Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto

Universidade de São Paulo

Avaliação de Arredondamento de Valores de Atributos Contínuos na Indução de Árvores de Decisão

Rogério Nunes Lemos
José Augusto Baranauskas

Relatórios Técnicos Do
Departamento de Física e Matemática
Da FFCLRP-USP

Ribeirão Preto
Fevereiro/2006

Avaliação de Arredondamento de Valores de Atributos Contínuos na Indução de Árvores de Decisão

Rogério Nunes Lemos[1,2]
rnlemos@fmrp.usp.br

José Augusto Baranauskas[1]
augusto@fmrp.usp.br

[1]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

[2]Universidade de São Paulo
Faculdade de Medicina de Ribeirão Preto
Avenida do Café, 3900
14049-900 - Ribeirão Preto, SP - Brasil

Resumo: A maior parte das operações para construir uma árvore de decisão cresce linearmente com o número de exemplos de treinamento. Entretanto, o processo de escolha de um atributo contínuo contendo d valores distintos requer a ordenação desses valores, crescendo como $d \log_2 d$. Assim, o tempo requerido para construir uma árvore de decisão a partir de um conjunto de treinamento grande pode ser dominado pela ordenação dos atributos contínuos. Neste relatório técnico é avaliado o arredondamento de valores de atributos contínuos no processo de indução de árvores de decisão, considerando não só o tempo de indução e a taxa de erro como também o tamanho final do classificador induzido.

Este documento foi preparado com o formatador de textos LaTeX. O sistema de citações de referências bibliográficas utiliza o padrão *Chicago* do sistema 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

Lista de Figuras

Lista de Tabelas

Lista de Algoritmos

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 & 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, $\overrightarrow{z} = (\overrightarrow{x}, y)$. 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, $h(x) \approx f(x)$.

Dentre os algoritmos de AM supervisionado utilizando classificação, tema desta pesquisa, existe uma família de algoritmos de AM indutivo conhecida como *Top Down Induction of Decision Trees* — TDIDT. De modo simplificado a indução de uma árvore de decisão realiza-se da seguinte forma (Breiman, Friedman, Olshen & 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 então atribuída ao novo exemplo.

Para exemplificar o processo de classificação em termos práticos, suponha que se queira aprender uma forma para predizer se um paciente tem problemas cardíacos. Para isso, é necessário verificar os históricos dos pacientes nos quais seriam encontrados registros contendo atributos, tais como idade, sexo, dor no peito, nível de colesterol, taxa máxima de batimentos cardíacos, a presença de angina induzida por exercícios, entre outros. Presume-se que cada registro histórico tenha sido diagnosticado (rotulado) por um especialista médico como um paciente *saudável* ou *doente*. O conjunto de exemplos composto por históricos de pacientes é então fornecido como entrada para um algoritmo de indução. A saída resultante, ou seja, a hipótese induzida, normalmente consiste em algumas regras que permitem classificar novos pacientes,

isto é, que permitem determinar se um novo paciente apresenta ou não problema cardíacos. Na Figura 1 é mostrada parte de uma árvore de decisão induzida a partir de dados reais provenientes do conjunto de exemplos cleve — Cleveland *heart disease* (Newman, Hettich, Blake & Merz 1998). Essa árvore pode ser utilizada para classificar novos pacientes: começando pela raiz da árvore, repetidamente segue-se o ramo de acordo com o atributo testado até que um nó folha seja encontrado, o qual rotula o paciente como saudável (*healthy*) ou doente (*sick*).

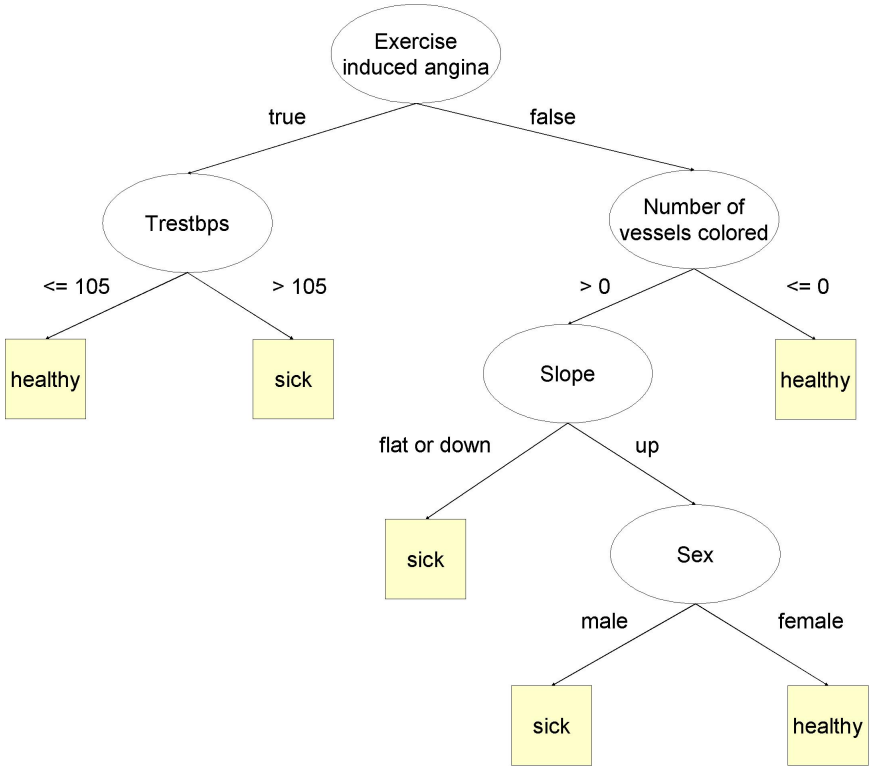

Figura 1: Parte da árvore de decisão induzida por J48/C4.5 para o conjunto de exemplos Cleveland *heart disease*

Um ponto importante é que enquanto a maior parte das operações para construir uma árvore de decisão 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 & Frank 1999) fazem uso do algoritmo *quicksort* para ordenar valores contínuos (Cormen, Leiserson, Rivest & Stein 2002)[Cap. 7], (Wirth 1986)[Cap. 2].

Outro fator importante que também deve ser considerado é o grau de compreensibilidade proporcionado ao ser humano. De acordo com Michalski (1983a) e Kubat, Bratko & Michalski (1998), os sistemas de aprendizado são classificados em duas grandes categorias:

1. sistemas *caixa-preta* que desenvolvem sua própria representação do conceito, isto é, sua representação interna pode não ser facilmente interpretada por humanos e não fornecem nem esclarecimento, nem explicação do processo de reconhecimento;

2. sistemas *orientados a conhecimento* que objetivam a criação de estruturas simbólicas que sejam compreensíveis por humanos.

Assim, no aprendizado de conceitos, o interesse principal consiste em obter descrições simbólicas que sejam fáceis de serem compreendidas e utilizadas por meio de modelos mentais.

Segundo o *postulado da compreensibilidade* de Michalski (1983b):

> "Os resultados da indução por computador devem ser descrições simbólicas das entidades dadas, sendo semântica e estruturalmente similares àquelas que um especialista humano poderia produzir observando as mesmas entidades. Os componentes dessas descrições devem ser compreensíveis como simples 'pedaços' de informação, diretamente interpretáveis em linguagem natural, bem como reportar conceitos quantitativos e qualitativos de maneira integrada."

