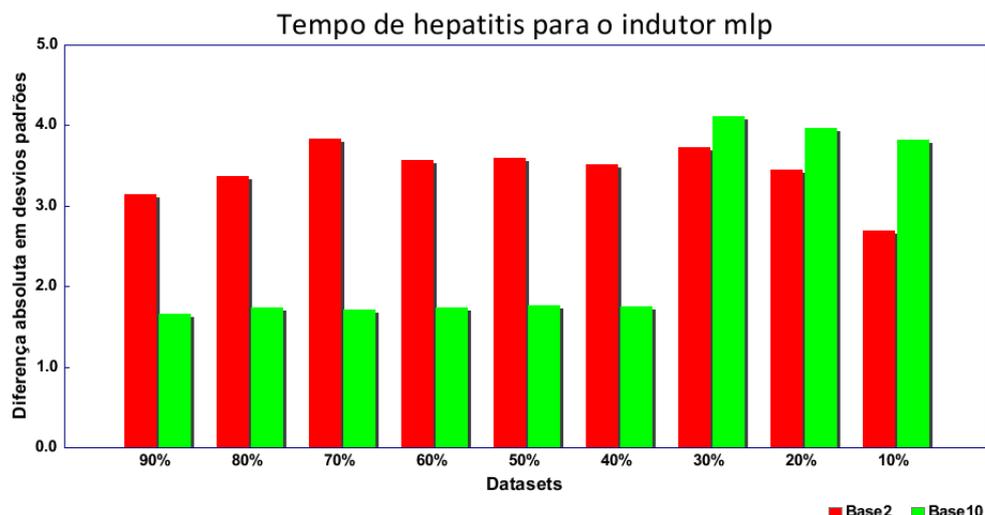


Figura 100: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

arredondamento com a base 2, também reduziu para os arredondados com a base 10, porém de forma significativa apenas para hepatitis-30%, hepatitis-20% e hepatitis-10%.

Analogamente, na Figura 101 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 18,042% (hepatitis), para 20,467% utilizando o arredondamento com a base 2, o que representa uma redução de 13,443% da taxa de erro. E aumentou para 20,426% utilizando o arredondamento com a base 10, representando assim um aumento de 13,212%.

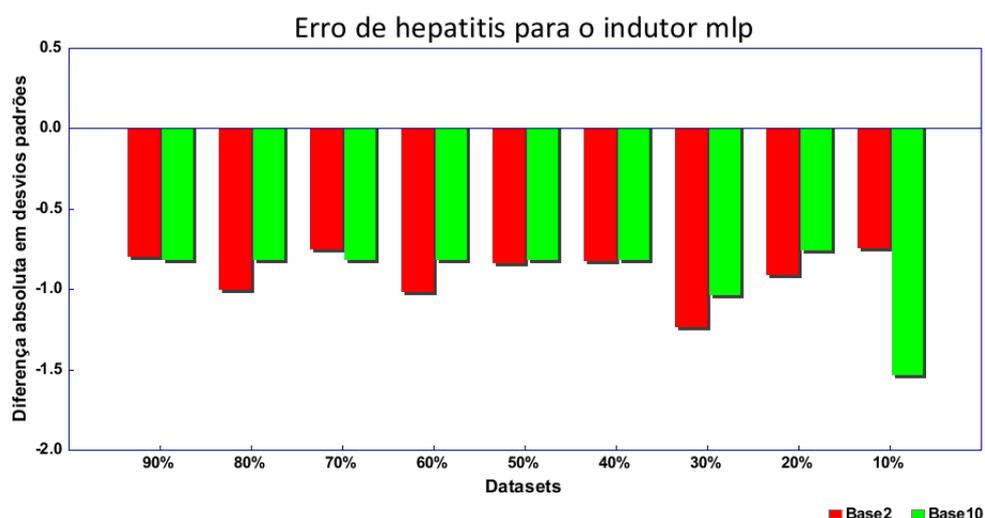


Figura 101: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

Como pode ser visto, a taxa de erro aumentou para todos os conjuntos de forma não significativa, utilizando arredondamento com a bases 2 e 10.

#### 6.4.4 Resultados ionosphere

Na Tabela 36 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *ionosphere* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
<i>ionosphere</i>	1,000 $\pm$ 0,018	1,000 $\pm$ 0,018	8,548 $\pm$ 1,127	8,548 $\pm$ 1,127
<i>ionosphere-90%</i>	1,001 $\pm$ 0,014	0,994 $\pm$ 0,005	9,103 $\pm$ 1,091	8,817 $\pm$ 1,219
<i>ionosphere-80%</i>	0,997 $\pm$ 0,009	0,994 $\pm$ 0,005	9,103 $\pm$ 1,091	8,817 $\pm$ 1,219
<i>ionosphere-70%</i>	0,998 $\pm$ 0,007	0,994 $\pm$ 0,005	9,389 $\pm$ 1,028	8,817 $\pm$ 1,219
<i>ionosphere-60%</i>	0,993 $\pm$ 0,008	0,994 $\pm$ 0,005	9,675 $\pm$ 0,951	8,817 $\pm$ 1,219
<i>ionosphere-50%</i>	1,004 $\pm$ 0,018	0,999 $\pm$ 0,016	9,389 $\pm$ 1,191	9,405 $\pm$ 1,283
<i>ionosphere-40%</i>	0,992 $\pm$ 0,005	0,992 $\pm$ 0,006	8,817 $\pm$ 1,060	7,127 $\pm$ 1,366
<i>ionosphere-30%</i>	0,993 $\pm$ 0,004	0,994 $\pm$ 0,004	8,817 $\pm$ 1,425	8,270 $\pm$ 1,444
<i>ionosphere-20%</i>	0,993 $\pm$ 0,007	0,993 $\pm$ 0,004	9,683 $\pm$ 1,138	8,270 $\pm$ 1,444
<i>ionosphere-10%</i>	0,994 $\pm$ 0,008	0,993 $\pm$ 0,004	9,984 $\pm$ 1,369	7,984 $\pm$ 1,585

Tabela 36: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 *ionosphere*

Na Figura 102 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *ionosphere* e *ionosphere-90%*, entre *ionosphere* e *ionosphere-80%* e assim por diante, utilizando base 2 e base 10.

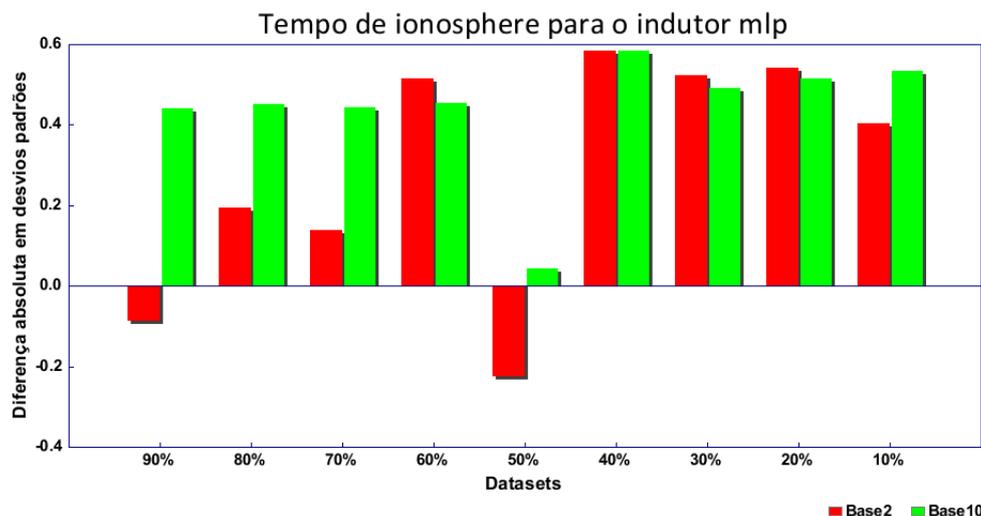


Figura 102: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *ionosphere*

No geral, o tempo de indução aumentou para todos os conjuntos de forma não significativa, tanto para base 2 como para a base 10.

Analogamente, na Figura 103 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 8,548% (*ionosphere*), para 9,329% utilizando o arredondamento com a base 2, o que representa um aumento de 9,135% da taxa de erro. E reduziu para 8,480% utilizando o arredondamento com a base 10, representando assim uma redução de 0,790%.

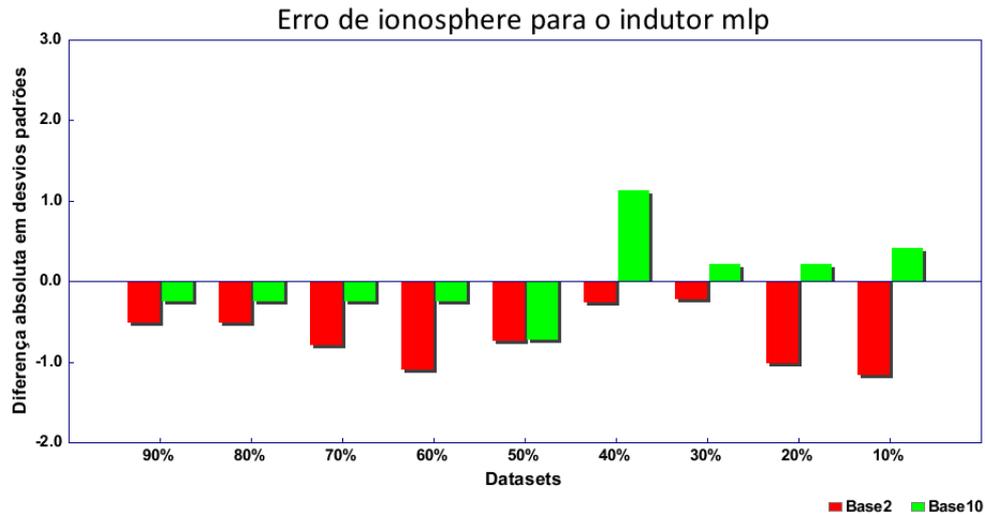


Figura 103: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) ionosphere

Como pode ser observado, no geral, a taxa de erro aumentou para todos os conjuntos de forma não significativa, utilizando arredondamento com a base 2. Já para a base 10, não se observou um padrão, pois alguns conjuntos tiveram a taxa de erro aumentada e outros reduzida.

#### 6.4.5 Resultados lymph

Na Tabela 37 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos lymph original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
lymph	1,000 $\pm$ 0,008	1,000 $\pm$ 0,008	14,905 $\pm$ 2,215	14,905 $\pm$ 2,215
lymph-90%	0,997 $\pm$ 0,008	1,001 $\pm$ 0,006	14,905 $\pm$ 2,624	20,286 $\pm$ 2,664
lymph-80%	0,997 $\pm$ 0,007	1,001 $\pm$ 0,006	14,905 $\pm$ 2,624	20,286 $\pm$ 2,664
lymph-70%	0,997 $\pm$ 0,007	1,001 $\pm$ 0,006	14,905 $\pm$ 2,624	20,286 $\pm$ 2,664
lymph-60%	0,997 $\pm$ 0,008	1,001 $\pm$ 0,006	16,238 $\pm$ 2,494	20,286 $\pm$ 2,664
lymph-50%	0,997 $\pm$ 0,007	1,001 $\pm$ 0,006	16,238 $\pm$ 2,494	20,286 $\pm$ 2,664
lymph-40%	0,997 $\pm$ 0,008	1,002 $\pm$ 0,006	17,619 $\pm$ 3,072	20,286 $\pm$ 2,664
lymph-30%	1,009 $\pm$ 0,017	1,000 $\pm$ 0,006	17,619 $\pm$ 3,072	20,286 $\pm$ 2,664
lymph-20%	0,999 $\pm$ 0,008	1,001 $\pm$ 0,006	18,333 $\pm$ 2,598	17,667 $\pm$ 2,636
lymph-10%	1,008 $\pm$ 0,007	1,005 $\pm$ 0,003	17,667 $\pm$ 2,636	17,667 $\pm$ 2,636

Tabela 37: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 lymph

Na Figura 104 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre lymph e lymph-90%, entre lymph e lymph-80% e assim por diante, utilizando base 2 e base 10.

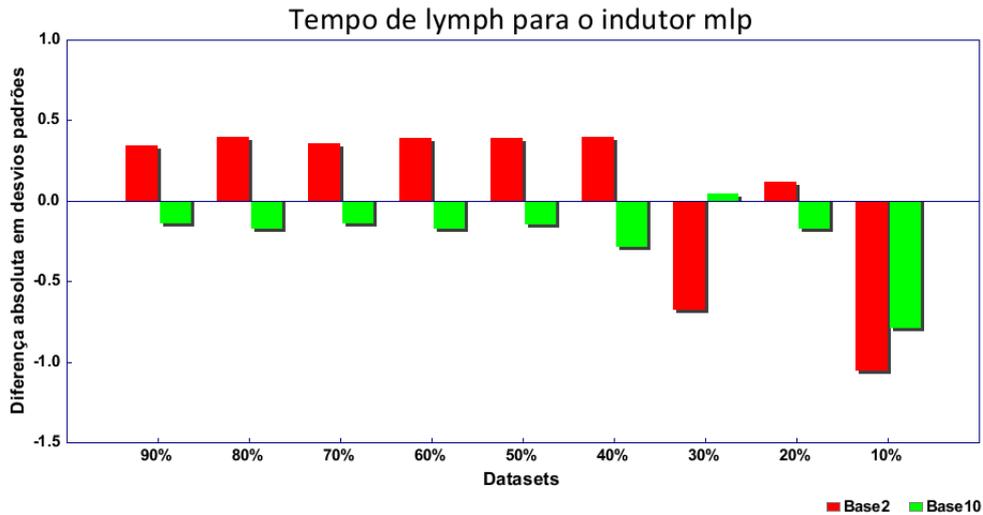


Figura 104: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

No geral, o tempo de indução reduziu para todos os conjuntos utilizando arredondamento com base 2, de forma não significativa. E para a base 10 o tempo de indução aumentou para a maioria dos conjuntos também de forma não significativa.

Analogamente, na Figura 105 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 14,905% (lymph), para 16,492% utilizando o arredondamento com a base 2, o que representa um aumento de 10,648% da taxa de erro. E aumentou para 219,704% utilizando o arredondamento com a base 10, representando assim um aumento de 32,197%.

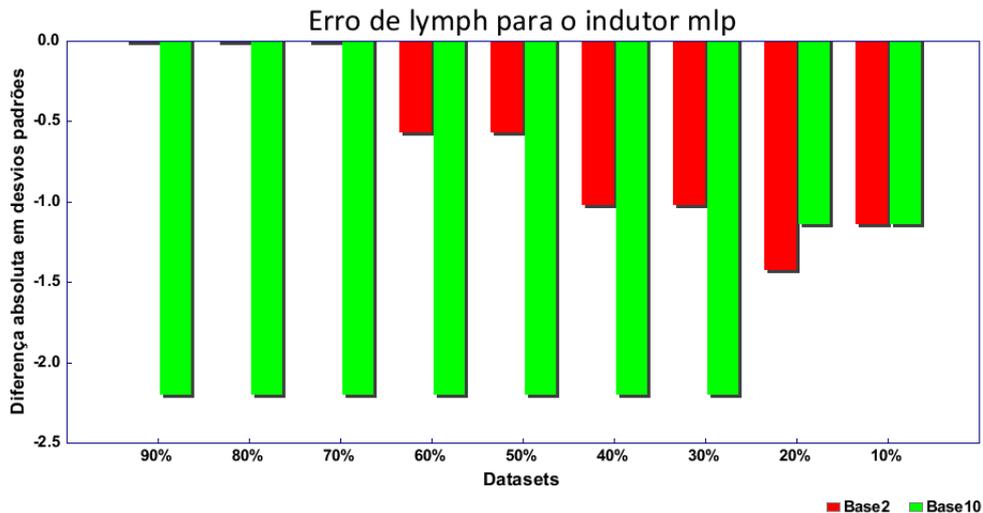


Figura 105: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

Como pode ser notado, a taxa de erro aumentou de forma significativa para 7 dos 9 conjuntos arredondados utilizando arredondamento a base 2. Já para a base 10 todos os conjuntos tiveram um aumento não significativo.

### 6.4.6 Resultados sick

Na Tabela 38 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sick original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
sick	1,000 $\pm$ 0,014	1,000 $\pm$ 0,014	2,996 $\pm$ 0,296	2,996 $\pm$ 0,296
sick-90%	1,006 $\pm$ 0,008	0,995 $\pm$ 0,013	3,579 $\pm$ 0,296	4,851 $\pm$ 0,466
sick-80%	1,002 $\pm$ 0,004	0,995 $\pm$ 0,014	4,427 $\pm$ 0,559	4,851 $\pm$ 0,466
sick-70%	0,999 $\pm$ 0,003	0,995 $\pm$ 0,013	4,693 $\pm$ 0,744	4,851 $\pm$ 0,466
sick-60%	1,003 $\pm$ 0,008	0,999 $\pm$ 0,011	4,136 $\pm$ 0,557	4,692 $\pm$ 0,423
sick-50%	1,007 $\pm$ 0,008	0,995 $\pm$ 0,006	4,534 $\pm$ 0,557	4,692 $\pm$ 0,423
sick-40%	1,007 $\pm$ 0,013	0,995 $\pm$ 0,007	4,852 $\pm$ 0,507	4,692 $\pm$ 0,423
sick-30%	1,002 $\pm$ 0,013	0,994 $\pm$ 0,008	4,745 $\pm$ 0,614	4,878 $\pm$ 0,444
sick-20%	1,011 $\pm$ 0,015	0,998 $\pm$ 0,008	9,624 $\pm$ 0,951	4,878 $\pm$ 0,444
sick-10%	1,002 $\pm$ 0,006	1,020 $\pm$ 0,026	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105

Tabela 38: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 sick

Na Figura 106 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre sick e sick-90%, entre sick e sick-80% e assim por diante, utilizando base 2 e base 10.

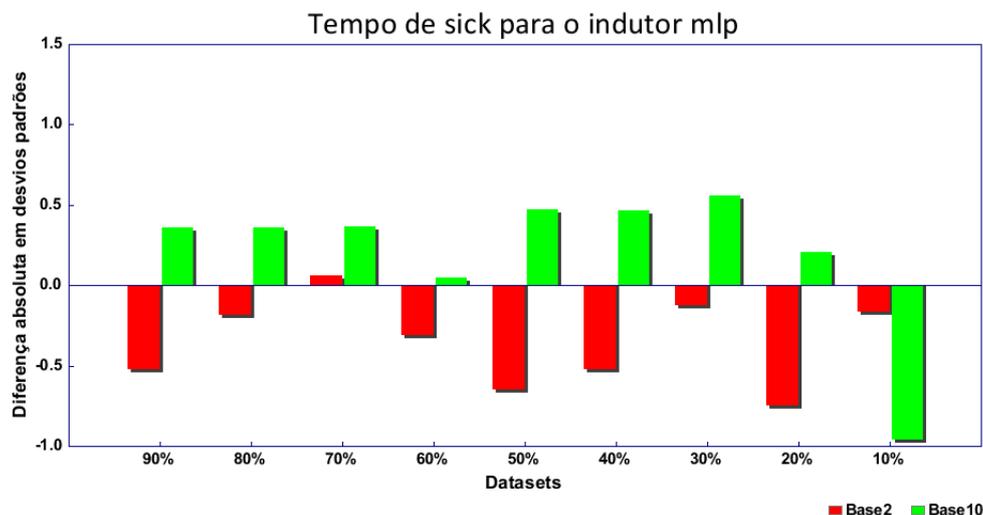


Figura 106: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 versus conjunto original) sick

Como pode ser visto, o tempo de indução aumentou para a maioria dos conjuntos utilizando arredondamento com base 2, sendo de forma não significativa. Já para a base 10, o tempo de indução reduziu para a maioria dos conjuntos, também de forma não significativa.

Analogamente, na Figura 107 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 2,996% (sick), para 5,370% utilizando o arredondamento com a base 2, o que representa um aumento de

79,239% da taxa de erro. E aumentou para 4,987% utilizando o arredondamento com a base 10, representando assim um aumento de 66,444%.

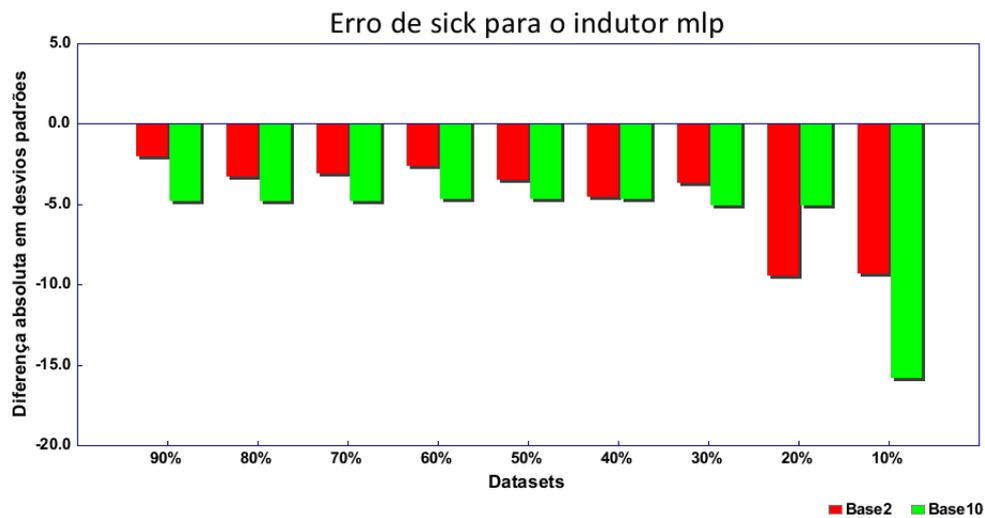


Figura 107: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

A taxa de erro aumentou para todos os conjuntos utilizando arredondamento com ambas as bases, sendo de forma significativa para todos os conjuntos, exceto para sick-80%, utilizando arredondamento com a base 2.

#### 6.4.7 Resultados sonar

Na Tabela 39 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sonar original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
sonar	1,000 $\pm$ 0,058	1,000 $\pm$ 0,058	2,996 $\pm$ 0,296	2,996 $\pm$ 0,296
sonar-90%	0,972 $\pm$ 0,058	1,000 $\pm$ 0,064	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105
sonar-80%	0,993 $\pm$ 0,060	1,000 $\pm$ 0,064	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105
sonar-70%	0,983 $\pm$ 0,061	1,001 $\pm$ 0,064	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105
sonar-60%	0,984 $\pm$ 0,060	1,001 $\pm$ 0,064	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105
sonar-50%	0,987 $\pm$ 0,066	1,000 $\pm$ 0,063	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105
sonar-40%	0,919 $\pm$ 0,042	0,961 $\pm$ 0,060	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105
sonar-30%	0,906 $\pm$ 0,005	1,020 $\pm$ 0,061	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105
sonar-20%	0,906 $\pm$ 0,004	0,985 $\pm$ 0,065	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105
sonar-10%	0,907 $\pm$ 0,005	0,899 $\pm$ 0,002	7,740 $\pm$ 0,663	6,495 $\pm$ 0,105

Tabela 39: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 sonar

Na Figura 108 é mostrada a diferença absoluta em desvios padrões do tempo de indução

entre o conjunto original e os conjuntos derivados, ou seja, entre sonar e sonar-90%, entre sonar e sonar-80% e assim por diante, utilizando base 2 e base 10.

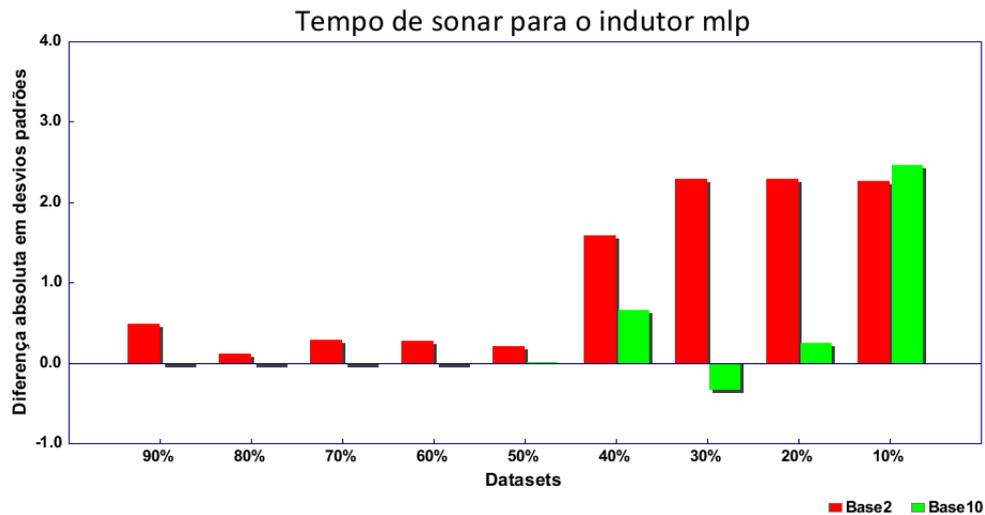


Figura 108: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

O tempo de indução reduziu para a maioria dos conjuntos, sendo de forma significativa para os conjuntos sonar-30%, sonar-20% e sonar-10% utilizando arredondamento com a base 2 e para sonar-10% utilizando arredondamento com a base 10.

Analogamente, na Figura 109 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 2,996% (sonar), para 7,740% utilizando o arredondamento com a base 2, o que representa um aumento de 158,344% da taxa de erro. E aumentou para 6,495% utilizando o arredondamento com a base 10, representando assim um aumento de 116,789%.

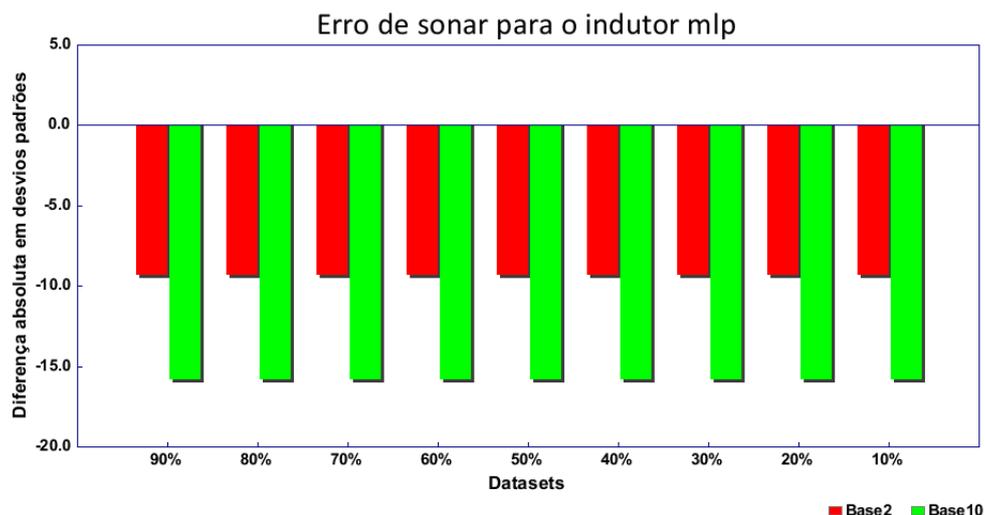


Figura 109: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

Como pode ser observado, a taxa de erro aumentou para todos os conjuntos de forma significativa, utilizando arredondamento com ambas as base.

### 6.4.8 Resultados vehicle

Na Tabela 40 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *vehicle* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
vehicle	1,000 $\pm$ 0,003	1,000 $\pm$ 0,003	17,147 $\pm$ 1,263	17,147 $\pm$ 1,263
vehicle-90%	1,008 $\pm$ 0,014	1,003 $\pm$ 0,008	23,293 $\pm$ 0,789	29,658 $\pm$ 1,677
vehicle-80%	1,009 $\pm$ 0,014	1,003 $\pm$ 0,007	23,293 $\pm$ 0,789	29,658 $\pm$ 1,677
vehicle-70%	1,017 $\pm$ 0,019	1,004 $\pm$ 0,008	22,950 $\pm$ 1,274	29,658 $\pm$ 1,677
vehicle-60%	1,028 $\pm$ 0,047	1,003 $\pm$ 0,008	27,291 $\pm$ 1,503	29,658 $\pm$ 1,677
vehicle-50%	1,008 $\pm$ 0,003	1,020 $\pm$ 0,042	30,139 $\pm$ 0,927	29,658 $\pm$ 1,677
vehicle-40%	1,008 $\pm$ 0,004	1,003 $\pm$ 0,008	29,080 $\pm$ 0,870	29,658 $\pm$ 1,677
vehicle-30%	1,006 $\pm$ 0,007	1,002 $\pm$ 0,006	38,199 $\pm$ 2,172	29,078 $\pm$ 1,396
vehicle-20%	1,009 $\pm$ 0,015	1,011 $\pm$ 0,007	47,380 $\pm$ 3,227	27,884 $\pm$ 2,083
vehicle-10%	1,007 $\pm$ 0,008	1,011 $\pm$ 0,017	64,297 $\pm$ 2,217	63,452 $\pm$ 1,858

Tabela 40: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 *vehicle*

Na Figura 110 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *vehicle* e *vehicle-90%*, entre *vehicle* e *vehicle-80%* e assim por diante, utilizando base 2 e base 10.

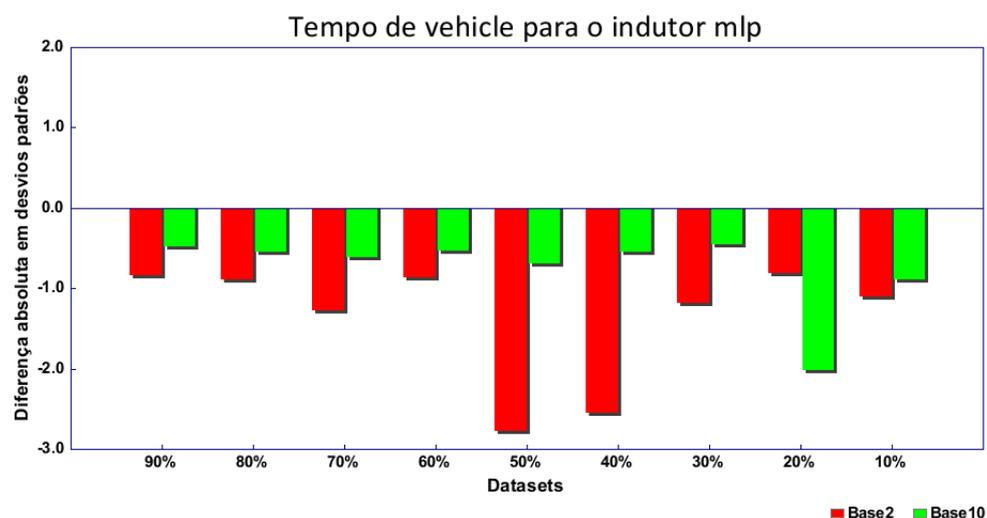


Figura 110: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *vehicle*

Podemos observar que o tempo de indução aumentou para todos os conjuntos, sendo de forma significativa para os conjuntos *vehicle-50%* e *vehicle-40%* utilizando arredondamento com a base 2 e para *vehicle-20%* utilizando arredondamento com a base 10.

Analogamente, na Figura 111 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 17,147% (*vehicle*), para 33,991% utilizando o arredondamento com a base 2, o que representa um aumento de

98,235% da taxa de erro. E aumentou para 33,151% utilizando o arredondamento com a base 10, representando assim um aumento de 93,336%.

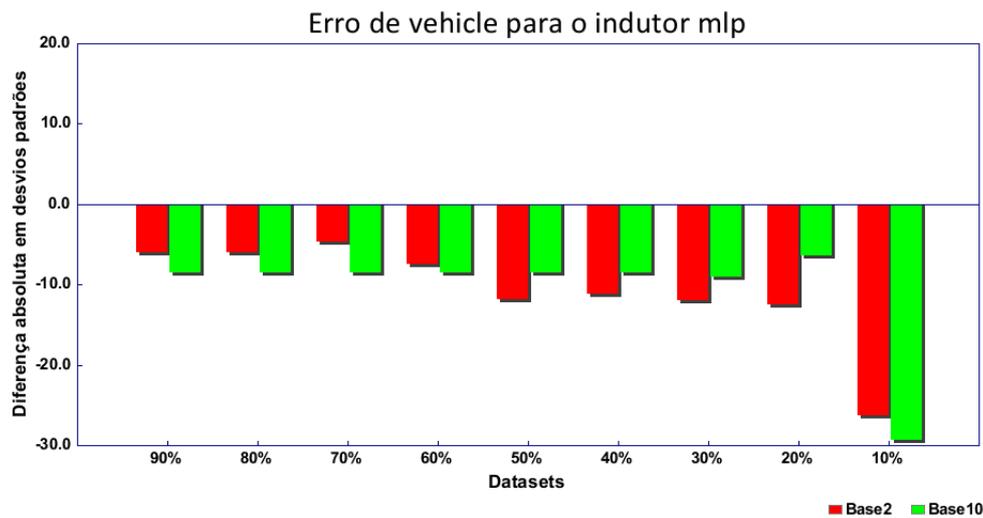


Figura 111: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vehicle

A taxa de erro aumentou para todos os conjuntos de forma significativa para ambas as bases.