Como regra prática, Michalski assume que os componentes de descrição, tais como regras ou nós em uma árvore de decisão, devem ser expressões contendo menos de cinco condições em uma conjunção; poucas condições em uma disjunção; no máximo um nível de parênteses; no máximo uma implicação; não mais de dois quantificadores e nenhuma recursão. Embora esses valores possam ser flexíveis, descrições geradas por indução dentro dos limites propostos são similares à representação do conhecimento humano e, portanto, fáceis de serem compreendidas. Embora tais medidas sejam simples de serem avaliadas, é importante salientar que elas são meramente sintáticas e que, muitas vezes, medidas semânticas devam ser consideradas (Pazzani 2000).

Em Aprendizado de Máquina existem muitos algoritmos de aprendizado que induzem classificadores. Este trabalho se concentra em indutores que contribuem para a compreensão dos dados em contraste com indutores que visam apenas uma grande precisão. Por exemplo, a indução de regras ou árvores de decisão pode auxiliar médicos a compreenderem melhor os dados, enquanto uma rede neural convencional, mesmo com precisão similar, pode ser extremamente difícil de ser compreendida por seres humanos[1]. Por exemplo, no desenvolvimento de sistemas especialistas é importante que especialistas humanos possam, fácil e confiavelmente, verificar o conhecimento extraído e relacioná-lo ao seu próprio domínio de conhecimento. Além disso, algoritmos de aprendizado que induzem estruturas compreensíveis, contribuindo para a compreensão do domínio considerado, podem produzir conhecimento novo (Dietterich 1986).

O objetivo deste trabalho consiste na avaliação do arredondamento de valores de atributos no processo de indução de árvores de decisão, ou seja, neste trabalho é tratado o aprendizado 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 é utilizado o indutor de árvores de decisão J48 da biblioteca Weka (Witten & Frank 1999) – Waikato Environment for Knowledge Analysis, uma reimplementação na linguagem Java do indutor C4.5 (Quinlan 1993).

O restante deste trabalho está organizado da seguinte forma: Na Seção 2 são descritos os conjuntos de exemplos utilizados nos experimentos realizados: Experimento 1 (Seção 3) e Experimento 2 (Seção 4) e Experimento 3 (Seção 6). Os Experimentos 1 e 2 foram conduzidos utilizando arredondamento científica usando redução de casas decimais. Na Seção 5 é mostrada uma metodologia diferente da adotada nesses experimentos, proposta por Weiss & Indurkhya (1998). Por último, na Seção 6 são mostrados os experimentos aplicando a metodologia descrita na Seção 5, bem com uma discussão dos resultados. 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 sonar,

[1]Existem, entretanto, vários métodos desenvolvidos para a extração de regras a partir de redes neurais.

ionosphere, vowel e wine foram obtidos a partir do repositório UCI Irvine (Newman, Hettich, Blake & Merz 1998). O conjunto aml-all foi obtido de Golub (1999).

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

sonar Este conjunto de exemplos foi usado por Gorman & 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 freqüência 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).

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 freqüência 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.

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).

aml-all 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 & Tolar 2004).

Na Tabela 1 são resumidas algumas características dos conjuntos de exemplos utilizados. 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.

Conjunto de Exemplos	#Exemplos	#Atributos (cont.;nom.)	#Classes	Erro Majoritário	Valor Desconhecido
sonar	208	60 (60;0)	2	46,63%	não
ionosphere	351	34 (34;0)	2	35,90%	não
vowel	990	13 (10;3)	11	90,91%	não
wine	178	13 (13;0)	3	60,11%	não
aml-all	72	7129 (7129;0)	2	28,95%	não

Tabela 1: Características dos conjuntos de exemplos

3 Experimento 1

O primeiro experimento foi realizado inicialmente utilizando somente o conjunto de exemplos sonar. Este experimento teve como objetivo avaliar o comportamento do tempo de indução[2] utilizando ou não arredondamento.

O conjunto de exemplos sonar foi submetido ao indutor J48, a indução foi realizada sem poda, obtendo-se uma árvore de decisão. Com base nisso, foram anotados os atributos que apareceram no classificador induzido e, utilizando arredondamento científico, usando redução de casas decimais, foram gerados três conjuntos derivados de sonar denotados como sonar-p (sonar-p3, sonar-p2 e sonar-p1) a partir do conjunto original (nenhum arredondamento aplicado), com seus valores arredondados para 3, 2 e 1 casas decimais, respectivamente, somente para aqueles atributos que apareceram no classificador J48. Por exemplo, o valor 0,3109 foi arredondado para 0,311 em sonar-p3, 0,31 em sonar-p2 e 0,3 em sonar-p1.

Analisando os três classificadores (obtidos a partir dos conjuntos derivados sonar-p) foi possível notar um conjunto diferente de atributos daquele obtido a partir da árvore induzida utilizando o conjunto original de exemplos sonar. Isso significa que, ao realizar o arredondamento, nos conjuntos derivados sonar-p alguns atributos foram substituídos por outros na árvore quando comparada à árvore induzida a partir de sonar.

Diante dessa situação foram gerados três conjuntos derivados adicionais denotados como sonar-t (sonar-t3, sonar-t2 e sonar-t1) a partir do conjunto original, com seus valores arredondados para 3, 2 e 1 casas decimais, respectivamente, para todos os atributos.

Ainda nesse experimento inicial dois outros conjuntos de exemplos foram utilizados: ionosphere e vowel. Como observado no caso de sonar, o arredondamento apenas de atributos que aparecem na árvore sem poda induzida a partir do conjunto original de exemplo pode resultar na escolha de outros atributos; assim sendo essa estratégia não foi utilizada para os estes conjuntos de exemplos.

Para o conjunto ionosphere foram gerados quatro conjuntos derivados ionosphere-t (ionosphere-t4, ionosphere-t3, ionosphere-t2 e ionosphere-t1), que tiveram seus valores arredondados para 4, 3, 2 e 1 casas decimais, respectivamente, para todos os atributos. De forma análoga para o conjunto vowel foram gerados dois conjuntos derivados vowel-t (vowel-t2 e vowel-t1), que tiveram seus valores arredondados para 2 e 1 casas decimais, respectivamente, para todos os atributos contínuos. Nota-se que o número máximo de casas decimais no conjunto original é peculiar a cada conjunto original de exemplos, resultando em um número diferente de conjuntos derivados para sonar, ionosphere e vowel.

Nesta fase inicial, para avaliar o desempenho foi utilizado *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 (em segundos). Adicionalmente, a taxa de erro e o tamanho do classificador — em número total de nós (tanto nós internos de teste quanto nós folhas) — foram também analisados, mesmo considerando o fato que, apenas

[2]Ressalta-se que todos os experimentos reportados nesse Relatório Técnico foram realizados no mesmo computador.

para a taxa de erro, a estimativa de desempenho obtida da forma proposta pode ter um *bias* otimista, já que os exemplos em todos os *folds* tiveram seus valores arredondados, incluindo o *fold* de teste.

3.1 Resultados sonar

Na Tabela 2 é mostrado o número de valores distintos para cada atributo tanto no conjunto original sonar, como nos derivados sonar-p (sonar-p1, p2, p3), sonar-t (sonar-t3, t2, t1).

Número do Atributo	Nome do Atributo	#Valores Distintos (vlr. relativo) sonar	#Valores Distintos (vlr. relativo) sonar-p3	#Valores Distintos (vlr. relativo) sonar-p2	#Valores Distintos (vlr. relativo) sonar-p1	#Valores Distintos (vlr. relativo) sonar-t3	#Valores Distintos (vlr. relativo) sonar-t2	#Valores Distintos (vlr. relativo) sonar-t1
#1	a01	177 (0,85)	66 (0,32)	15 (0,07)	2 (0,01)	66 (0,32)	15 (0,07)	2 (0,01)
#2	a02	182 (0,88)	82 (0,39)	16 (0,08)	3 (0,01)	82 (0,39)	16 (0,08)	3 (0,01)
#3	a03	190 (0,91)	190 (0,91)	190 (0,91)	190 (0,91)	90 (0,43)	20 (0,10)	4 (0,02)
#4	a04	181 (0,87)	93 (0,45)	19 (0,09)	5 (0,02)	93 (0,45)	19 (0,09)	5 (0,02)
#5	a05	193 (0,93)	193 (0,93)	193 (0,93)	193 (0,93)	112 (0,54)	23 (0,11)	5 (0,02)
#6	a06	196 (0,94)	196 (0,94)	196 (0,94)	196 (0,94)	132 (0,63)	27 (0,13)	5 (0,02)
#7	a07	195 (0,94)	195 (0,94)	195 (0,94)	195 (0,94)	134 (0,64)	31 (0,15)	5 (0,02)
#8	a08	201 (0,97)	142 (0,68)	36 (0,17)	6 (0,03)	142 (0,68)	36 (0,17)	6 (0,03)
#9	a09	205 (0,99)	205 (0,99)	205 (0,99)	205 (0,99)	156 (0,75)	45 (0,22)	8 (0,04)
#10	a10	207 (1,00)	207 (1,00)	207 (1,00)	207 (1,00)	165 (0,79)	53 (0,25)	8 (0,04)
#11	a11	203 (0,98)	164 (0,79)	52 (0,25)	8 (0,04)	164 (0,79)	52 (0,25)	8 (0,04)
#12	a12	206 (0,99)	206 (0,99)	206 (0,99)	206 (0,99)	165 (0,79)	54 (0,26)	8 (0,04)
#13	a13	198 (0,95)	198 (0,95)	198 (0,95)	198 (0,95)	167 (0,80)	57 (0,27)	8 (0,04)
#14	a14	202 (0,97)	202 (0,97)	202 (0,97)	202 (0,97)	171 (0,82)	57 (0,27)	10 (0,05)
#15	a15	203 (0,98)	203 (0,98)	203 (0,98)	203 (0,98)	176 (0,85)	73 (0,35)	11 (0,05)
#16	a16	203 (0,98)	203 (0,98)	203 (0,98)	203 (0,98)	182 (0,88)	76 (0,37)	11 (0,05)
#17	a17	202 (0,97)	202 (0,97)	202 (0,97)	202 (0,97)	176 (0,85)	77 (0,37)	11 (0,05)
#18	a18	204 (0,98)	178 (0,86)	79 (0,38)	11 (0,05)	178 (0,86)	79 (0,38)	11 (0,05)
#19	a19	206 (0,99)	206 (0,99)	206 (0,99)	206 (0,99)	175 (0,84)	85 (0,41)	11 (0,05)
#20	a20	203 (0,98)	203 (0,98)	203 (0,98)	203 (0,98)	182 (0,88)	79 (0,38)	10 (0,05)
#21	a21	200 (0,96)	185 (0,89)	77 (0,37)	10 (0,05)	185 (0,89)	77 (0,37)	10 (0,05)
#22	a22	203 (0,98)	203 (0,98)	203 (0,98)	203 (0,98)	184 (0,88)	83 (0,40)	11 (0,05)
#23	a23	199 (0,96)	176 (0,85)	75 (0,36)	10 (0,05)	176 (0,85)	75 (0,36)	10 (0,05)
#24	a24	201 (0,97)	201 (0,97)	201 (0,97)	201 (0,97)	174 (0,84)	78 (0,38)	11 (0,05)
#25	a25	198 (0,95)	198 (0,95)	198 (0,95)	198 (0,95)	182 (0,88)	80 (0,38)	11 (0,05)
#26	a26	194 (0,93)	194 (0,93)	194 (0,93)	194 (0,93)	175 (0,84)	74 (0,36)	10 (0,05)
#27	a27	190 (0,91)	172 (0,83)	75 (0,36)	11 (0,05)	172 (0,83)	75 (0,36)	11 (0,05)
#28	a28	194 (0,93)	171 (0,82)	74 (0,36)	11 (0,05)	171 (0,82)	74 (0,36)	11 (0,05)
#29	a29	197 (0,95)	197 (0,95)	197 (0,95)	197 (0,95)	178 (0,86)	79 (0,38)	11 (0,05)
#30	a30	202 (0,97)	202 (0,97)	202 (0,97)	202 (0,97)	182 (0,88)	76 (0,37)	10 (0,05)
#31	a31	207 (1,00)	207 (1,00)	207 (1,00)	207 (1,00)	190 (0,91)	77 (0,37)	11 (0,05)
#32	a32	205 (0,99)	205 (0,99)	205 (0,99)	205 (0,99)	182 (0,88)	75 (0,36)	10 (0,05)
#33	a33	205 (0,99)	205 (0,99)	205 (0,99)	205 (0,99)	188 (0,90)	77 (0,37)	11 (0,05)
#34	a34	206 (0,99)	206 (0,99)	206 (0,99)	206 (0,99)	183 (0,88)	75 (0,36)	11 (0,05)
#35	a35	205 (0,99)	205 (0,99)	205 (0,99)	205 (0,99)	186 (0,89)	75 (0,36)	11 (0,05)
#36	a36	205 (0,99)	205 (0,99)	205 (0,99)	205 (0,99)	186 (0,89)	83 (0,40)	11 (0,05)
#37	a37	206 (0,99)	206 (0,99)	206 (0,99)	206 (0,99)	181 (0,87)	81 (0,39)	10 (0,05)
#38	a38	206 (0,99)	206 (0,99)	206 (0,99)	206 (0,99)	173 (0,83)	74 (0,36)	11 (0,05)
#39	a39	204 (0,98)	204 (0,98)	204 (0,98)	204 (0,98)	170 (0,82)	64 (0,31)	11 (0,05)
#40	a40	206 (0,99)	206 (0,99)	206 (0,99)	206 (0,99)	184 (0,88)	64 (0,31)	10 (0,05)
#41	a41	204 (0,98)	204 (0,98)	204 (0,98)	204 (0,98)	175 (0,84)	63 (0,30)	10 (0,05)
#42	a42	208 (1,00)	208 (1,00)	208 (1,00)	208 (1,00)	174 (0,84)	60 (0,29)	9 (0,04)
#43	a43	205 (0,99)	205 (0,99)	205 (0,99)	205 (0,99)	176 (0,85)	58 (0,28)	8 (0,04)
#44	a44	196 (0,94)	196 (0,94)	196 (0,94)	196 (0,94)	156 (0,75)	52 (0,25)	8 (0,04)
#45	a45	205 (0,99)	205 (0,99)	205 (0,99)	205 (0,99)	162 (0,78)	55 (0,26)	8 (0,04)
#46	a46	199 (0,96)	199 (0,96)	199 (0,96)	199 (0,96)	152 (0,73)	52 (0,25)	8 (0,04)
#47	a47	202 (0,97)	202 (0,97)	202 (0,97)	202 (0,97)	145 (0,70)	38 (0,18)	6 (0,03)
#48	a48	204 (0,98)	204 (0,98)	204 (0,98)	204 (0,98)	133 (0,64)	29 (0,14)	4 (0,02)
#49	a49	193 (0,93)	193 (0,93)	193 (0,93)	193 (0,93)	98 (0,47)	19 (0,09)	3 (0,01)
#50	a50	154 (0,74)	154 (0,74)	154 (0,74)	154 (0,74)	50 (0,24)	8 (0,04)	2 (0,01)
#51	a51	160 (0,77)	45 (0,22)	8 (0,04)	2 (0,01)	45 (0,22)	8 (0,04)	2 (0,01)
#52	a52	144 (0,69)	144 (0,69)	144 (0,69)	144 (0,69)	39 (0,19)	7 (0,03)	2 (0,01)
#53	a53	134 (0,64)	31 (0,15)	5 (0,02)	1 (0,00)	31 (0,15)	5 (0,02)	1 (0,00)
#54	a54	134 (0,64)	31 (0,15)	5 (0,02)	1 (0,00)	31 (0,15)	5 (0,02)	1 (0,00)
#55	a55	129 (0,62)	129 (0,62)	129 (0,62)	129 (0,62)	29 (0,14)	5 (0,02)	1 (0,00)
#56	a56	122 (0,59)	122 (0,59)	122 (0,59)	122 (0,59)	26 (0,13)	5 (0,02)	1 (0,00)
#57	a57	121 (0,58)	121 (0,58)	121 (0,58)	121 (0,58)	27 (0,13)	5 (0,02)	1 (0,00)
#58	a58	124 (0,60)	124 (0,60)	124 (0,60)	124 (0,60)	29 (0,14)	5 (0,02)	1 (0,00)
#59	a59	119 (0,57)	119 (0,57)	119 (0,57)	119 (0,57)	29 (0,14)	5 (0,02)	1 (0,00)
#60	a60	109 (0,52)	109 (0,52)	109 (0,52)	109 (0,52)	24 (0,12)	4 (0,02)	1 (0,00)
Média		187,60				137,35	49,64	7,35