#### 6.4.9 Resultados vowel

Na Tabela 41 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos vowel original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
vowel	1,000 $\pm$ 0,003	1,000 $\pm$ 0,003	8,283 $\pm$ 1,104	8,283 $\pm$ 1,104
vowel-90%	1,005 $\pm$ 0,003	1,000 $\pm$ 0,007	7,374 $\pm$ 1,167	7,778 $\pm$ 0,866
vowel-80%	1,024 $\pm$ 0,031	1,031 $\pm$ 0,036	8,384 $\pm$ 1,167	7,778 $\pm$ 0,866
vowel-70%	1,006 $\pm$ 0,004	1,001 $\pm$ 0,007	7,475 $\pm$ 1,129	7,778 $\pm$ 0,866
vowel-60%	1,028 $\pm$ 0,020	1,000 $\pm$ 0,007	7,576 $\pm$ 1,169	7,778 $\pm$ 0,866
vowel-50%	1,014 $\pm$ 0,020	1,002 $\pm$ 0,007	7,576 $\pm$ 1,057	7,879 $\pm$ 0,849
vowel-40%	1,009 $\pm$ 0,012	1,006 $\pm$ 0,001	7,677 $\pm$ 0,955	7,273 $\pm$ 0,780
vowel-30%	1,007 $\pm$ 0,001	0,999 $\pm$ 0,005	8,384 $\pm$ 1,137	7,071 $\pm$ 1,107
vowel-20%	1,003 $\pm$ 0,004	0,997 $\pm$ 0,005	8,485 $\pm$ 0,881	7,273 $\pm$ 1,084
vowel-10%	1,003 $\pm$ 0,009	0,997 $\pm$ 0,005	8,889 $\pm$ 1,052	7,273 $\pm$ 1,084

Tabela 41: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 vowel

Na Figura 112 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre vowel e vowel-90%, entre vowel e vowel-80% e assim por diante, utilizando base 2 e base 10.

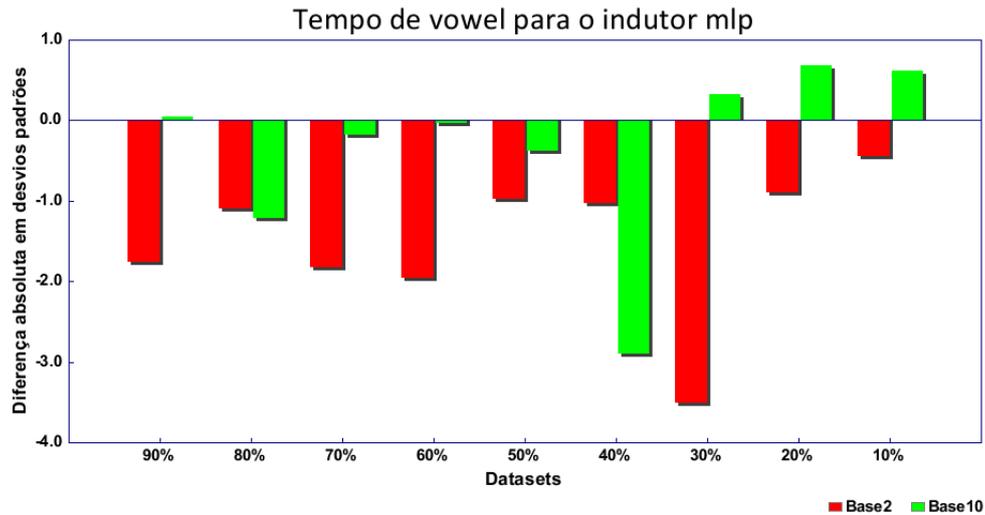


Figura 112: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

No geral, o tempo de indução reduziu para todos os conjuntos, sendo de forma significativa para vowel-30%, utilizando arredondamento com base 2 e para vowel-40%, utilizando arredondamento com a base 10.

Analogamente, na Figura 113 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 8,283% (vowel), para 7,980% utilizando o arredondamento com a base 2, o que representa uma redução de 3,658% da taxa de erro. E reduziu para 7,542% utilizando o arredondamento com a base 10, representando assim uma redução de 8,942%.

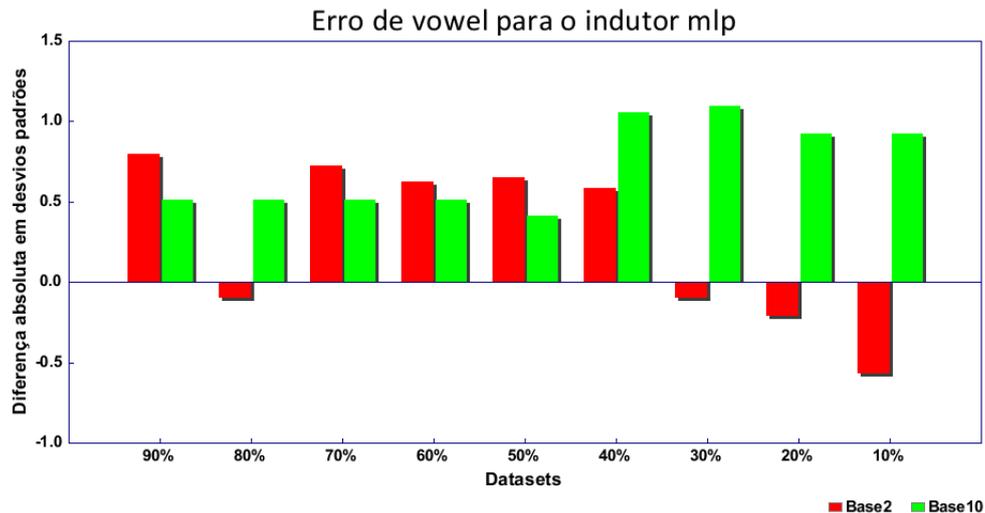


Figura 113: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

No geral, a taxa de erro reduziu para a maioria dos conjuntos, porém de forma não significativa.

#### 6.4.10 Resultados wine

Na Tabela 42 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos wine original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
wine	1,000 $\pm$ 0,018	1,000 $\pm$ 0,018	1,667 $\pm$ 0,849	1,667 $\pm$ 0,849
wine-90%	0,987 $\pm$ 0,005	0,987 $\pm$ 0,006	2,222 $\pm$ 1,228	2,222 $\pm$ 1,228
wine-80%	0,986 $\pm$ 0,004	0,987 $\pm$ 0,007	1,667 $\pm$ 0,849	2,222 $\pm$ 1,228
wine-70%	0,986 $\pm$ 0,004	0,987 $\pm$ 0,008	2,222 $\pm$ 1,228	2,222 $\pm$ 1,228
wine-60%	0,985 $\pm$ 0,005	0,991 $\pm$ 0,018	2,222 $\pm$ 1,228	2,222 $\pm$ 1,228
wine-50%	0,987 $\pm$ 0,007	0,990 $\pm$ 0,017	2,222 $\pm$ 1,228	2,222 $\pm$ 1,228
wine-40%	0,986 $\pm$ 0,006	0,991 $\pm$ 0,018	2,222 $\pm$ 1,228	2,222 $\pm$ 1,228
wine-30%	0,984 $\pm$ 0,004	1,022 $\pm$ 0,026	2,222 $\pm$ 1,228	1,667 $\pm$ 0,849
wine-20%	0,984 $\pm$ 0,004	1,003 $\pm$ 0,008	2,222 $\pm$ 1,228	2,222 $\pm$ 0,907
wine-10%	0,996 $\pm$ 0,019	0,988 $\pm$ 0,008	3,333 $\pm$ 1,228	2,810 $\pm$ 1,251

Tabela 42: Tempo de indução e taxa de erro do classificador para o indutor MLP utilizando arredondamento com bases 2 e 10 wine

Na Figura 114 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre wine e wine-90%, entre wine e wine-80% e assim por diante, utilizando base 2 e base 10.

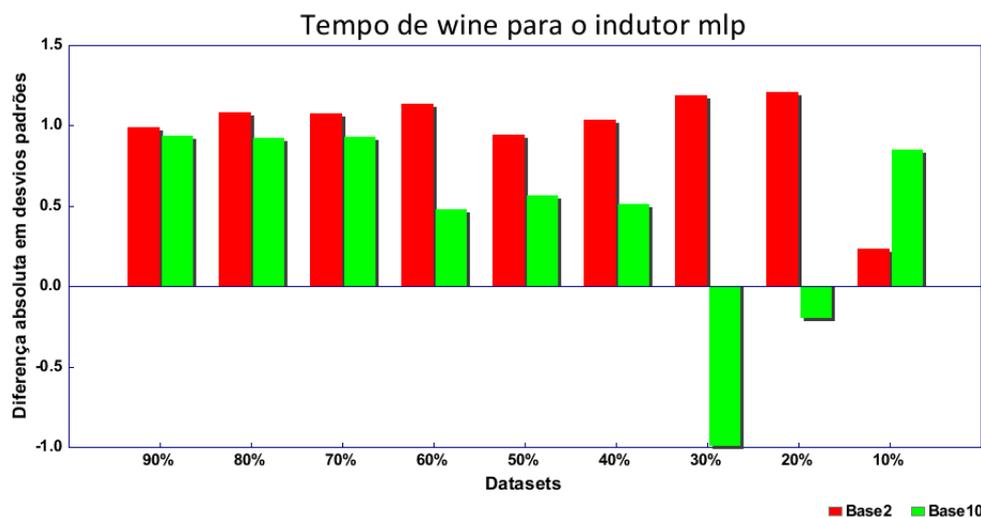


Figura 114: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

O tempo de indução reduziu para a maioria dos conjuntos utilizando arredondamento, sendo de forma não significativa, tanto para base 2 como para a base 10.

Analogamente, na Figura 115 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 1,667% (wine), para 2,284% utilizando o arredondamento com a base 2, o que representa uma redução de 36,999% da taxa de erro. E aumentou para 2,226% utilizando o arredondamento com a base 10, representando assim um aumento de 33,513%.

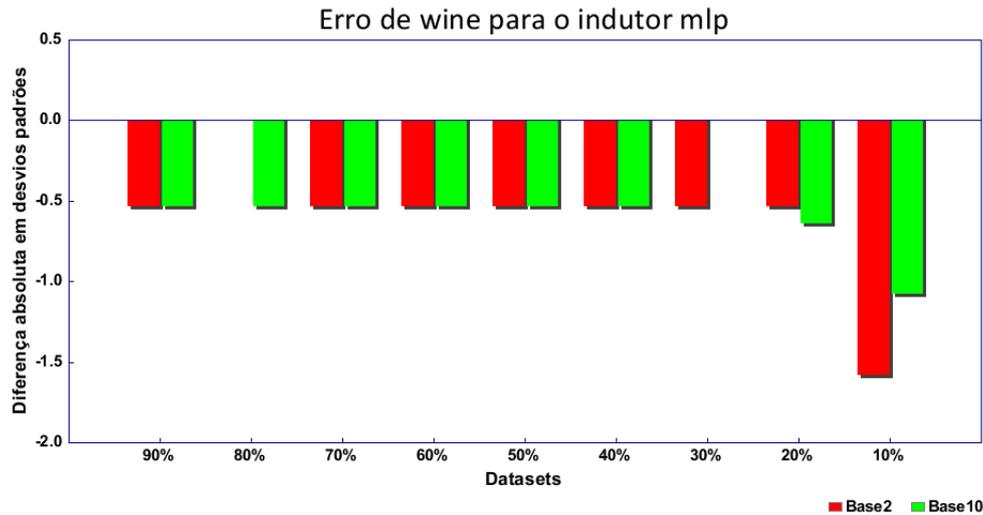


Figura 115: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

No geral, a taxa de erro aumentou para a maioria dos conjuntos de forma não significativa, para o arredondamento com ambas as bases.

#### 6.4.11 Discussão

Nas Figuras 116, 117, 118 e 119 é mostrado um resumo dos resultados obtidos por meio do indutor MLP do tempo de indução utilizando base 2, tempo de indução utilizando base 10, taxa de erro utilizando base 2, taxa de erro utilizando base 10, respectivamente. No total são 81 conjuntos (9 conjuntos arredondados  $\times$  9 conjuntos de exemplos). No eixo  $y$  do gráfico é mostrado o número conjuntos que ficaram em cada uma das categorias mostradas no eixo  $x$  do gráfico, que são:

- aqueles que tiveram o tempo significativamente maior para os conjuntos derivados em relação ao conjunto original (barras maior que -2);
- os que tiveram o tempo maior para os conjuntos derivados (barras variando entre -2 e 0);
- os que tiveram o tempo igual para os conjuntos derivados e o conjunto original (barras igual 0);
- os que tiveram o tempo menor para os conjuntos derivados (barras variando entre 0 e 2);
- os que tiveram o tempo significativamente menor para os conjuntos derivados em relação ao conjunto original (barras maior que 2);

Analogamente para a taxa de erro.

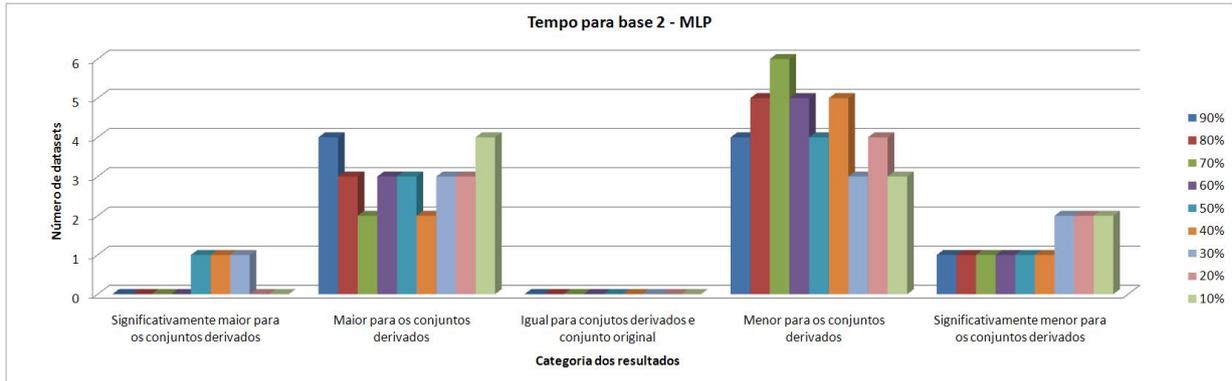


Figura 116: Resumo dos resultados do tempo de indução utilizando base 2 de MLP

Como pode ser observado, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma não significativa.

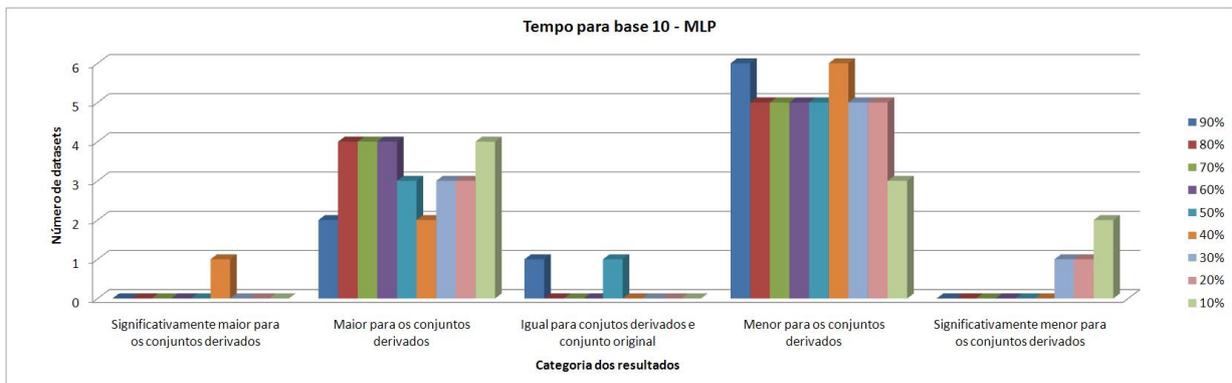


Figura 117: Resumo dos resultados do tempo de indução utilizando base 10 de MLP

Como pode ser notado, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma não significativa.

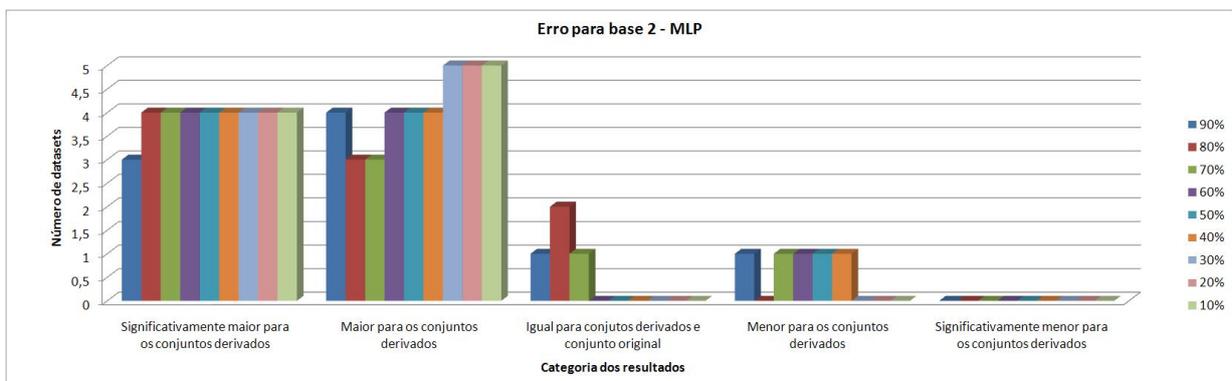


Figura 118: Resumo dos resultados da taxa de erro utilizando base 2 de MLP

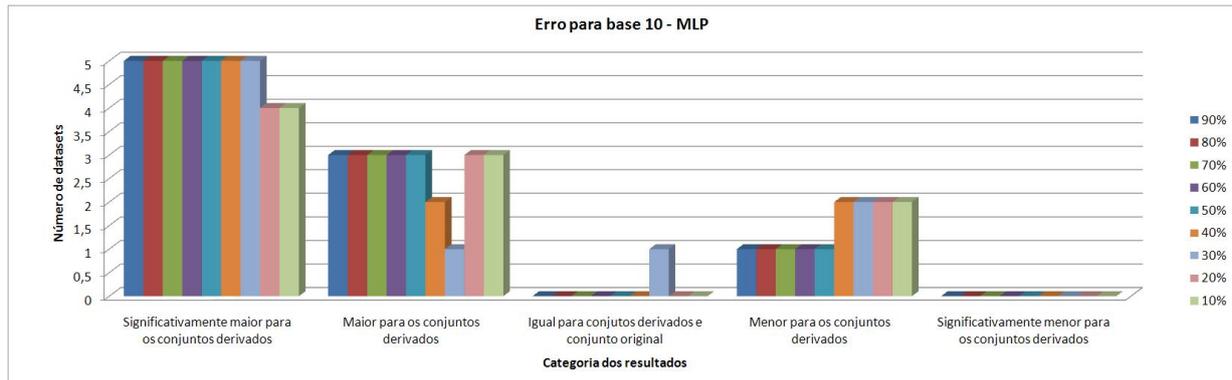


Figura 119: Resumo dos resultados da taxa de erro utilizando base 10 de MLP

Como pode ser visto, houve uma tendência da taxa de erro aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma significativa, tanto para os conjuntos arredondados com a base 2 quanto para os de base 10.