Tabela 2: Número de valores distintos dos atributos sonar

Como pode ser observado o arredondamento científico usando redução de casas decimais, diminui drasticamente o número de valores distintos. Em média, de 187,60 (sonar) para 137,35 (sonar-t3), para 49,64 (sonar-t2) e para 7,35 (sonar-t1). Isso corresponde a uma redução média de 26,79% de sonar para sonar-t3, de 73,54% de sonar para sonar-t2 e de 96,08% de sonar para sonar-t1.

Na Tabela 3 são mostrados os atributos que aparecem nas árvores induzidas, número de atributos (#A) e porcentagem do total de atributos (%A), usando todo o conjunto de exemplos, tanto para o conjunto original como para os derivados sonar. Analogamente, essa informação é mostrada de outra forma na Tabela 4. É possível notar que há diferenças entre os atributos que aparecem nas árvores, mesmo entre conjuntos com um mínimo de arredondamento de valores, por exemplo, entre sonar, sonar-p3 e sonar-t3.

Conjunto	Atributos	#A	%A
sonar	1, 2, 4, 8, 11, 18, 21, 23, 27, 28, 51, 53, 54	13	21,67%
sonar-p3	1, 2, 4, 6, 8, 11, 18, 21, 27, 28, 43, 51, 53, 54	14	23,33%
sonar-p2	2, 4, 8, 9, 11, 27, 37, 39, 43, 45, 54, 55	12	20,00%
sonar-p1	3, 4, 8, 11, 12, 17, 23, 36, 45, 47, 52	11	18,33%
sonar-t3	1, 2, 4, 8, 11, 18, 21, 27, 28, 50, 51, 53, 54, 55	14	23,33%
sonar-t2	2, 5, 8, 11, 15, 20, 21, 23, 27, 39, 49, 52, 57, 58, 59	15	25,00%
sonar-t1	4, 8, 11, 12, 17, 19, 20, 23, 31, 32, 36, 37, 41, 44, 45, 48	16	26,67%

Tabela 3: Atributos que aparecem na árvore induzida sonar

É possível observar que a quantidade de atributos aumentou para todos os conjuntos sonar-t; e também que os atributos são diferentes, por exemplo, o atributo #23 deixou de aparecer no classificador de sonar-t3, voltando a ser importante tanto para os classificadores de sonar-t2 e sonar-t1

Na Tabela 5 são mostrados os resultados (média ± desvio padrão) obtidos em relação aos conjuntos de exemplos sonar original e derivados usando *10-fold stratified cross-validation*. Em média, o tempo de indução diminuiu de 0,22 (sonar) para 0,16 (sonar-p) e para 0,13 (sonar-t). Isso significa uma redução de 27,27% e 40,91% do tempo de indução, respectivamente. É possível notar, nos conjuntos sonar-p que o tempo de indução foi similar, já que o indutor ordenou mais valores, ou seja, tanto os atributos arredondados como aqueles que não foram arredondados. Isso não aconteceu com os conjuntos sonar-t, já que menos valores tiveram que ser ordenados.

Na Figura 2 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 sonar e sonar-t1, entre sonar e sonar-t2 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 mostrados nas Figuras 3 e 4, respectivamente.

Número do Atributo	sonar	sonar-p3	sonar-p2	sonar-p1	sonar-t3	sonar-t2	sonar-t1
#1	•	•			•		
#2	•	•	•		•	•	
#3				•			
#4	•	•	•	•	•		•
#5						•	
#6		•					
#7							
#8	•	•	•	•	•	•	•
#9			•				
#10							
#11	•	•	•	•	•	•	•
#12				•			•
#13							
#14							
#15						•	
#16							
#17				•			•
#18	•	•			•		
#19							•
#20						•	•
#21	•	•			•	•	
#22							
#23	•			•		•	•
#24							
#25							
#26							
#27	•	•	•		•	•	
#28	•	•			•		
#29							
#30							
#31							•
#32							•
#33							
#34							
#35							
#36				•			•
#37			•				•
#38							
#39			•			•	
#40							
#41							•
#42							
#43		•	•				
#44							•
#45			•	•			•
#46							
#47				•			
#48							•
#49						•	
#50					•		
#51	•	•			•		
#52				•		•	
#53	•	•			•		
#54	•	•	•		•		
#55			•		•		
#56							
#57						•	
#58						•	
#59						•	
#60							
Total 60 100%	13 21,67%	14 23,33%	12 20,00%	11 18,33%	14 23,33%	15 25,00%	16 26,67%

Tabela 4: Atributos selecionados pelo classificador sonar

Conjunto	Tempo (s)	Erro	Tamanho
sonar	$0,22 \pm 0,11$	$28,83 \pm 2,24$	$29,20 \pm 3,58$
sonar-p3	$0,16 \pm 0,02$	$29,95 \pm 2,26$	$29,20 \pm 3,82$
sonar-p2	$0,16 \pm 0,01$	$27,88 \pm 2,43$	$29,40 \pm 3,24$
sonar-p1	$0,16 \pm 0,01$	$23,02 \pm 3,62$	$30,20 \pm 2,70$
sonar-t3	$0,16 \pm 0,02$	$27,40 \pm 2,48$	$29,60 \pm 4,12$
sonar-t2	$0,13 \pm 0,02$	$25,98 \pm 3,13$	$32,40 \pm 2,50$
sonar-t1	$0,11 \pm 0,05$	$23,98 \pm 3,65$	$35,60 \pm 3,66$

Tabela 5: Tempo de indução, taxa de erro e tamanho do classificador sonar

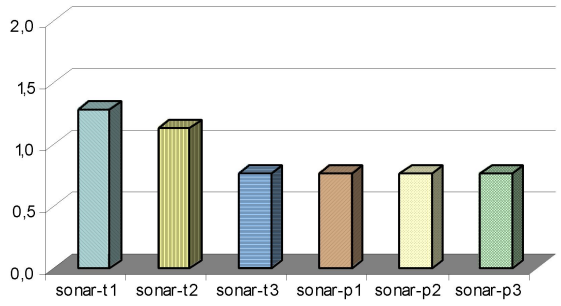

Figura 2: Diferença absoluta do tempo de indução sonar

Como esperado, o tempo de indução reduziu para todos os conjuntos utilizando arredondamento, embora não de forma significativa (com grau de confiança de 95%).