## 6.5 Resultados do indutor Naïve Bayes

### 6.5.1 Resultados aml-all-completo

Na Tabela 43 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos aml-all-completo original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
aml-all-completo	1,000 $\pm$ 0,060	1,000 $\pm$ 0,060	0,000 $\pm$ 0,000	0,000 $\pm$ 0,000
aml-all-completo-90%	0,743 $\pm$ 0,008	0,737 $\pm$ 0,008	0,000 $\pm$ 0,000	0,000 $\pm$ 0,000
aml-all-completo-80%	0,745 $\pm$ 0,006	0,738 $\pm$ 0,008	0,000 $\pm$ 0,000	0,000 $\pm$ 0,000
aml-all-completo-70%	0,744 $\pm$ 0,011	0,732 $\pm$ 0,009	1,429 $\pm$ 1,429	0,000 $\pm$ 0,000
aml-all-completo-60%	0,740 $\pm$ 0,010	0,726 $\pm$ 0,008	1,429 $\pm$ 1,429	1,429 $\pm$ 1,429
aml-all-completo-50%	0,743 $\pm$ 0,014	0,718 $\pm$ 0,008	1,429 $\pm$ 1,429	1,429 $\pm$ 1,429
aml-all-completo-40%	0,730 $\pm$ 0,008	0,709 $\pm$ 0,008	1,429 $\pm$ 1,429	1,429 $\pm$ 1,429
aml-all-completo-30%	0,720 $\pm$ 0,007	0,712 $\pm$ 0,026	1,429 $\pm$ 1,429	1,429 $\pm$ 1,429
aml-all-completo-20%	0,719 $\pm$ 0,014	0,702 $\pm$ 0,006	1,429 $\pm$ 1,429	1,429 $\pm$ 1,429
aml-all-completo-10%	0,695 $\pm$ 0,005	0,706 $\pm$ 0,015	15,179 $\pm$ 4,400	4,286 $\pm$ 3,049

Tabela 43: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 aml-all-completo

Na Figura 120 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre aml-all-completo e aml-all-completo-90%, entre aml-all-completo e aml-all-completo-80% e assim por diante, utilizando base 2 e base 10.

O tempo de indução reduziu de forma significativa para todos os conjuntos utilizando arredondamento com as base 2 e 10.

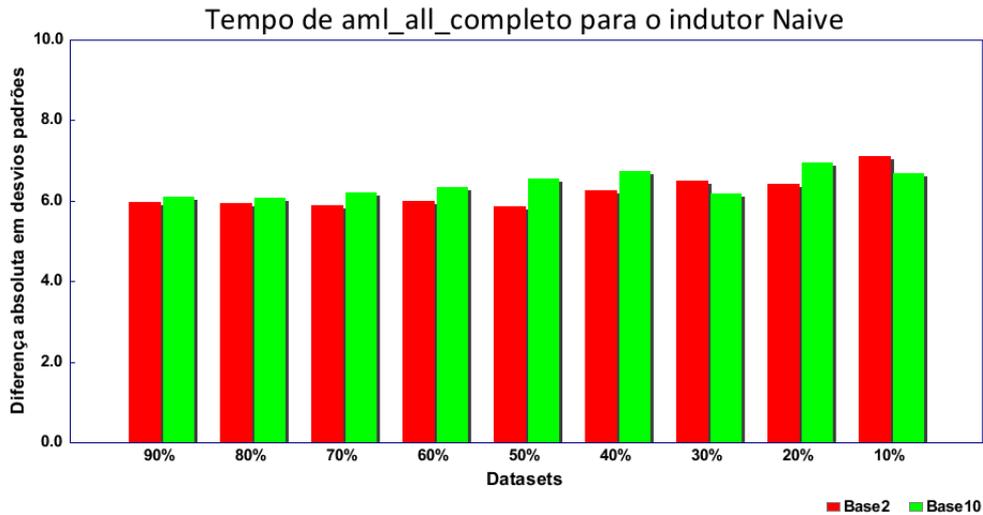


Figura 120: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) aml-all-completo

Analogamente, na Figura 121 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 9,643% (aml-all-completo), para 13,075% utilizando o arredondamento com a base 2, o que representa um aumento de 35,595% da taxa de erro. E aumentou para 11,349% utilizando o arredondamento com a base 10, representando assim um aumento de 17,693%.

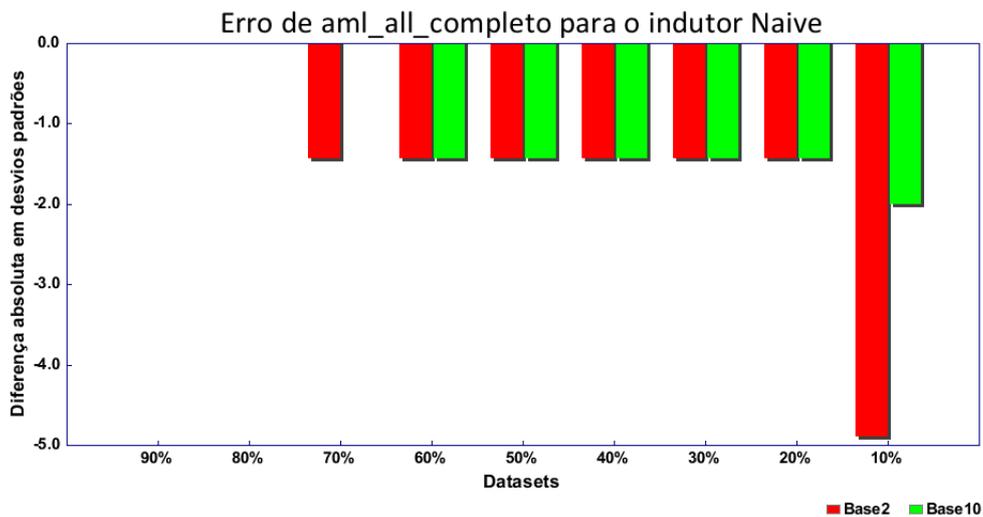


Figura 121: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) aml-all-completo

Como pode ser visto, a taxa de erro permaneceu constante ou aumentou para todos os conjuntos de forma não significativa, exceto para aml-all-completo-10% que a taxa de erro aumentou significativamente utilizando arredondamento tanto para a base 2 quanto para a base 10.

### 6.5.2 Resultados eucalyptus

Na Tabela 44 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *eucalyptus* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
<i>eucalyptus</i>	1,000 $\pm$ 0,106	1,000 $\pm$ 0,106	45,365 $\pm$ 2,084	45,365 $\pm$ 2,084
<i>eucalyptus</i> -90%	0,985 $\pm$ 0,080	0,985 $\pm$ 0,080	44,961 $\pm$ 1,819	47,819 $\pm$ 1,421
<i>eucalyptus</i> -80%	0,955 $\pm$ 0,073	1,030 $\pm$ 0,064	46,046 $\pm$ 1,656	47,819 $\pm$ 1,421
<i>eucalyptus</i> -70%	0,939 $\pm$ 0,096	1,045 $\pm$ 0,086	45,233 $\pm$ 1,405	47,819 $\pm$ 1,421
<i>eucalyptus</i> -60%	0,955 $\pm$ 0,073	1,015 $\pm$ 0,073	45,098 $\pm$ 1,148	47,001 $\pm$ 1,631
<i>eucalyptus</i> -50%	0,924 $\pm$ 0,048	1,106 $\pm$ 0,073	44,421 $\pm$ 1,485	46,872 $\pm$ 1,662
<i>eucalyptus</i> -40%	1,015 $\pm$ 0,073	0,955 $\pm$ 0,073	44,424 $\pm$ 1,356	47,281 $\pm$ 1,723
<i>eucalyptus</i> -30%	0,970 $\pm$ 0,078	0,924 $\pm$ 0,048	46,042 $\pm$ 2,036	46,059 $\pm$ 1,597
<i>eucalyptus</i> -20%	0,939 $\pm$ 0,096	0,879 $\pm$ 0,096	48,375 $\pm$ 1,000	48,240 $\pm$ 1,652
<i>eucalyptus</i> -10%	0,864 $\pm$ 0,102	0,909 $\pm$ 0,071	75,542 $\pm$ 0,054	52,710 $\pm$ 1,940

Tabela 44: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 *eucalyptus*

Na Figura 122 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *eucalyptus* e *eucalyptus*-90%, entre *eucalyptus* e *eucalyptus*-80% e assim por diante, utilizando base 2 e base 10.

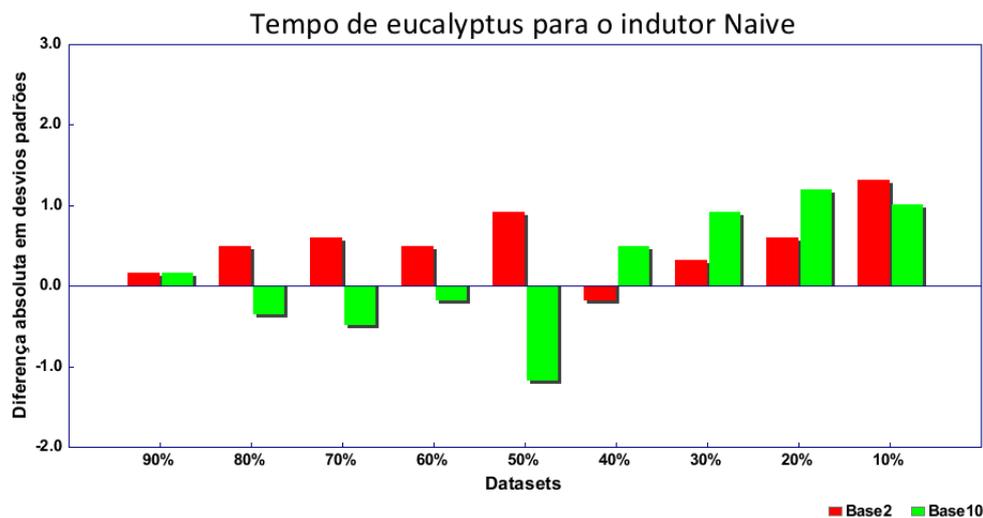


Figura 122: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *eucalyptus*

Para os conjuntos arredondados com a base 2, o tempo de indução aumentou para a maioria dos conjuntos, embora de forma não significativa. Para aqueles arredondados com a base 10, o tempo de indução não manteve um padrão, aumentado para alguns e reduzindo para outros, sendo de forma não significativa.

Analogamente, na Figura 123 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 45,365% (*eucalyptus*), para 48,905% utilizando o arredondamento com a base 2, o que representa um aumento de

7,803% da taxa de erro. E aumentou para 47,958% utilizando o arredondamento com a base 10, representando assim um aumento de 5,715%.

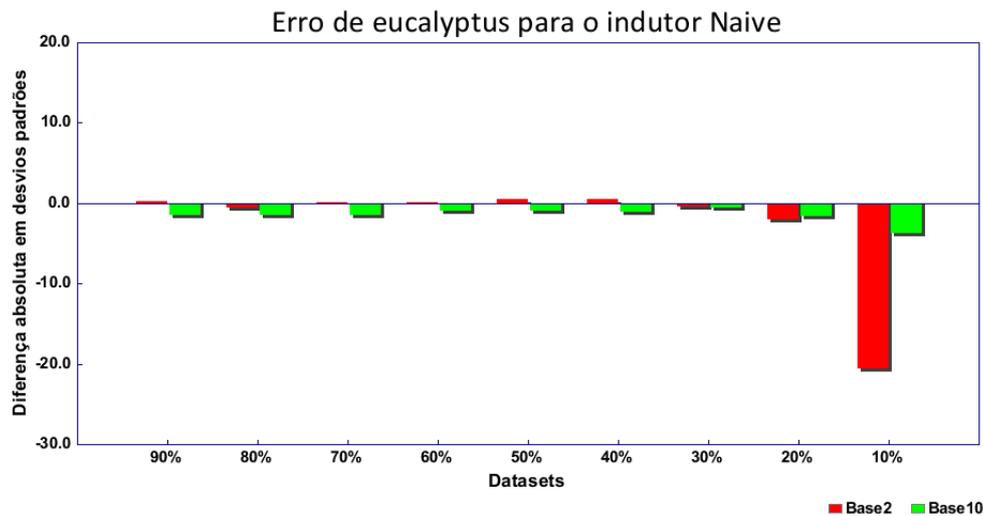


Figura 123: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) eucalyptus

Como pode ser notado, no geral, a taxa de erro aumentou para todos os conjuntos de forma não significativa, exceto para *eucalyptus-10%* que a taxa de erro aumentou significativamente utilizando arredondamento com ambas as bases.

### 6.5.3 Resultados hepatitis

Na Tabela 45 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *hepatitis* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
<i>hepatitis</i>	1,000 $\pm$ 0,194	1,000 $\pm$ 0,194	15,458 $\pm$ 3,686	15,458 $\pm$ 3,686
<i>hepatitis-90%</i>	1,611 $\pm$ 1,587	1,028 $\pm$ 0,187	17,417 $\pm$ 4,168	14,792 $\pm$ 3,995
<i>hepatitis-80%</i>	1,083 $\pm$ 0,423	1,056 $\pm$ 0,117	15,458 $\pm$ 3,802	14,792 $\pm$ 3,995
<i>hepatitis-70%</i>	1,083 $\pm$ 0,158	1,111 $\pm$ 0,227	15,458 $\pm$ 3,802	14,792 $\pm$ 3,995
<i>hepatitis-60%</i>	0,917 $\pm$ 0,134	1,278 $\pm$ 0,527	15,458 $\pm$ 3,802	14,792 $\pm$ 3,995
<i>hepatitis-50%</i>	1,028 $\pm$ 0,134	1,028 $\pm$ 0,134	15,458 $\pm$ 3,686	14,792 $\pm$ 3,995
<i>hepatitis-40%</i>	1,167 $\pm$ 0,176	1,194 $\pm$ 0,187	15,417 $\pm$ 3,781	14,792 $\pm$ 3,995
<i>hepatitis-30%</i>	1,167 $\pm$ 0,117	1,056 $\pm$ 0,176	14,792 $\pm$ 3,772	15,417 $\pm$ 3,781
<i>hepatitis-20%</i>	1,111 $\pm$ 0,185	1,028 $\pm$ 0,134	15,458 $\pm$ 3,786	17,375 $\pm$ 3,762
<i>hepatitis-10%</i>	1,222 $\pm$ 0,268	1,139 $\pm$ 0,158	20,625 $\pm$ 2,294	17,375 $\pm$ 3,762

Tabela 45: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 *hepatitis*

Na Figura 124 é mostrada a diferença absoluta em desvios padrões do tempo de indução

entre o conjunto original e os conjuntos derivados, ou seja, entre hepatitis e hepatitis-90%, entre hepatitis e hepatitis-80% e assim por diante, utilizando base 2 e base 10.

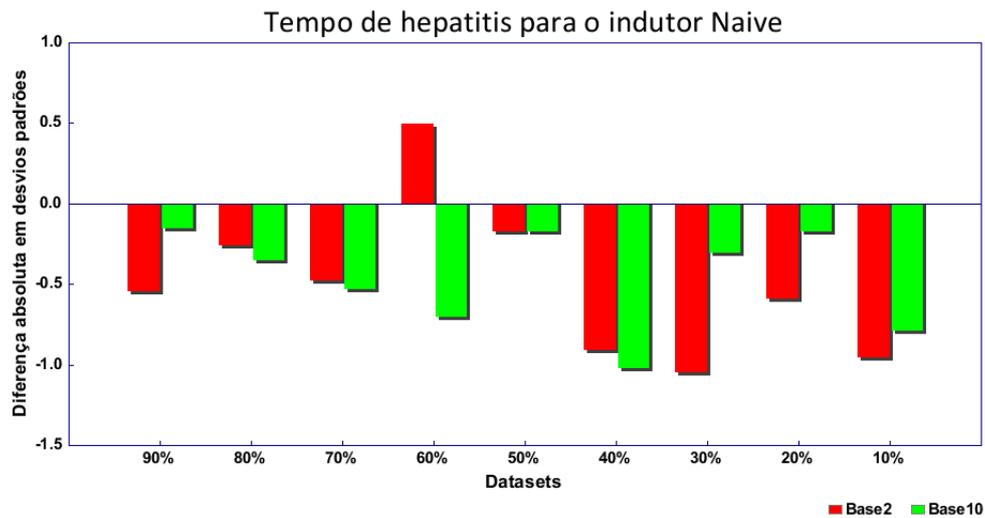


Figura 124: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

Na maioria dos conjuntos o tempo de indução aumentou, sendo de forma não significativa, tanto para base 2 como para a base 10.