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 diminuiu de 28,83% (sonar) para 26,95% (sonar-p) e para 25,79% (sonar-t). Isso significa uma redução de 6,52% e 10,54% da taxa de erro, respectivamente. Como pode ser observado, a taxa de erro reduziu para todos os conjuntos utilizando arredondamento, embora não de forma significativa.

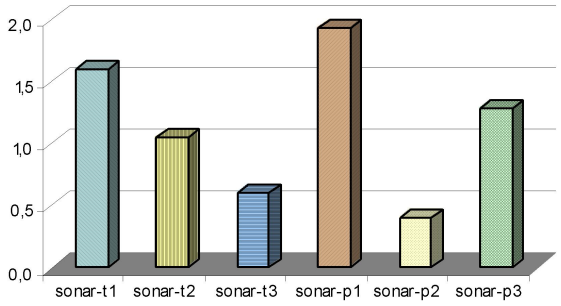

Figura 3: Diferença absoluta da taxa de erro sonar

Na Figura 4 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 29,20 (sonar) para 29,60 (sonar-p) e para 32,53 (sonar-t). Isso significa um aumento de 1,35% e 10,24% no tamanho da árvore, respectivamente. Como pode ser observado o tamanho da árvore aumentou para todos os conjuntos, exceto sonar-p3, embora não de forma significativa.

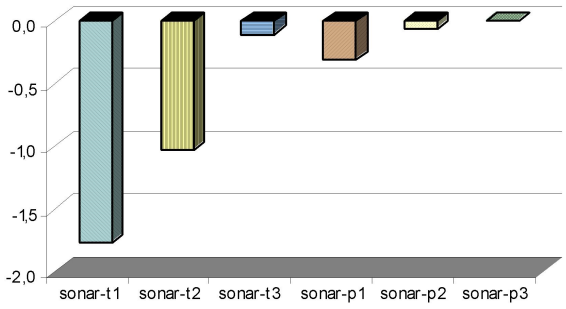

Figura 4: Diferença absoluta do tamanho do classificador sonar

3.2 Resultados ionosphere

Na Tabela 6 é mostrado o número de valores distintos para cada atributo tanto no conjunto original ionosphere, como nos derivados ionosphere-t (ionosphere-t4, t3, t2, t1).

Número do Atributo	Nome do Atributo	#Valores Distintos (vlr. relativo) ionosphere	#Valores Distintos (vlr. relativo) ionosphere-t4	#Valores Distintos (vlr. relativo) ionosphere-t3	#Valores Distintos (vlr. relativo) ionosphere-t2	#Valores Distintos (vlr. relativo) ionosphere-t1
#1	a01	2 (0,00)	2 (0,00)	2 (0,00)	2 (0,00)	2 (0,00)
#2	a02	1 (0,00)	1 (0,00)	1 (0,00)	1 (0,00)	1 (0,00)
#3	a03	219 (0,62)	216 (0,61)	184 (0,52)	79 (0,22)	17 (0,04)
#4	a04	269 (0,76)	265 (0,75)	228 (0,64)	92 (0,26)	18 (0,05)
#5	a05	204 (0,58)	201 (0,57)	177 (0,50)	81 (0,23)	17 (0,04)
#6	a06	259 (0,73)	257 (0,73)	226 (0,64)	104 (0,29)	20 (0,05)
#7	a07	231 (0,65)	227 (0,64)	204 (0,58)	106 (0,30)	17 (0,04)
#8	a08	260 (0,74)	257 (0,73)	226 (0,64)	116 (0,33)	20 (0,05)
#9	a09	244 (0,69)	241 (0,68)	220 (0,62)	110 (0,31)	19 (0,05)
#10	a10	267 (0,76)	261 (0,74)	222 (0,63)	111 (0,31)	18 (0,05)
#11	a11	246 (0,70)	245 (0,69)	225 (0,64)	114 (0,32)	20 (0,05)
#12	a12	269 (0,76)	266 (0,75)	235 (0,66)	118 (0,33)	20 (0,05)
#13	a13	238 (0,67)	238 (0,67)	214 (0,60)	119 (0,33)	21 (0,05)
#14	a14	266 (0,75)	261 (0,74)	235 (0,66)	114 (0,32)	21 (0,05)
#15	a15	234 (0,66)	234 (0,66)	217 (0,61)	127 (0,36)	21 (0,05)
#16	a16	270 (0,76)	266 (0,75)	235 (0,66)	112 (0,31)	20 (0,05)
#17	a17	254 (0,72)	251 (0,71)	237 (0,67)	118 (0,33)	21 (0,05)
#18	a18	280 (0,79)	279 (0,79)	258 (0,73)	135 (0,38)	20 (0,05)
#19	a19	254 (0,72)	250 (0,71)	233 (0,66)	130 (0,37)	21 (0,05)
#20	a20	266 (0,75)	260 (0,74)	228 (0,64)	127 (0,36)	21 (0,05)
#21	a21	248 (0,70)	246 (0,70)	232 (0,66)	129 (0,36)	21 (0,05)
#22	a22	265 (0,75)	261 (0,74)	236 (0,67)	124 (0,35)	21 (0,05)
#23	a23	248 (0,70)	245 (0,69)	234 (0,66)	128 (0,36)	21 (0,05)
#24	a24	264 (0,75)	261 (0,74)	235 (0,66)	131 (0,37)	21 (0,05)
#25	a25	256 (0,72)	253 (0,72)	236 (0,67)	129 (0,36)	21 (0,05)
#26	a26	273 (0,77)	271 (0,77)	246 (0,70)	126 (0,35)	21 (0,05)
#27	a27	256 (0,72)	256 (0,72)	234 (0,66)	120 (0,34)	20 (0,05)
#28	a28	281 (0,80)	276 (0,78)	244 (0,69)	123 (0,35)	21 (0,05)
#29	a29	244 (0,69)	243 (0,69)	217 (0,61)	116 (0,33)	20 (0,05)
#30	a30	266 (0,75)	263 (0,74)	240 (0,68)	117 (0,33)	21 (0,05)
#31	a31	243 (0,69)	239 (0,68)	216 (0,61)	113 (0,32)	20 (0,05)
#32	a32	263 (0,74)	262 (0,74)	239 (0,68)	127 (0,36)	21 (0,05)
#33	a33	245 (0,69)	245 (0,69)	220 (0,62)	115 (0,32)	21 (0,05)
#34	a34	263 (0,74)	260 (0,74)	229 (0,65)	120 (0,34)	21 (0,05)
Média		239,65	237,03	213,68	109,82	19,00

Tabela 6: Número de valores distintos dos atributos ionosphere

Como pode ser observado o arredondamento científico usando redução de casas decimais, diminui acentuadamente o número de valores distintos, principalmente para ionosphere-t1. Em média, de 239,65 (ionosphere) para 237,03 (ionosphere-t4), para 213,68 (ionosphere-t3), para 109,82 (ionosphere-t2) e para 19 (ionosphere-t1), (ou seja, 4, 3, 2 e 1 casas decimais), respectivamente. Isso corresponde a uma redução média de 1,09% de ionosphere para ionosphere-t4, de 10,84% de ionosphere para ionosphere-t3, de 54,17% de ionosphere para ionosphere-t2 e de 92,07% de ionosphere para ionosphere-t1.

Na Tabela 7 são mostrados os atributos que aparecem nas árvores induzidas, número de atributos, representado por (#A) e porcentagem do total de atributos, representado por (%A),

usando todo o conjunto de exemplos, tanto para o conjunto original como para os derivados para ionosphere. Analogamente, essa informação é mostrada de outra forma na Tabela 8.

Conjunto	Atributos	#A	%A
ionosphere	1, 3, 4, 5, 6, 7, 8, 10, 16, 17, 19, 21, 27, 28	14	41,18%
ionosphere-t4	1, 3, 4, 5, 6, 7, 8, 10, 16, 17, 19, 21, 27, 28	14	41,18%
ionosphere-t3	1, 3, 5, 6, 8, 9, 16, 18, 21, 27, 28, 33	12	35,29%
ionosphere-t2	1, 3, 5, 6, 8, 12, 14, 16, 20, 21, 23, 27, 28	13	38,23%
ionosphere-t1	1, 3, 5, 6, 8, 16, 23, 24, 27, 28	10	29,41%

Tabela 7: Atributos que aparecem na árvore induzida ionosphere

Número do Atributo	ionosphere	ionosphere-t4	ionosphere-t3	ionosphere-t2	ionosphere-t1
#1	•	•	•	•	•
#2					
#3	•	•	•	•	•
#4	•	•			
#5	•	•	•	•	•
#6	•	•	•	•	•
#7	•	•			
#8	•	•	•	•	•
#9			•		
#10	•	•			
#11					
#12				•	
#13					
#14				•	
#15					
#16	•	•	•	•	•
#17	•	•			
#18			•		
#19	•	•			
#20				•	
#21	•	•	•	•	
#22					
#23				•	•
#24					•
#25					
#26					
#27	•	•	•	•	•
#28	•	•	•	•	•
#29					
#30					
#31					
#32					
#33			•		
#34					
Total 34 100%	14 41,18%	14 41,18%	12 35,29%	13 38,23%	10 29,41%

Tabela 8: Atributos selecionados pelo classificador ionosphere

Na Tabela 9 são mostrados os resultados obtidos em relação aos conjuntos de exemplos relacionados ao ionosphere usando o *10-fold stratified cross-validation*. Em média, o tempo de indução diminuiu de 0,21 (ionosphere) para 0,17 (ionosphere-t). Isso significa uma redução de 19,05% do tempo de indução.

Conjunto	Tempo (s)	Erro	Tamanho
ionosphere	$0,21 \pm 0,02$	$8,54 \pm 1,03$	$27,40 \pm 3,75$
ionosphere-t4	$0,20 \pm 0,02$	$8,54 \pm 1,03$	$27,40 \pm 3,75$
ionosphere-t3	$0,20 \pm 0,02$	$8,25 \pm 1,07$	$27,60 \pm 2,67$
ionosphere-t2	$0,15 \pm 0,01$	$8,83 \pm 1,37$	$28,00 \pm 3,16$
ionosphere-t1	$0,11 \pm 0,01$	$7,11 \pm 0,86$	$25,00 \pm 4,11$

Tabela 9: Tempo de indução, taxa de erro e tamanho do classificador ionosphere

Na Figura 5 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-t1, entre ionosphere e ionosphere-t2 e assim por diante. Analogamente para taxa de erro e tamanho da árvore mostrados nas Figuras 6 e 7, respectivamente.

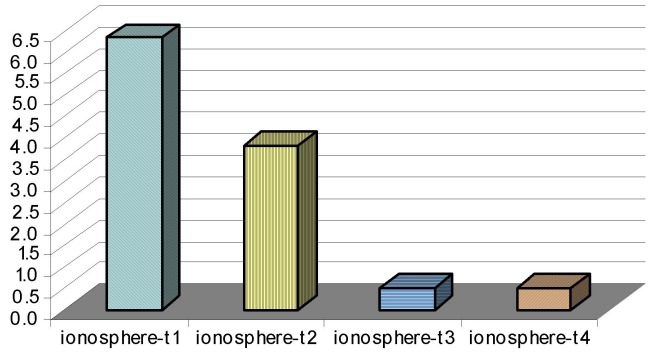

Figura 5: Diferença absoluta do tempo de indução ionosphere

Como esperado, o tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo significativa (com grau de confiança de 95%) para ionosphere-t1 e ionosphere-t2.

Na Figura 6 é mostrada a diferença absoluta em desvios padrões da taxa de erro. Em média, a taxa de erro diminuiu de 8,54% (ionosphere) para 8,18% (ionosphere-t). Isso significa uma redução de 4,22% da taxa de erro. Como pode ser observado, a taxa de erro reduziu, embora não de forma significativa, para todos os conjuntos derivados, exceto ionosphere-t2 que aumentou (de forma não significativa) e ionosphere-t4 que se manteve constante.

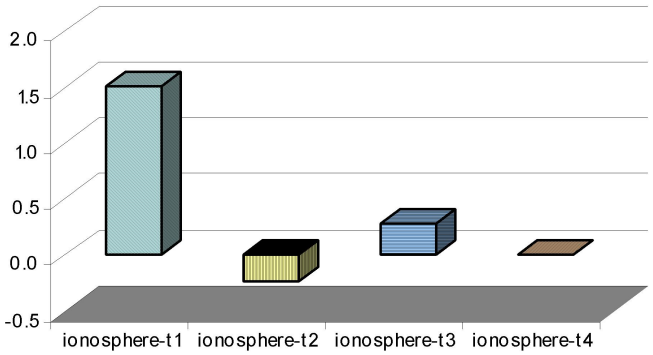

Figura 6: Diferença absoluta da taxa de erro ionosphere

Na Figura 7 é mostrada a diferença absoluta em desvios padrões do tamanho da árvore. Em média, o tamanho da árvore diminuiu de 27,40 (ionosphere) para 27,00 (ionosphere-t). Isso significa uma redução de 1,46% do tamanho da árvore. Como pode ser observado o tamanho da árvore diminuiu para o conjunto ionosphere-t1, aumentou para ionosphere-t2 e ionosphere-t3 e se manteve constante em ionosphere-t4.

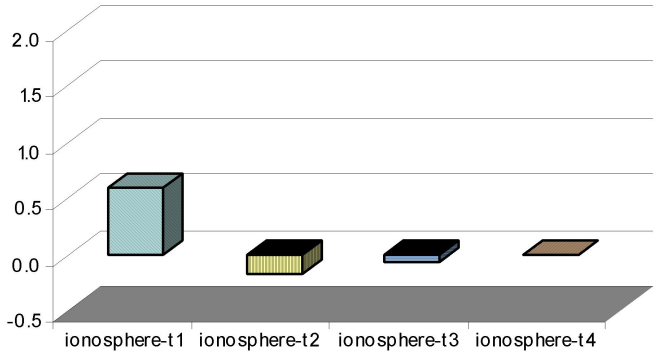

Figura 7: Diferença absoluta do tamanho do classificador ionosphere

3.3 Resultados vowel

Na Tabela 10 é mostrado o número de valores distintos para cada atributo tanto no conjunto original vowel, como nos derivados vowel-t (vowel-t2, t1).