Analogamente, na Figura 125 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 15,458% (hepatitis), para 16,171% utilizando o arredondamento com a base 2, o que representa um aumento de 4,614% da taxa de erro. E reduziu para 15,435% utilizando o arredondamento com a base 10, representando assim uma redução de 0,146%.

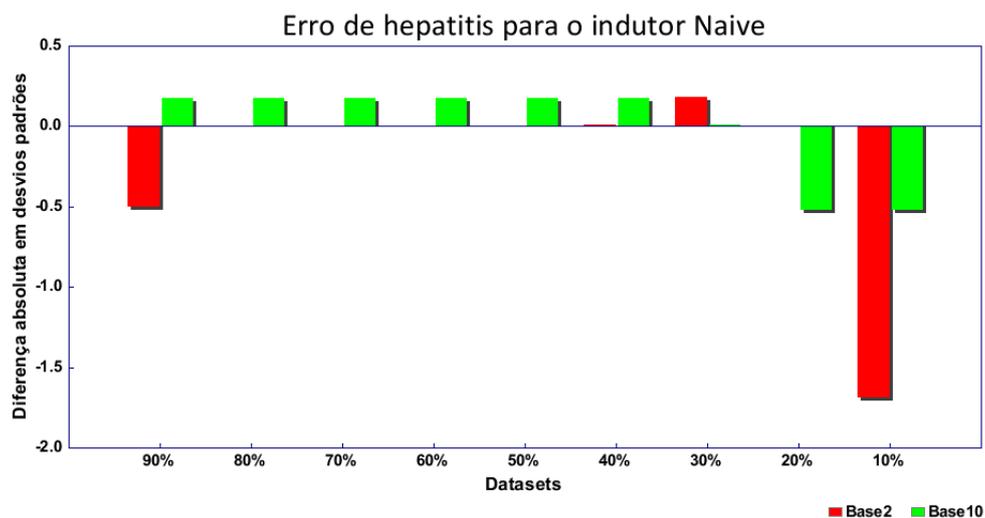


Figura 125: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) hepatitis

Para os conjuntos arredondados com a base 2, a taxa de erro não manteve um padrão, aumentado para alguns e reduzindo para outros, sendo de forma não significativa. Para aqueles

arredondados com a base 10, a taxa de erro reduziu para a maioria dos conjuntos, embora de forma não significativa.

#### 6.5.4 Resultados ionosphere

Na Tabela 46 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *ionosphere* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
<i>ionosphere</i>	1,000 $\pm$ 0,092	1,000 $\pm$ 0,092	16,810 $\pm$ 2,388	16,810 $\pm$ 2,388
<i>ionosphere-90%</i>	0,929 $\pm$ 0,075	1,000 $\pm$ 0,092	17,381 $\pm$ 2,642	17,381 $\pm$ 2,642
<i>ionosphere-80%</i>	0,946 $\pm$ 0,086	0,964 $\pm$ 0,092	17,381 $\pm$ 2,642	17,381 $\pm$ 2,642
<i>ionosphere-70%</i>	0,929 $\pm$ 0,075	1,036 $\pm$ 0,075	16,810 $\pm$ 2,388	17,381 $\pm$ 2,642
<i>ionosphere-60%</i>	1,000 $\pm$ 0,092	0,964 $\pm$ 0,092	16,810 $\pm$ 2,388	17,381 $\pm$ 2,642
<i>ionosphere-50%</i>	1,018 $\pm$ 0,086	0,964 $\pm$ 0,092	17,095 $\pm$ 2,447	16,810 $\pm$ 2,388
<i>ionosphere-40%</i>	1,000 $\pm$ 0,092	0,946 $\pm$ 0,086	16,810 $\pm$ 2,388	17,381 $\pm$ 2,537
<i>ionosphere-30%</i>	0,964 $\pm$ 0,092	0,946 $\pm$ 0,086	16,810 $\pm$ 2,388	17,381 $\pm$ 2,537
<i>ionosphere-20%</i>	0,946 $\pm$ 0,086	0,982 $\pm$ 0,094	16,817 $\pm$ 2,468	17,381 $\pm$ 2,537
<i>ionosphere-10%</i>	1,054 $\pm$ 0,449	0,929 $\pm$ 0,075	16,238 $\pm$ 2,591	17,381 $\pm$ 2,537

Tabela 46: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 *ionosphere*

Na Figura 126 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *ionosphere* e *ionosphere-90%*, entre *ionosphere* e *ionosphere-80%* e assim por diante, utilizando base 2 e base 10.

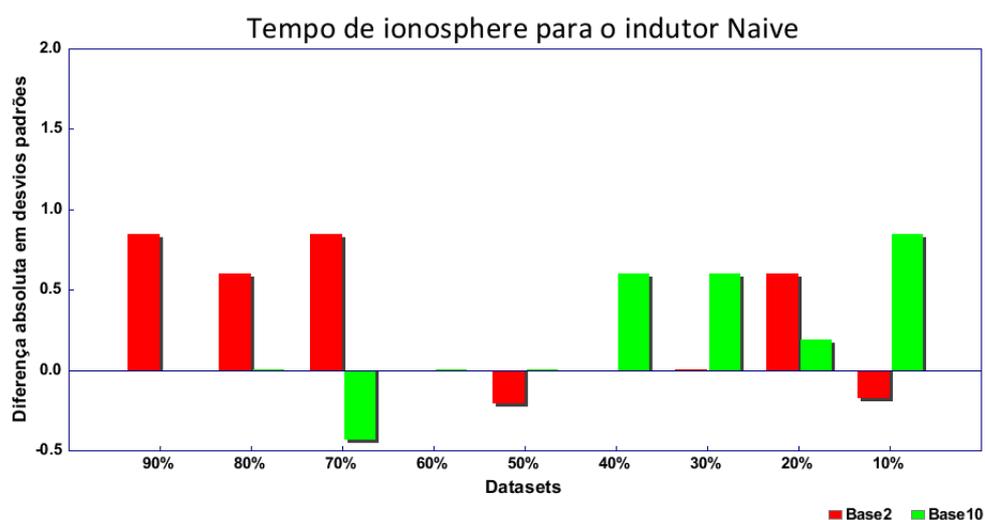


Figura 126: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) *ionosphere*

O tempo de indução aumentou para a maioria dos conjuntos, sendo de forma não significativa, utilizando arredondamento para ambas as bases.

Analogamente, na Figura 127 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 16,810% (ionosphere), para 16,906% utilizando o arredondamento com a base 2, o que representa um aumento de 0,570% da taxa de erro. E aumentou para 17,318% utilizando o arredondamento com a base 10, representando assim um aumento de 3,019%.

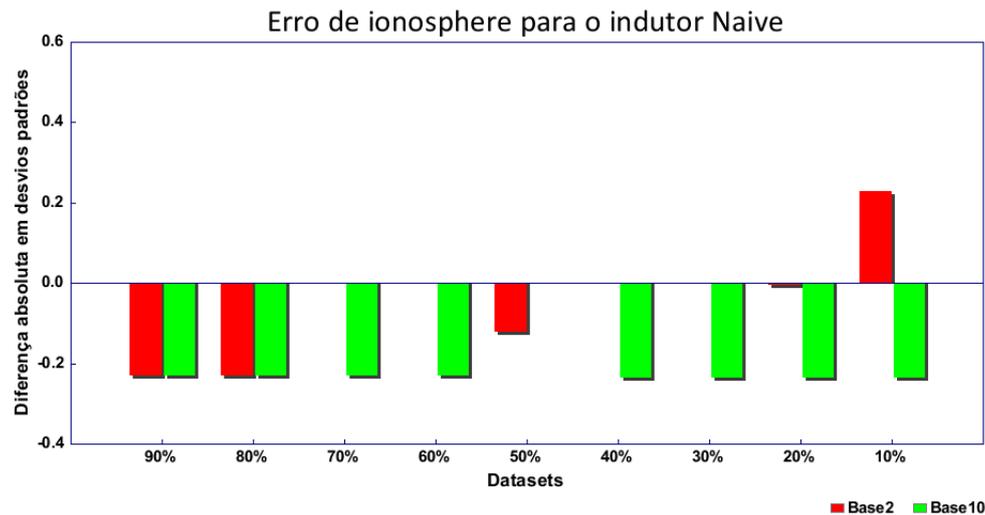


Figura 127: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) ionosphere

Como pode ser observado, a taxa de erro aumentou para a maioria dos conjuntos de forma não significativa, utilizando arredondamento com as bases 2 e 10.

### 6.5.5 Resultados lymph

Na Tabela 47 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos lymph original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Na Figura 128 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre lymph e lymph-90%, entre lymph e lymph-80% e assim por diante, utilizando base 2 e base 10.

O tempo de indução aumentou para todos os conjuntos utilizando arredondamento com ambas as bases, sendo de forma significativa apenas para lymph-10%, utilizando arredondamento com base 10.

Analogamente, na Figura 129 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 16,238% (lymph), para 20,005% utilizando o arredondamento com a base 2, o que representa um aumento de 23,200% da taxa de erro. E reduziu para 15,386% utilizando o arredondamento com a base 10, representando assim uma redução de 5,244%.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
lymph	$1,000 \pm 0,198$	$1,000 \pm 0,198$	$16,238 \pm 2,739$	$16,238 \pm 2,739$
lymph-90%	$0,938 \pm 0,000$	$0,969 \pm 0,099$	$17,571 \pm 2,726$	$15,524 \pm 2,910$
lymph-80%	$0,938 \pm 0,000$	$0,906 \pm 0,099$	$17,571 \pm 2,726$	$15,524 \pm 2,910$
lymph-70%	$0,938 \pm 0,000$	$0,938 \pm 0,000$	$17,571 \pm 2,726$	$15,524 \pm 2,910$
lymph-60%	$0,938 \pm 0,000$	$0,938 \pm 0,000$	$14,905 \pm 2,857$	$15,524 \pm 2,910$
lymph-50%	$0,969 \pm 0,099$	$0,938 \pm 0,000$	$14,905 \pm 2,857$	$15,524 \pm 2,910$
lymph-40%	$0,938 \pm 0,000$	$0,938 \pm 0,147$	$33,857 \pm 3,273$	$15,524 \pm 2,910$
lymph-30%	$0,813 \pm 0,161$	$0,875 \pm 0,132$	$33,857 \pm 3,273$	$15,524 \pm 2,910$
lymph-20%	$0,906 \pm 0,099$	$0,750 \pm 0,161$	$14,905 \pm 2,215$	$14,905 \pm 2,215$
lymph-10%	$0,781 \pm 0,165$	$0,656 \pm 0,099$	$14,905 \pm 2,215$	$14,905 \pm 2,215$

Tabela 47: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 lymph

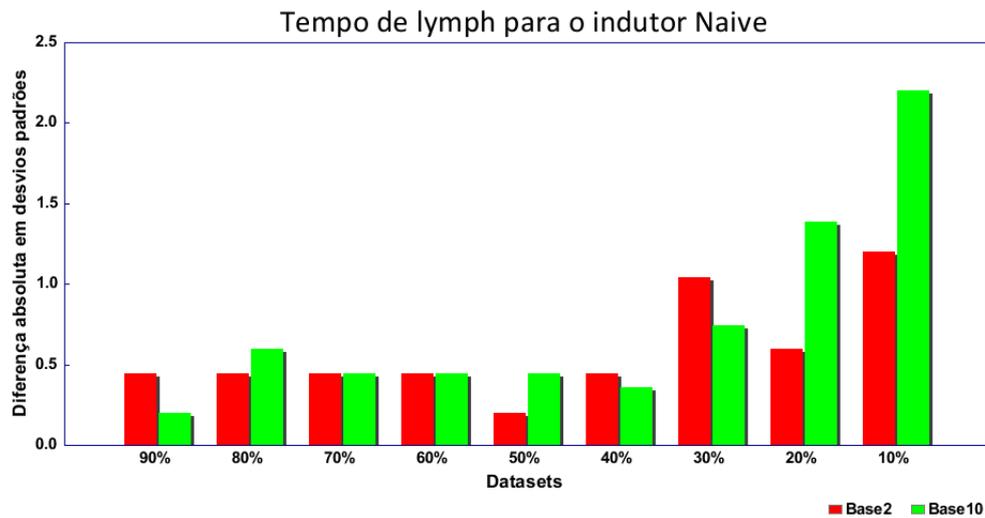


Figura 128: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

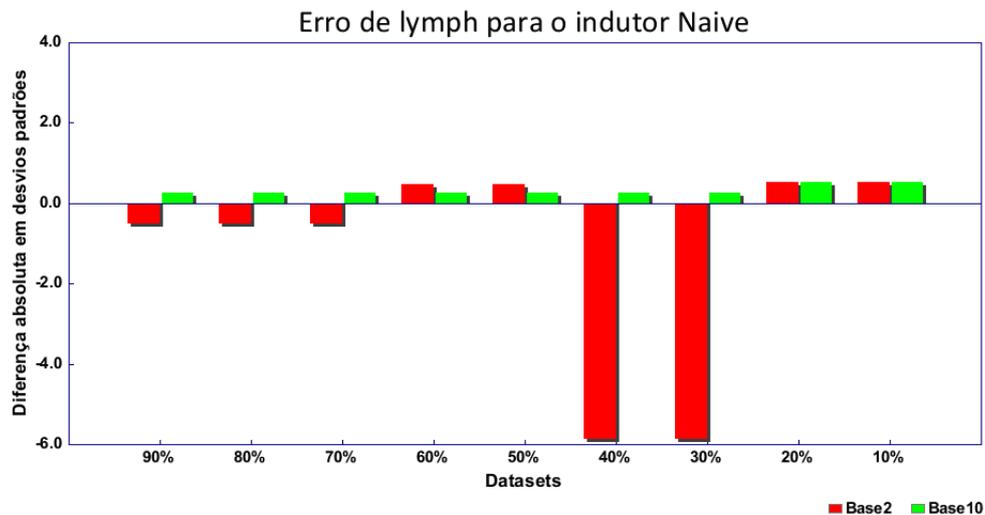


Figura 129: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) lymph

Como pode ser notado, para o arredondamento utilizando a base 2, a taxa de erro aumentou para todos os conjuntos, sendo de forma significativa apenas para *lymph-40%* e *lymph-30%*. Para o arredondamento utilizando a base 10, todos os conjuntos tiveram uma redução na taxa de erro, embora de forma não significativa.

### 6.5.6 Resultados sick

Na Tabela 48 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos *sick* original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
<i>sick</i>	1,000 $\pm$ 0,038	1,000 $\pm$ 0,038	7,158 $\pm$ 0,439	7,158 $\pm$ 0,439
<i>sick-90%</i>	1,291 $\pm$ 0,041	1,016 $\pm$ 0,025	9,755 $\pm$ 0,714	10,790 $\pm$ 0,806
<i>sick-80%</i>	1,323 $\pm$ 0,033	1,047 $\pm$ 0,038	10,603 $\pm$ 0,695	10,790 $\pm$ 0,806
<i>sick-70%</i>	1,299 $\pm$ 0,041	1,024 $\pm$ 0,000	10,815 $\pm$ 0,627	10,790 $\pm$ 0,806
<i>sick-60%</i>	1,291 $\pm$ 0,041	1,024 $\pm$ 0,000	10,948 $\pm$ 0,584	10,710 $\pm$ 0,783
<i>sick-50%</i>	1,291 $\pm$ 0,041	1,024 $\pm$ 0,037	10,736 $\pm$ 0,569	10,710 $\pm$ 0,783
<i>sick-40%</i>	1,276 $\pm$ 0,033	1,024 $\pm$ 0,037	14,050 $\pm$ 0,584	10,710 $\pm$ 0,783
<i>sick-30%</i>	1,276 $\pm$ 0,033	1,047 $\pm$ 0,112	14,342 $\pm$ 0,583	10,524 $\pm$ 0,777
<i>sick-20%</i>	1,276 $\pm$ 0,033	1,000 $\pm$ 0,038	15,800 $\pm$ 0,555	10,524 $\pm$ 0,777
<i>sick-10%</i>	1,260 $\pm$ 0,000	0,961 $\pm$ 0,033	7,264 $\pm$ 0,481	13,203 $\pm$ 0,880

Tabela 48: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 *sick*

Na Figura 130 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre *sick* e *sick-90%*, entre *sick* e *sick-80%* e assim por diante, utilizando base 2 e base 10.