Número do Atributo	Nome do Atributo	#Valores Distintos (vlr. relativo) vowel	#Valores Distintos (vlr. relativo) vowel-t2	#Valores Distintos (vlr. relativo) vowel-t1
#1	a01	2 (0,00)	2 (0,00)	2 (0,00)
#2	a02	15 (0,01)	15 (0,01)	15 (0,01)
#3	a03	2 (0,00)	2 (0,00)	2 (0,00)
#4	a04	853 (0,86)	335 (0,33)	44 (0,04)
#5	a05	877 (0,88)	414 (0,41)	60 (0,06)
#6	a06	815 (0,82)	286 (0,28)	39 (0,03)
#7	a07	836 (0,84)	296 (0,29)	39 (0,03)
#8	a08	803 (0,81)	272 (0,27)	36 (0,03)
#9	a09	798 (0,80)	258 (0,26)	32 (0,03)
#10	a10	748 (0,75)	214 (0,21)	30 (0,03)
#11	a11	794 (0,80)	239 (0,24)	34 (0,03)
#12	a12	788 (0,79)	246 (0,24)	30 (0,03)
#13	a13	775 (0,78)	243 (0,24)	31 (0,03)
Média		623,54	217,08	30,31

Tabela 10: Número de valores distintos dos atributos vowel

Como pode ser observado o arredondamento científico usando redução de casas decimais, diminui muito o número de valores distintos. Em média, de 623,54 (vowel) para 217,08 (vowel-t2) e para 30,31 (vowel-t1), (ou seja, 2 e 1 casas decimais), respectivamente. Isso corresponde a uma redução média de 65,19% de vowel para vowel-t2 e de 95,14% de vowel para vowel-t1.

Na Tabela 11 são mostrados os atributos que aparecem nas árvores induzidas, número de atributos (#A) e porcentagem do total de atributos (%A), usando todo o conjunto de exemplos, tanto para o conjunto original como para os derivados para vowel. Analogamente, essa informação é mostrada de outra forma na Tabela 12. Como pode ser observado, não houve alteração nos atributos que aparecem nas árvores.

Conjunto	Atributos	#A	%A
vowel	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-t2	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-t1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%

Tabela 11: Atributos que aparecem na árvore induzida vowel

Número do Atributo	vowel	vowel-t2	vowel-t1
#1	•	•	•
#2	•	•	•
#3	•	•	•
#4	•	•	•
#5	•	•	•
#6	•	•	•
#7	•	•	•
#8	•	•	•
#9	•	•	•
#10	•	•	•
#11	•	•	•
#12	•	•	•
#13	•	•	•
Total 13 100%	13 100,00%	13 100,00%	13 100,00%

Tabela 12: Atributos selecionados pelo classificador vowel

Na Tabela 13 são mostrados os resultados obtidos em relação aos conjuntos de exemplos relacionados ao vowel usando o *10-fold stratified cross-validation*. Em média, o tempo de indução diminuiu de 0,50 (vowel) para 0,34 (vowel-t). Isso significa uma redução de 32,00% do tempo de indução.

Conjunto	Tempo (s)	Erro	Tamanho
vowel	$0,50 \pm 0,07$	$18,48 \pm 1,49$	$213,40 \pm 17,54$
vowel-t2	$0,46 \pm 0,31$	$20,51 \pm 1,35$	$216,90 \pm 16,39$
vowel-t1	$0,22 \pm 0,01$	$18,99 \pm 1,06$	$228,40 \pm 21,81$

Tabela 13: Tempo de indução, taxa de erro e tamanho do classificador vowel

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 vowel e vowel-t1, entre vowel e vowel-t2. Analogamente para taxa de erro e tamanho da árvore mostrados nas Figuras 9 e 10, respectivamente.

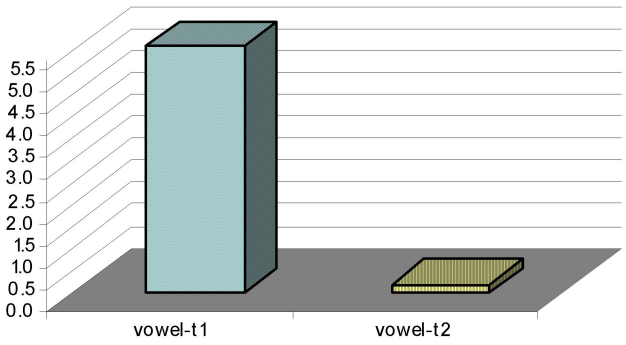

Figura 8: Diferença absoluta do tempo de indução vowel

Como esperado, o tempo de indução reduziu para os dois conjuntos derivados, sendo de forma significativa para vowel-t1.

Na Figura 9 é mostrada a diferença absoluta em desvios padrões da taxa de erro. Em média, a taxa de erro aumentou de 18,48% (vowel) para 19,75% (vowel-t). Isso significa um aumento de 6,87% da taxa de erro. Como pode ser observado, a taxa de erro aumentou para os dois

conjuntos derivados, embora não de forma significativa.

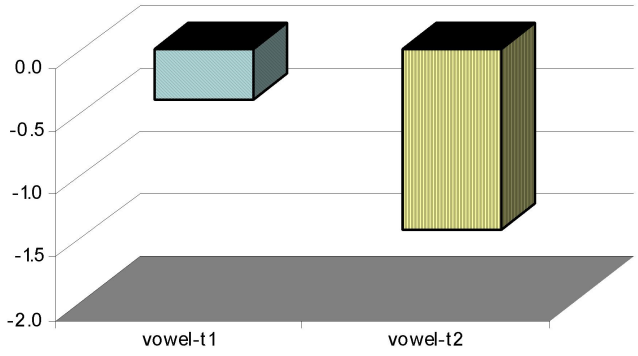

Figura 9: Diferença absoluta da taxa de erro vowel

Na Figura 10 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 213,40 (vowel) para 222,65 (vowel-t). Isso significa um aumento de 4,33% do tamanho da árvore. Como pode ser notado o tamanho da árvore aumentou para os dois os conjuntos, embora não de forma significativa.

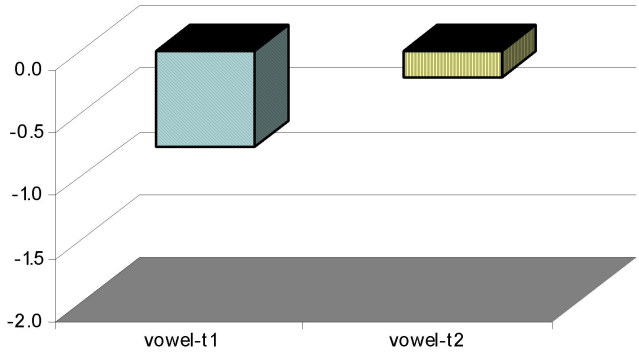

Figura 10: Diferença absoluta do tamanho do classificador vowel

4 Experimento 2

Como já mencionado, no Experimento 1 a taxa de erro obtida da forma proposta pode ter um *bias* otimista, já que os exemplos em todos os *folds* tiveram seus valores arredondados[3].

Assim sendo, no Experimento 2 foi avaliada a taxa de erro utilizando *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 taxa de erro, com o objetivo de excluir o *bias* otimista.

O 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) foi avaliada a taxa de erro do

[3]Esse *bias* otimista não se aplica nem ao tempo de indução e nem ao tamanho do classificador, portanto, os valores das métricas de tempo de indução e tamanho do classificador para este Experimento 2 são os mesmos daqueles reportados no Experimento 1.

classificador. Esse processo foi repetido um total de 10 vezes, cada vez utilizando um *fold* diferente de teste (sem arredondamento).