Para o arredondamento utilizando a base 2, o tempo de indução aumentou para todos os conjuntos de forma significativa. Já para a base 10, a maioria dos conjunto também tiveram um aumento no tempo de indução porém de forma não significativa.

Analogamente, na Figura 131 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 7,158% (*sick*), para 11,590% utilizando o arredondamento com a base 2, o que representa um aumento de 61,921% da taxa de erro. E aumentou para 10,972% utilizando o arredondamento com a base 10, representando assim um aumento de 53,288%.

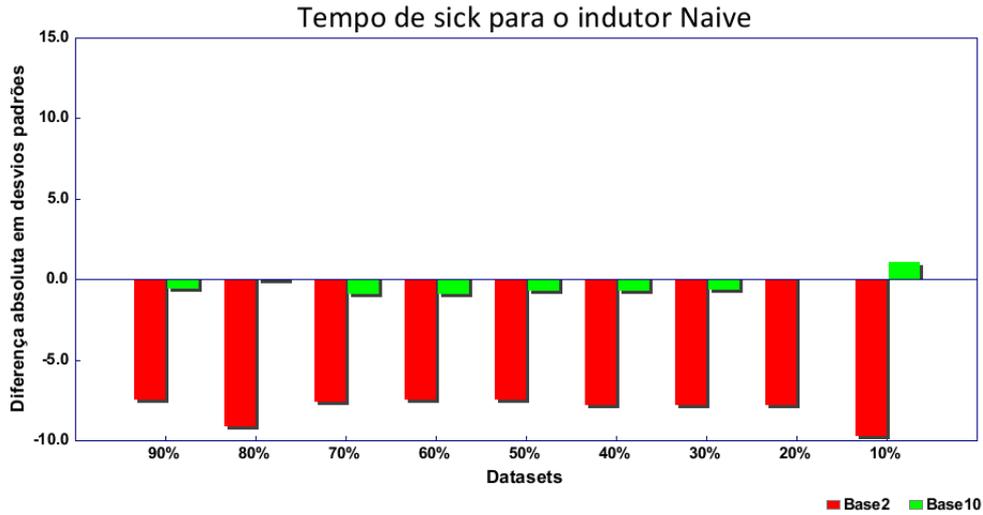


Figura 130: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

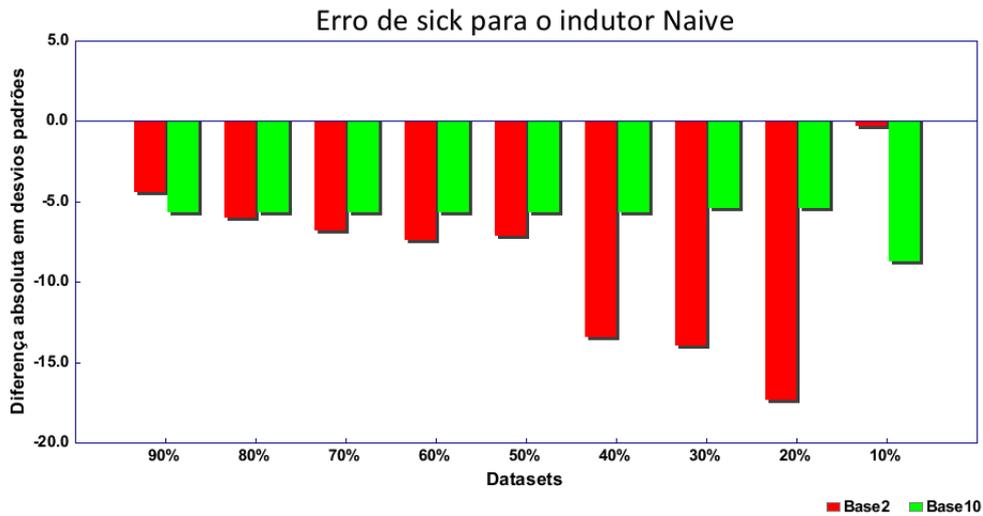


Figura 131: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sick

A taxa de erro aumentou para todos os conjunto, de forma significativa para o arredondamento com ambas as bases, exceto para o conjunto sick-10%, que não aumentou significativamente para o arredondamento com a base 2.

### 6.5.7 Resultados sonar

Na Tabela 49 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos sonar original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária

e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
sonar	$1,000 \pm 0,081$	$1,000 \pm 0,081$	$30,810 \pm 2,938$	$30,810 \pm 2,938$
sonar-90%	$0,938 \pm 0,049$	$0,954 \pm 0,065$	$29,833 \pm 2,685$	$29,833 \pm 3,183$
sonar-80%	$0,969 \pm 0,074$	$0,954 \pm 0,065$	$29,833 \pm 2,685$	$29,833 \pm 3,183$
sonar-70%	$0,923 \pm 0,000$	$0,938 \pm 0,049$	$30,310 \pm 2,703$	$29,833 \pm 3,183$
sonar-60%	$0,923 \pm 0,000$	$0,985 \pm 0,079$	$30,310 \pm 2,703$	$29,833 \pm 3,183$
sonar-50%	$0,938 \pm 0,049$	$0,938 \pm 0,049$	$30,786 \pm 2,309$	$29,833 \pm 3,183$
sonar-40%	$0,954 \pm 0,065$	$0,954 \pm 0,065$	$30,310 \pm 2,509$	$29,833 \pm 3,183$
sonar-30%	$0,969 \pm 0,104$	$0,923 \pm 0,000$	$30,786 \pm 2,309$	$29,833 \pm 3,183$
sonar-20%	$0,938 \pm 0,049$	$0,954 \pm 0,065$	$31,738 \pm 2,698$	$31,286 \pm 2,910$
sonar-10%	$0,985 \pm 0,079$	$0,954 \pm 0,065$	$32,714 \pm 2,939$	$31,333 \pm 4,015$

Tabela 49: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 sonar

Na Figura 132 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre sonar e sonar-90%, entre sonar e sonar-80% e assim por diante, utilizando base 2 e base 10.

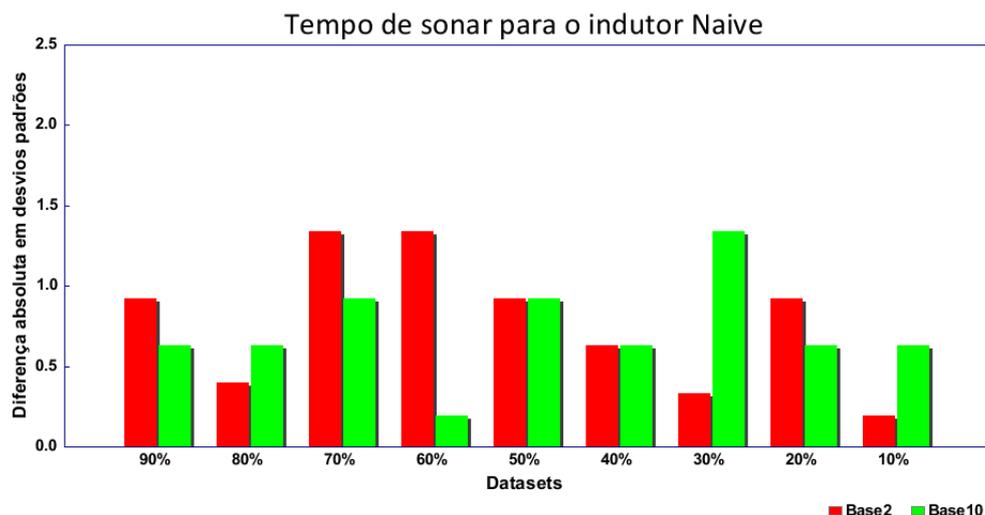


Figura 132: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

O tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo de forma não significativa tanto para base 2 como para a base 10.

Analogamente, na Figura 133 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 30,810% (sonar), para 30,736% utilizando o arredondamento com a base 2, o que representa uma redução de 0,242% da taxa de erro. E reduziu para 30,161% utilizando o arredondamento com a base 10, representando assim um aumento de 2,106%.

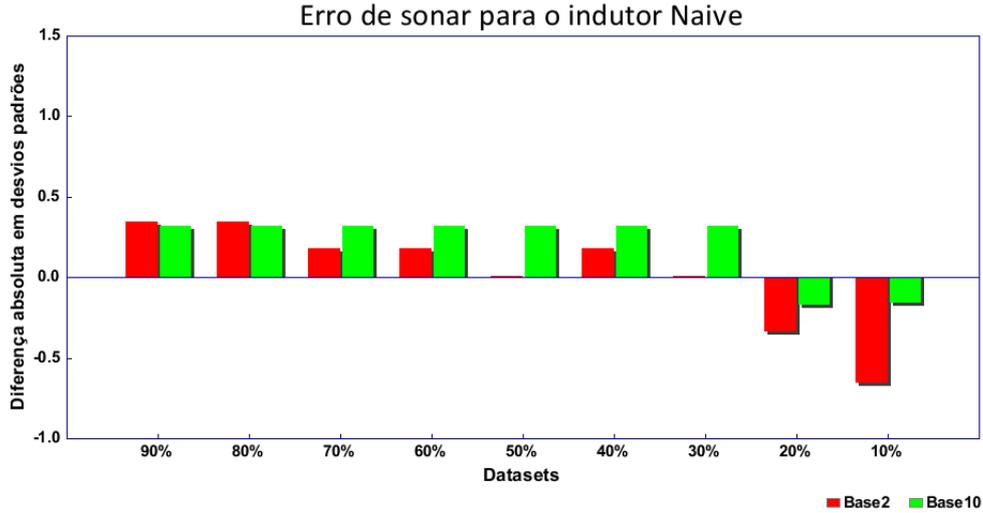


Figura 133: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) sonar

Como pode ser visto, a taxa de erro reduziu para todos os conjuntos de forma não significativa, exceto para sonar-20% e sonar-10% que a taxa de erro aumentou também de forma não significativa utilizando arredondamento com a base 2.

### 6.5.8 Resultados vehicle

Na Tabela 50 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos vehicle original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
vehicle	1,000 $\pm$ 0,000	1,000 $\pm$ 0,000	56,503 $\pm$ 1,437	56,503 $\pm$ 1,437
vehicle-90%	1,086 $\pm$ 0,074	1,000 $\pm$ 0,000	54,487 $\pm$ 1,389	54,368 $\pm$ 1,356
vehicle-80%	1,043 $\pm$ 0,069	0,971 $\pm$ 0,060	54,487 $\pm$ 1,389	54,368 $\pm$ 1,356
vehicle-70%	1,043 $\pm$ 0,069	0,943 $\pm$ 0,074	52,359 $\pm$ 1,358	54,368 $\pm$ 1,356
vehicle-60%	1,014 $\pm$ 0,045	0,957 $\pm$ 0,069	52,947 $\pm$ 1,408	54,368 $\pm$ 1,356
vehicle-50%	1,014 $\pm$ 0,045	0,986 $\pm$ 0,045	56,148 $\pm$ 1,528	54,368 $\pm$ 1,356
vehicle-40%	1,043 $\pm$ 0,069	0,957 $\pm$ 0,069	56,629 $\pm$ 1,620	54,368 $\pm$ 1,356
vehicle-30%	1,057 $\pm$ 0,074	0,943 $\pm$ 0,074	51,784 $\pm$ 2,248	55,668 $\pm$ 1,626
vehicle-20%	1,029 $\pm$ 0,060	0,957 $\pm$ 0,069	52,832 $\pm$ 1,843	57,915 $\pm$ 1,365
vehicle-10%	1,071 $\pm$ 0,075	0,957 $\pm$ 0,069	60,510 $\pm$ 1,100	62,296 $\pm$ 1,541

Tabela 50: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 vehicle

Na Figura 134 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre vehicle e vehicle-90%, entre vehicle e vehicle-80% e assim por diante, utilizando base 2 e base 10.

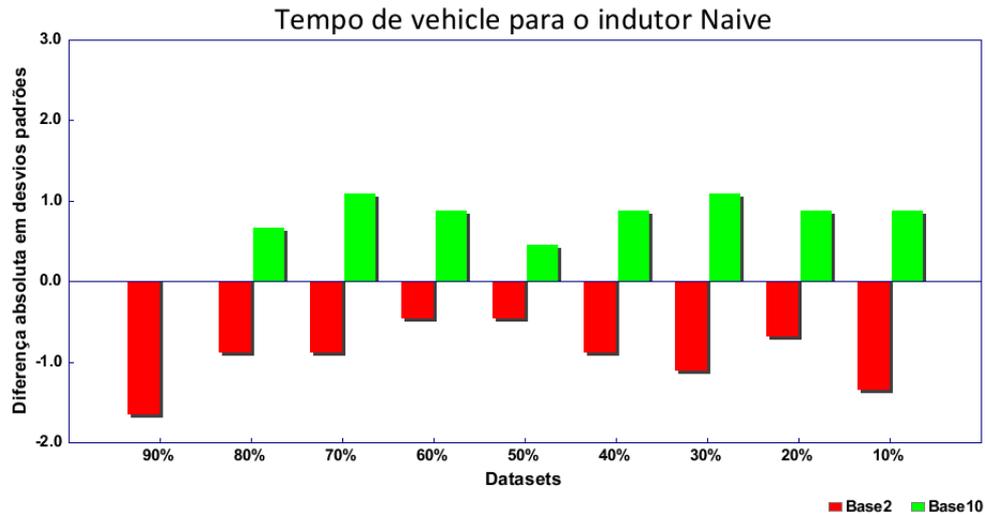


Figura 134: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vehicle

Podemos observar que o tempo de indução aumentou para todos os conjuntos utilizando arredondamento com a base 2, e reduziu para os conjuntos utilizando arredondamento com a base 10, sendo de forma não significativa para ambas as bases.

Analogamente, na Figura 135 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro reduziu de 56,503% (vehicle), para 54,687% utilizando o arredondamento com a base 2, o que representa uma redução de 3,214% da taxa de erro. E reduziu para 55,787% utilizando o arredondamento com a base 10, representando assim uma redução de 1,266%.

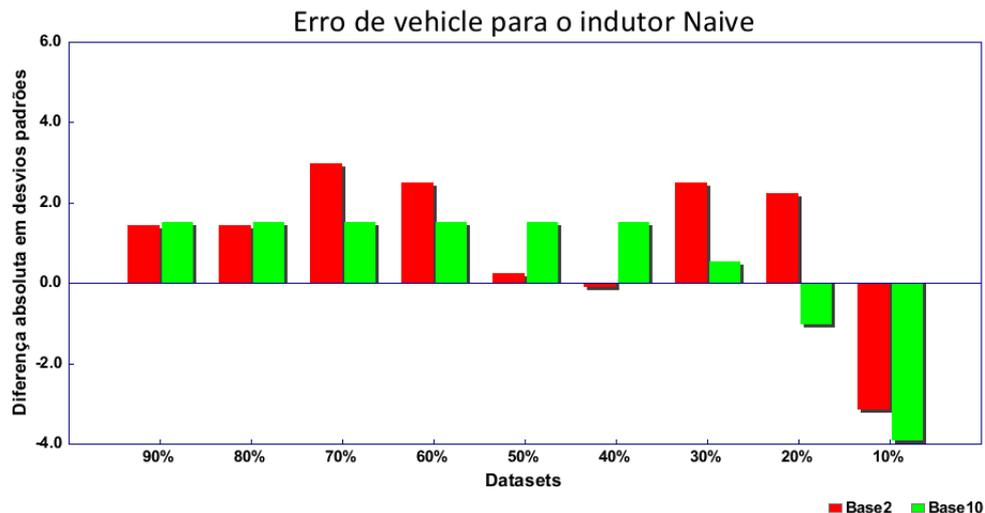


Figura 135: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vehicle

Para o arredondamento utilizando a base 2, a taxa de erro reduziu para a maioria dos conjuntos, sendo de forma significativa para vehicle-70%, vehicle-60%, vehicle-30% e vehicle-20%, e teve um aumento significativo para vehicle-10%. Já para o arredondamento utilizando a base 10, a taxa de erro também reduziu para a maioria dos conjuntos, embora de forma não

significativa, e o conjunto `vehicle-10%` teve um aumento significativo.

### 6.5.9 Resultados vowel

Na Tabela 51 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos `vowel` original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
<code>vowel</code>	1,000 $\pm$ 0,068	1,000 $\pm$ 0,068	37,172 $\pm$ 1,354	37,172 $\pm$ 1,354
<code>vowel-90%</code>	1,016 $\pm$ 0,078	0,984 $\pm$ 0,051	37,172 $\pm$ 1,328	37,273 $\pm$ 1,299
<code>vowel-80%</code>	1,032 $\pm$ 0,083	1,016 $\pm$ 0,078	37,172 $\pm$ 1,354	37,273 $\pm$ 1,299
<code>vowel-70%</code>	0,984 $\pm$ 0,051	1,000 $\pm$ 0,068	37,172 $\pm$ 1,354	37,273 $\pm$ 1,299
<code>vowel-60%</code>	1,000 $\pm$ 0,068	0,984 $\pm$ 0,051	37,273 $\pm$ 1,359	37,273 $\pm$ 1,299
<code>vowel-50%</code>	0,984 $\pm$ 0,051	0,984 $\pm$ 0,051	37,071 $\pm$ 1,339	37,273 $\pm$ 1,299
<code>vowel-40%</code>	0,984 $\pm$ 0,051	0,968 $\pm$ 0,000	37,172 $\pm$ 1,443	37,576 $\pm$ 1,655
<code>vowel-30%</code>	1,000 $\pm$ 0,068	0,984 $\pm$ 0,051	37,172 $\pm$ 1,403	37,576 $\pm$ 1,879
<code>vowel-20%</code>	0,968 $\pm$ 0,000	0,968 $\pm$ 0,000	37,778 $\pm$ 1,609	37,576 $\pm$ 1,879
<code>vowel-10%</code>	1,016 $\pm$ 0,078	0,984 $\pm$ 0,051	37,778 $\pm$ 1,552	37,576 $\pm$ 1,879

Tabela 51: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 `vowel`

Na Figura 136 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre `vowel` e `vowel-90%`, entre `vowel` e `vowel-80%` e assim por diante, utilizando base 2 e base 10.

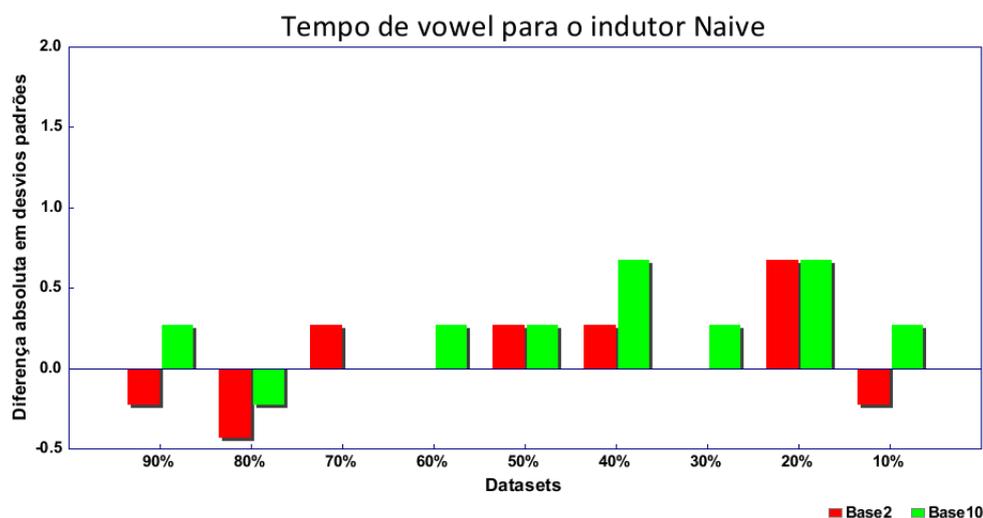


Figura 136: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) `vowel`

No geral, o tempo de indução reduziu para a maioria dos conjuntos utilizando arredondamento tanto para base 2 como para a base 10.

Analogamente, na Figura 137 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 37,172% (vowel), para 37,307% utilizando o arredondamento com a base 2, o que representa um aumento de 0,362% da taxa de erro. E aumentou para 37,408% utilizando o arredondamento com a base 10, representando assim um aumento de 0,634%.

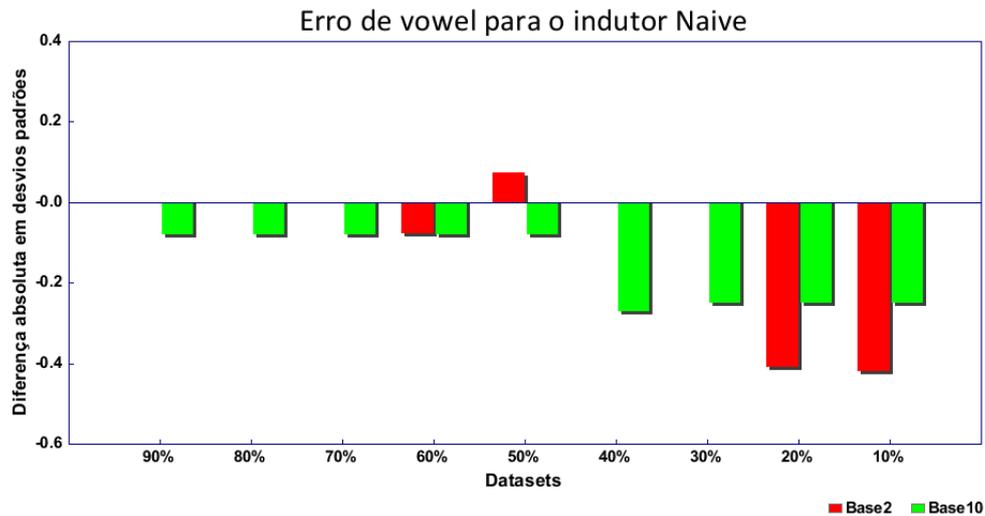


Figura 137: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) vowel

No geral, a taxa de erro aumentou para a maioria dos conjuntos, porém de forma não significativa.

### 6.5.10 Resultados wine

Na Tabela 52 são mostrados os resultados (média  $\pm$  desvio padrão) obtidos em relação aos conjuntos de exemplos wine original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente. A quarta e quinta colunas representam os resultados da taxa de erro, utilizando a base binária e a base decimal, respectivamente.

Na Figura 138 é mostrada a diferença absoluta em desvios padrões do tempo de indução entre o conjunto original e os conjuntos derivados, ou seja, entre wine e wine-90%, entre wine e wine-80% e assim por diante, utilizando base 2 e base 10.

Para o arredondamento utilizando a base 2, o tempo de indução não manteve um padrão, aumentando para alguns conjuntos e reduzindo para outros. Já para o arredondamento utilizando a base 10, o tempo de indução reduziu para a maioria dos conjuntos, sendo de forma não significativa para ambas as bases.

Analogamente, na Figura 139 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico. Em média, a taxa de erro aumentou de 2,222% (wine), para 3,653% utilizando o arredondamento com a base 2, o que representa um aumento de 64,396% da taxa de erro. E aumentou para 3,293% utilizando o arredondamento com a base 10, representando assim um aumento de 48,215%.

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)
wine	$1,000 \pm 0,077$	$1,000 \pm 0,077$	$2,222 \pm 1,228$	$2,222 \pm 1,228$
wine-90%	$1,000 \pm 0,077$	$1,000 \pm 0,077$	$2,778 \pm 1,242$	$2,778 \pm 1,242$
wine-80%	$1,049 \pm 0,118$	$1,000 \pm 0,077$	$2,222 \pm 1,228$	$2,778 \pm 1,242$
wine-70%	$0,976 \pm 0,000$	$1,000 \pm 0,077$	$2,778 \pm 1,242$	$2,222 \pm 0,907$
wine-60%	$1,024 \pm 0,103$	$1,000 \pm 0,077$	$3,366 \pm 1,235$	$2,222 \pm 0,907$
wine-50%	$1,000 \pm 0,077$	$0,976 \pm 0,000$	$3,366 \pm 1,235$	$2,222 \pm 0,907$
wine-40%	$1,024 \pm 0,154$	$0,976 \pm 0,000$	$2,778 \pm 1,242$	$2,222 \pm 0,907$
wine-30%	$0,976 \pm 0,000$	$0,976 \pm 0,000$	$3,889 \pm 1,446$	$3,954 \pm 1,197$
wine-20%	$1,024 \pm 0,103$	$1,024 \pm 0,103$	$3,889 \pm 1,446$	$3,366 \pm 0,917$
wine-10%	$0,976 \pm 0,000$	$0,976 \pm 0,000$	$7,810 \pm 1,693$	$7,876 \pm 1,708$

Tabela 52: Tempo de indução e taxa de erro do classificador para o indutor NB utilizando arredondamento com bases 2 e 10 wine

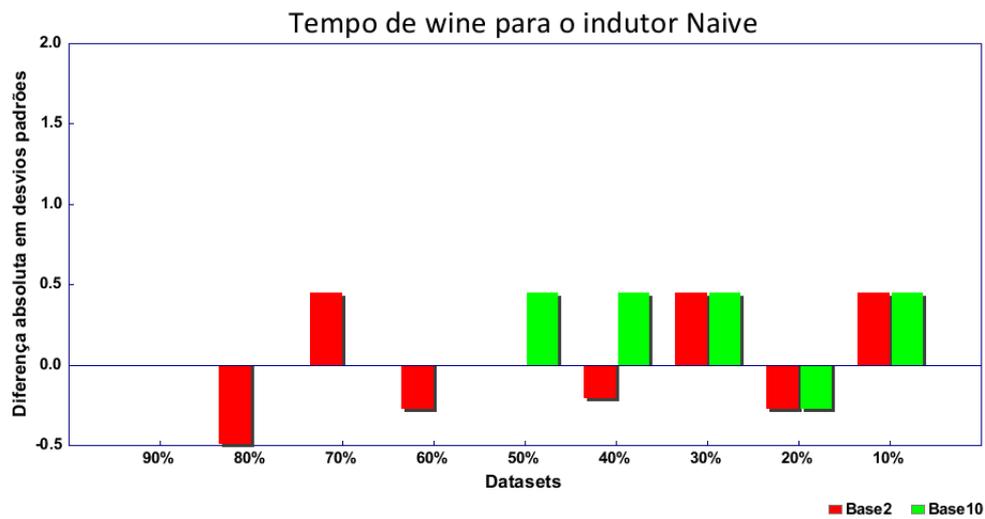


Figura 138: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

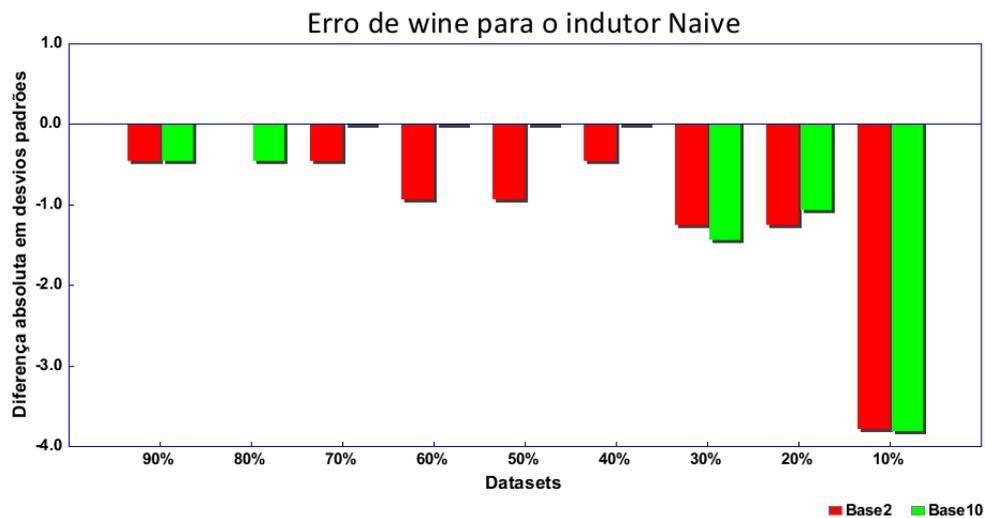


Figura 139: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 e base 10 *versus* conjunto original) wine

A taxa de erro aumentou para todos os conjuntos, sendo de forma significativa para wine-10%, utilizando arredondamento com ambas as bases.

### 6.5.11 Discussão

Nas Figuras 140, 141, 142 e 143 é mostrado um resumo dos resultados obtidos por meio do indutor NB do tempo de indução utilizando base 2, tempo de indução utilizando base 10, taxa de erro utilizando base 2, taxa de erro utilizando base 10, respectivamente. No total são 81 conjuntos (9 conjuntos arredondados  $\times$  9 conjuntos de exemplos). No eixo  $y$  do gráfico é mostrado o número conjuntos que ficaram em cada uma das categorias mostradas no eixo  $x$  do gráfico, que são:

- aqueles que tiveram o tempo significativamente maior para os conjuntos derivados em relação ao conjunto original (barras maior que -2);
- os que tiveram o tempo maior para os conjuntos derivados (barras variando entre -2 e 0);
- os que tiveram o tempo igual para os conjuntos derivados e o conjunto original (barras igual 0);
- os que tiveram o tempo menor para os conjuntos derivados (barras variando entre 0 e 2);
- os que tiveram o tempo significativamente menor para os conjuntos derivados em relação ao conjunto original (barras maior que 2);

Analogamente para a taxa de erro.

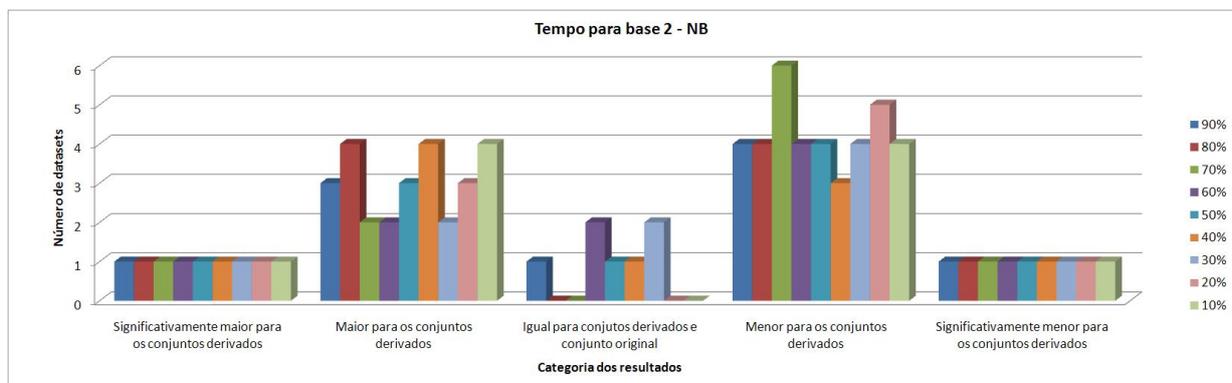


Figura 140: Resumo dos resultados do tempo de indução utilizando base 2 de NB

Como pode ser observado, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma não significativa.

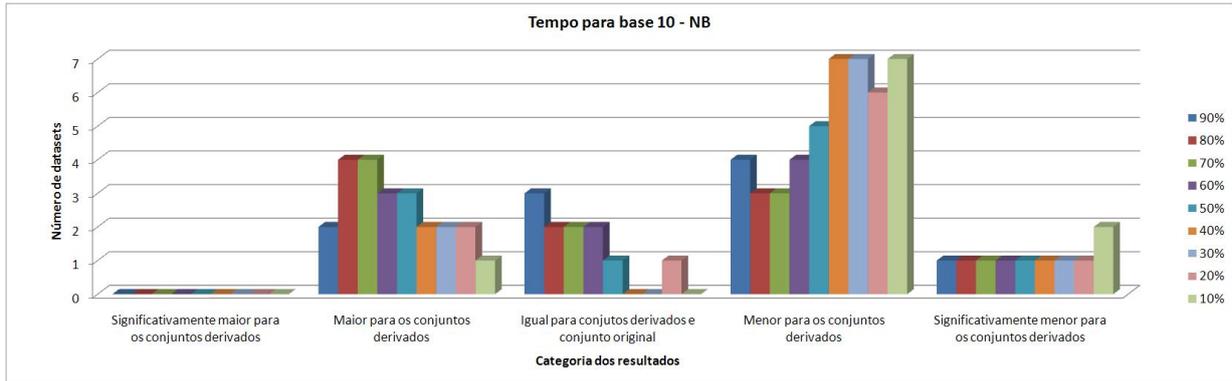


Figura 141: Resumo dos resultados do tempo de indução utilizando base 10 de NB

Como pode ser visto, houve uma tendência do tempo de indução reduzir para os conjuntos derivados, sendo que a maioria reduzindo de forma não significativa.

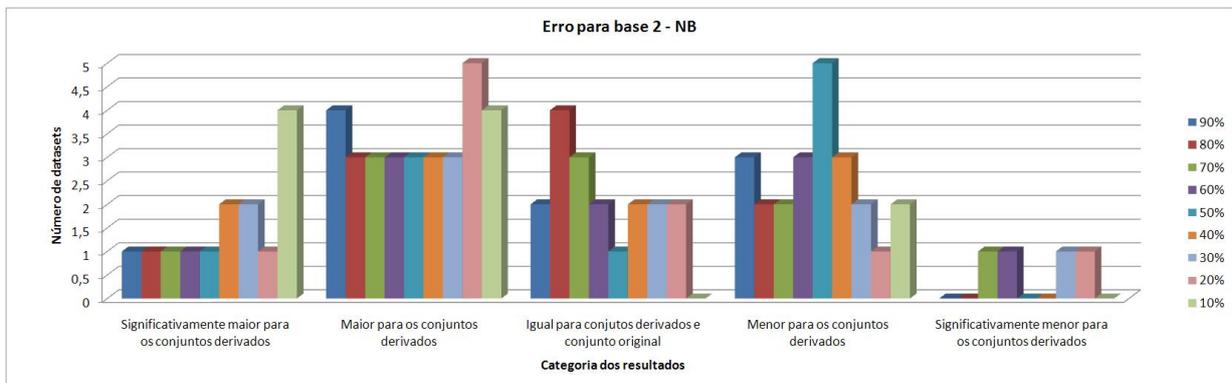


Figura 142: Resumo dos resultados da taxa de erro utilizando base 2 de NB

Como pode ser notado, houve uma tendência da taxa de erro aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa.

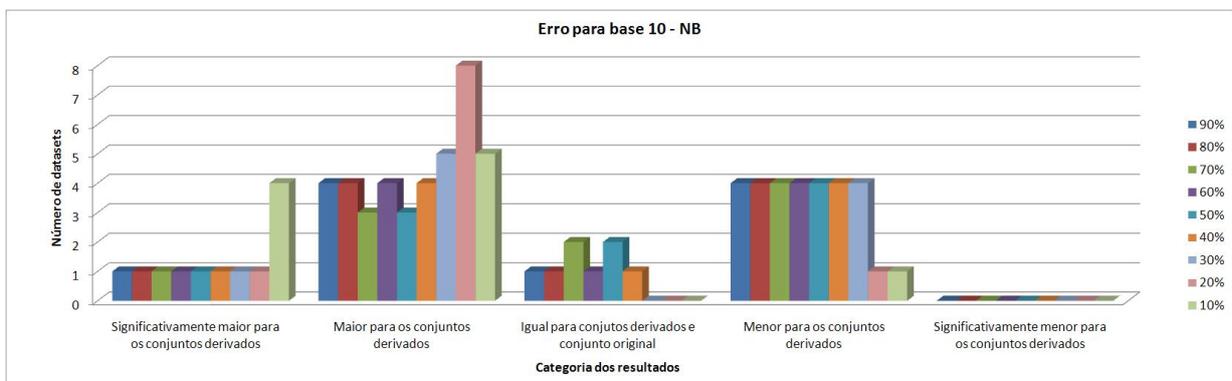


Figura 143: Resumo dos resultados da taxa de erro utilizando base 10 de NB

Como pode ser observado, houve uma tendência da taxa de erro aumentar para os conjuntos derivados, sendo que a maioria aumentado de forma não significativa.

## 7 Discussão Geral

Nas Figuras 144, 145, 146, 147, 148 e 149, é mostrado um resumo dos resultados obtidos para o tempo de indução utilizando base 2, tempo de indução utilizando base 10, taxa de erro utilizando base 2, taxa de erro utilizando base 10, tamanho do classificador utilizando base 2, tamanho do classificador utilizando base 10, respectivamente. Para se ter uma idéia do comportamento geral do arredondamento os resultados de todos os indutores foram mostrados no mesmo gráfico, sendo que para a métrica “tamanho do classificador” foram mostrados os resultados apenas dos indutores J48 e JRip.

No eixo  $y$  do gráfico é mostrado o número conjuntos que ficaram em cada uma das categorias mostradas no eixo  $x$  do gráfico, que são:

- aqueles que tiveram o tempo significativamente maior para os conjuntos derivados em relação ao conjunto original (barras maior que 2);
- os que tiveram o tempo maior para os conjuntos derivados (barras variando entre -2 e 0);
- os que tiveram o tempo igual para os conjuntos derivados e o conjunto original (barras igual 0);
- os que tiveram o tempo menor para os conjuntos derivados (barras variando entre 0 e 2);
- os que tiveram o tempo significativamente menor para os conjuntos derivados em relação ao conjunto original (barras maior que 2);

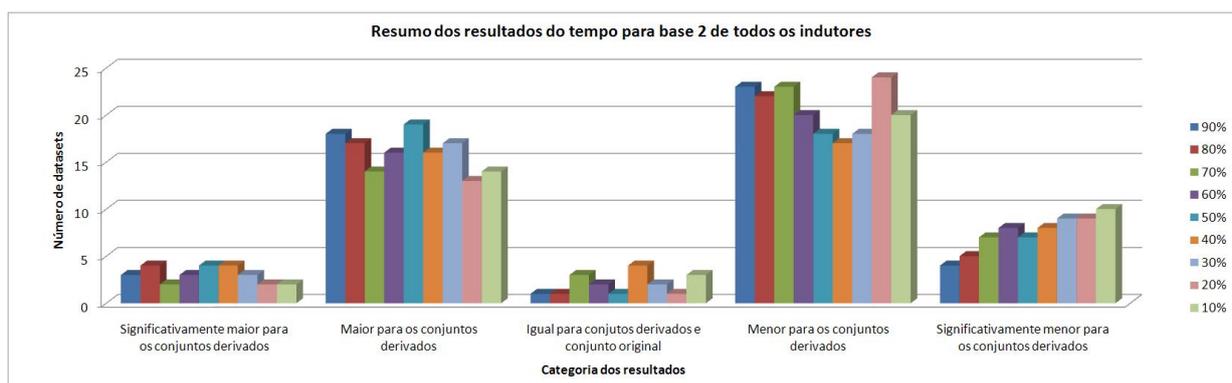


Figura 144: Resumo dos resultados do tempo de indução para a base 2 de todos os indutores

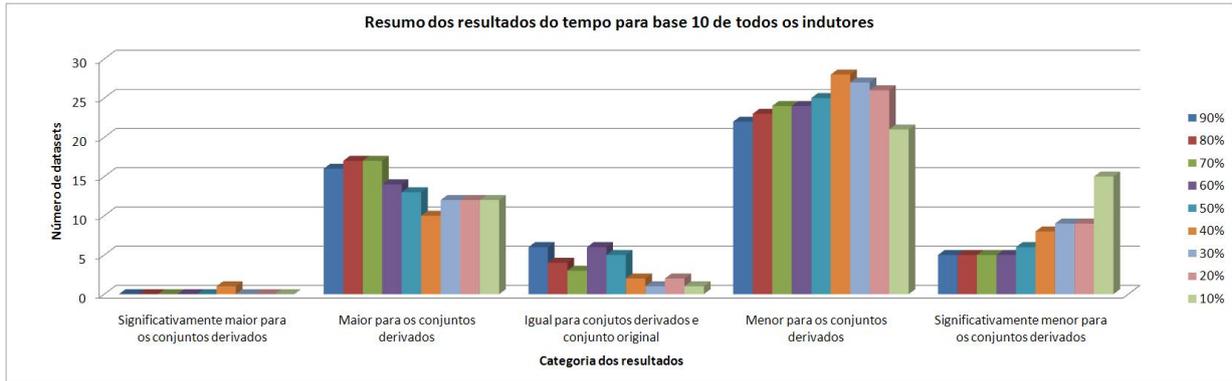


Figura 145: Resumo dos resultados do tempo de indução para a base 10 de todos os indutores



Figura 146: Resumo dos resultados da taxa de erro para a base 2 de todos os indutores

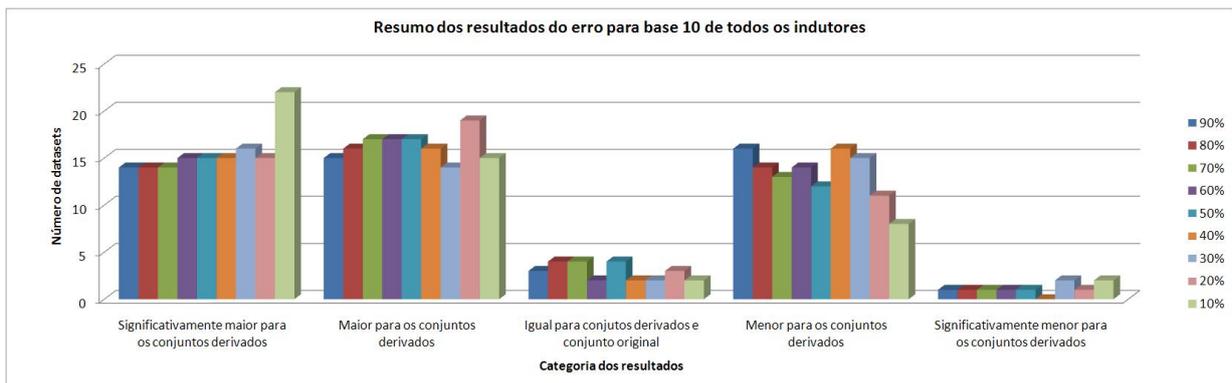


Figura 147: Resumo dos resultados da taxa de erro para a base 10 de todos os indutores

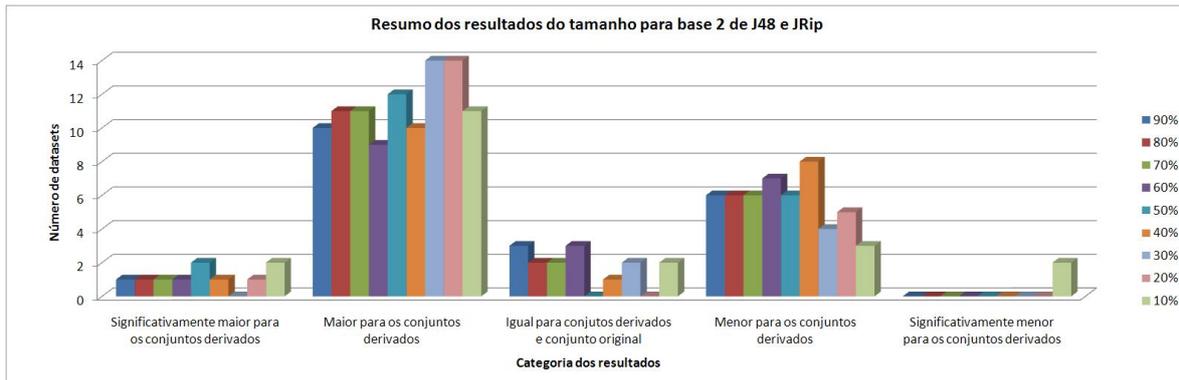


Figura 148: Resumo dos resultados do tamanho do classificador para a base 2 dos indutores J48 e JRip

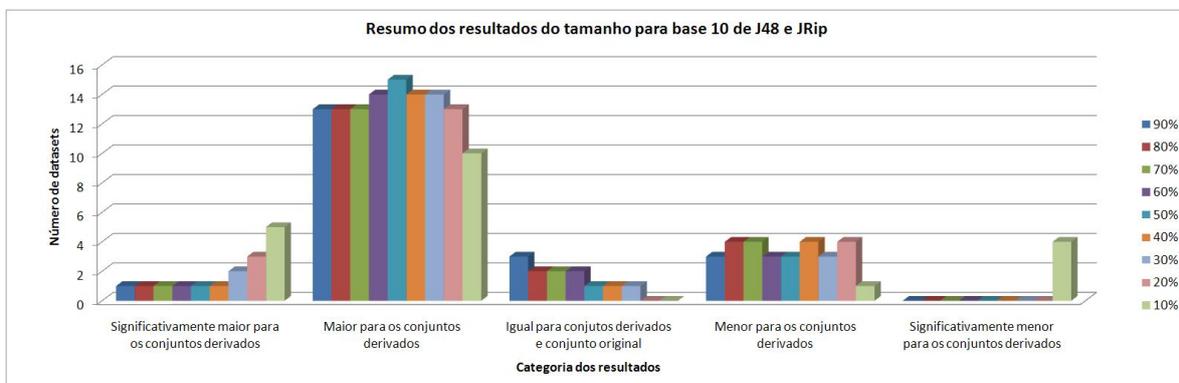


Figura 149: Resumo dos resultados do tamanho do classificador para a base 2 dos indutores J48 e JRip

Analisando o comportamento dos diversos paradigmas de AM estudados frente ao arredondamento de valores é possível observar que, no geral, o tempo de indução permanece o mesmo ou é reduzido em 66,55% das vezes (587 igualdades ou reduções *versus* 295 aumentos), para as bases decimais e binárias avaliadas; a redução de tempo ocorre em 71,88% das vezes, sendo maior para base decimal (317 igualdades ou reduções *versus* 124 aumentos) do que para base binária, que ocorre em 61,22% das vezes (270 igualdades ou reduções *versus* 171 aumentos). Esse resultado já era esperado para árvores de decisão (Quinlan 1993), e que se estendeu para os demais paradigmas. Isso indica que, para grandes conjuntos de exemplos, o arredondamento de valores pode ser considerado como uma forma de diminuir o tempo de indução, principalmente para indutores com grande tempo de aprendizado.

Com relação à taxa de erro, no geral, ocorre um aumento ao utilizar-se arredondamento; entretanto, o aumento do erro não é proporcional à quantidade de valores distintos (parâmetro  $p$  do Algoritmo 2), principalmente considerando a base decimal. No total, o aumento ocorreu em 64,74% dos casos: ocorreram 253 aumentos significativos e 318 aumentos não significativos *versus* 311 igualdades ou reduções na taxa de erro; o aumento na taxa de erro foi menor para a base binária, ocorrendo em 64,63% das vezes (113 aumentos significativos e 172 aumentos não significativos *versus* 156 igualdades ou reduções) do que para a base decimal, ocorrendo em 64,85% das vezes (140 aumentos significativos e 146 aumentos não significativos *versus* 155 igualdades ou reduções).

Considerando o tamanho dos classificadores J48 e JRip há uma tendência de aumento, embora

não significativo para as duas bases avaliadas, ocorrendo em 68,61% dos casos: ocorreram 26 aumentos significativos e 221 aumentos não significativos *versus* 113 igualdades ou reduções no tamanho do classificador; o aumento no tamanho foi maior para a base decimal, ocorrendo em 75,00% das vezes (16 aumentos significativos e 119 aumentos não significativos *versus* 45 igualdades ou reduções) do que para a base binária, ocorrendo em 62,22% das vezes (10 aumentos significativos e 102 aumentos não significativos *versus* 68 igualdades ou reduções).

## 8 Considerações Finais

Durante a pesquisa bibliográfica para este trabalho, foi possível encontrar alternativas de arredondamento, por exemplo em P. S. Miner and J. F. Leathrum (1996)[Definições 8, 9 e 10] que correspondem na prática às linhas 13 a 17 do Algoritmo 1, que podem ser investigadas em trabalhos futuros.

Os principais resultados desta pesquisa comprovam que há uma redução no tempo de indução, para diferentes conjuntos de exemplos e diferentes paradigmas de aprendizado. Entretanto, a redução do tempo pode ser acompanhada por um aumento tanto na taxa de erro como no tamanho do classificador induzido. Embora não seja possível aconselhar o arredondamento de valores para todas as situações, é importante observar que, para casos individuais de indutores e conjuntos de exemplos, sua utilização pode trazer benefícios quanto à redução da taxa de erro e tamanho do classificador.

## Referências

- Aha, D. W. (1992). Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies* 36, 267–287.
- Aha, D. W. (1997). Lazy learning. *Artificial Intelligence Review* 11, 7–10.
- Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984). *Classification and Regression Trees*. Pacific Grove, CA: Wadsworth & Books.
- Cohen, W. W. (1995). Fast effective rule induction. In *Proceedings of the Twelfth International Conference on Machine Learning*, San Francisco, CA, pp. 115–123. Morgan Kaufmann.
- Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein (2002). *Algoritmos: Teoria e Prática*. Campus. 2ª edição.
- Deitel, H. M. and P. J. Deitel (Eds.) (2005). *Java: Como Programar*. Prentice-Hall.
- Forina, M. (1991). An extendible package for data exploration, classification and correlation.
- Gamberger, D., N. Lavrac, F. Zelezny, and J. Tolar (2004). Induction of comprehensible models for gene expression datasets by subgroup discovery methodology. *Journal of Biomedical Informatics* 37, 269–284.
- Golub, T. R. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537.
- Gorman, R. P. and T. J. Sejnowski (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks* 1, 75–89.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall; 2nd edition.
- Heckerman, D. (1996). *Bayesian Networks for Knowledge Discovery*, pp. 273–306. Menlo Park, CA: American Association for Artificial Intelligence.

- Langley, P., W. Iba, and K. Thompson (1992). An analysis of bayesian classifiers. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pp. 223–228. AAAI Press and MIT Press.
- Mitchell, T. M. (1998). *Machine Learning*. McGraw–Hill.
- Moses, L. E. (Ed.) (1986). *Think and Explain with Statistics*. Addison–Wesley.
- Newman, D. J., S. Hettich, C. Blake, and C. Merz (1998). UCI repository of machine learning databases. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- P. S. Miner and J. F. Leathrum (1996). Verification of IEEE compliant subtractive division algorithms. In M. Srivas and A. Camilleri (Eds.), *First international conference on formal methods in computer-aided design*, Volume 1166, Palo Alto, CA, USA, pp. 64–78. Springer Verlag.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning 1*, 81–106. Reprinted in Shavlik and Dietterich (eds.), 1990. *Readings in Machine Learning*, Morgan Kaufmann Publishers, Inc.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann. San Francisco, CA.
- Rezende, S. O. (Ed.) (2003). *Sistemas Inteligentes*. Manole.
- Turney, P. (1993). Robust classification with context-sensitive features.
- Weiss, S. M. and N. Indurkha (1998). *Predictive Data Mining: A Practical Guide*. San Francisco, CA: Morgan Kaufmann.
- Weiss, S. M. and C. A. Kulikowski (1991). *Computer Systems that Learn*. San Mateo, CA: Morgan Kaufmann.
- Wirth, N. (1986). *Algoritmos e Estruturas de Dados*. Prentice Hall do Brasil.
- Witten, I. H. and E. Frank (1999, october). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Volume 1. Morgan Kaufmann.