Para tornar clara a distinção entre a metodologia utilizada neste Experimento 2 daquela utilizada no Experimento 1, será utilizado o termo *arredondamento parcial* para se referir ao arredondamento foi aplicado apenas ao conjunto de treinamento mas não ao conjunto de teste — Experimento 2 — e o termo *arredondamento completo* para se referir ao arredondamento aplicado tanto ao conjunto de treinamento como ao conjunto de teste — Experimento 1.

4.1 Resultados sonar

Na Tabela 14 são mostrados os resultados (média ± desvio padrão) obtidos em relação aos conjuntos de exemplos sonar original e derivados. A segunda coluna representa os resultados da taxa de erro já mostrados na Tabela 5, utilizando o arredondamento completo. A terceira coluna representa a taxa de erro utilizando o arredondamento parcial.

Conjunto	Erro (arredondando conj. treinamento e teste)	Erro (arredondando apenas conj. treinamento)
sonar	$28,83 \pm 2,24$	$28,83 \pm 2,24$
sonar-t3	$27,40 \pm 2,48$	$31,76 \pm 2,09$
sonar-t2	$25,98 \pm 3,13$	$28,81 \pm 2,55$
sonar-t1	$23,98 \pm 3,65$	$31,21 \pm 1,98$

Tabela 14: Taxa de erro do arredondamento completo e do arredondamento parcial sonar

Na Figura 11 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico para arredondamento parcial *versus* conjunto original. Em média, a taxa de erro aumentou de 28,83% (sonar), para 30,59% (sonar-t). Isso significa um aumento de 6,10% da taxa de erro. Como pode ser observado, a taxa de erro aumentou para todos os conjuntos utilizando arredondamento parcial, exceto para sonar-t2, embora não de forma significativa.

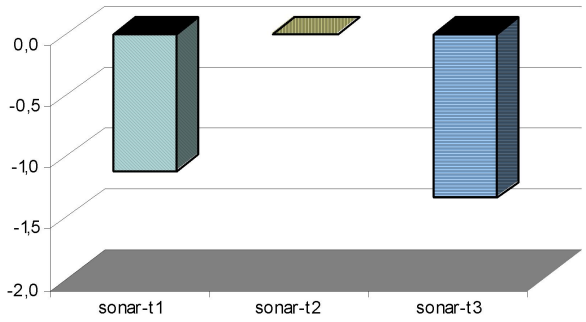

Figura 11: Diferença absoluta da taxa de erro (arredondamento parcial *versus* conjunto original) sonar

Na Figura 12 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico para arredondamento parcial *versus* arredondamento completo. A taxa de erro aumentou de 23,98% para 31,21% em (sonar-t1), de 25,98% para 28,81% em (sonar-t2), e de 27,40% para 31,76% em (sonar-t3). Isso significa um aumento de 30,15% em (sonar-t1), 10,89% em (sonar-t2), 15,91% em (sonar-t3) da taxa de erro. Nota-se, portanto, a confirmação do *bias* otimista, que é significativo no caso de sonar-t1.

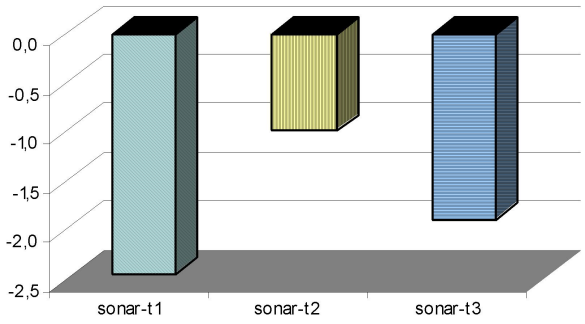

Figura 12: Diferença absoluta da taxa de erro (arredondamento parcial *versus* arredondamento completo) sonar

4.2 Resultados ionosphere

Na Tabela 15 são mostrados os resultados (média ± desvio padrão) obtidos em relação aos conjuntos de exemplos ionosphere original e derivados.

Conjunto	Erro (arredondando conj. treinamento e teste)	Erro (arredondando apenas conj. treinamento)
ionosphere	$8,54 \pm 1,03$	$8,54 \pm 1,03$
ionosphere-t4	$8,54 \pm 1,03$	$6,86 \pm 2,34$
ionosphere-t3	$8,25 \pm 1,07$	$7,96 \pm 2,24$
ionosphere-t2	$8,83 \pm 1,37$	$2,85 \pm 1,95$
ionosphere-t1	$7,11 \pm 0,86$	$4,82 \pm 1,99$

Tabela 15: Taxa de erro do arredondamento completo e taxa de erro do arredondamento parcial ionosphere

Na Figura 13 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico para arredondamento parcial *versus* conjunto original. Em média, a taxa de erro diminuiu de 8,54% (ionosphere), para 5,62% (ionosphere-t). Isso significa uma redução de 34,19% da taxa de erro. Como pode ser observado, a taxa de erro diminuiu para todos os conjuntos utilizando arredondamento no conjunto de treinamento e deixando do conjunto de teste intacto, embora de forma significativa apenas para os conjuntos ionosphere-t1 e ionosphere-t2.

Na Figura 14 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico para arredondamento parcial *versus* arredondamento completo. A taxa de erro diminuiu de 7,11% para 4,82% em (ionosphere-t1), de 8,83% para 2,85% em (ionosphere-t2), de 8,25% para 7,96% em (ionosphere-t3), e de 8,54% para 6,86% em (ionosphere-t4). Isso significa uma redução de 32,21% em (ionosphere-t1), 67,72% em (ionosphere-t2), 3,51% em (ionosphere-t3), e de 19,67% em (ionosphere-t4). Para ionosphere-t2 há uma diferença significativa.

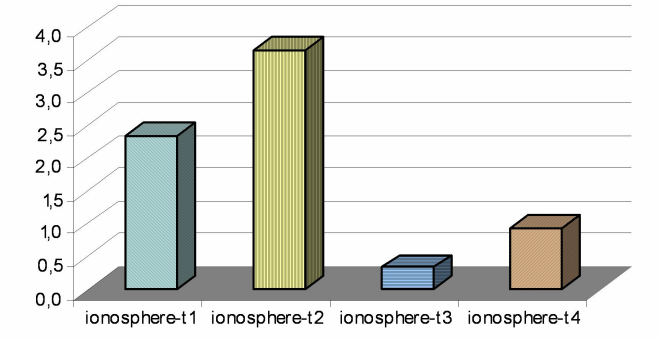

Figura 13: Diferença absoluta da taxa de erro (arredondamento parcial *versus* conjunto original) ionosphere

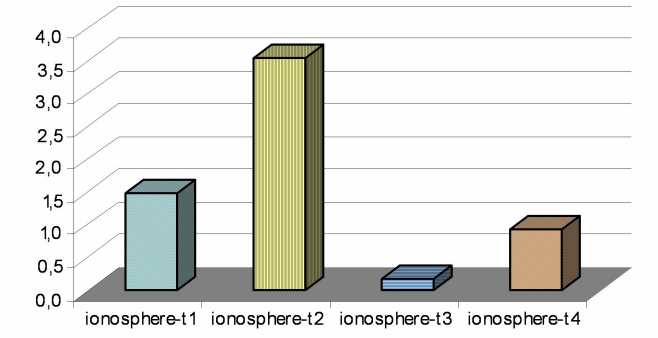

Figura 14: Diferença absoluta da taxa de erro (arredondamento parcial *versus* arredondamento completo) ionosphere

4.3 Resultados vowel

Na Tabela 16 são mostrados os resultados (média ± desvio padrão) obtidos em relação aos conjuntos de exemplos vowel original e derivados.

Conjunto	Erro (arredondando conj. treinamento e teste)	Erro (arredondando apenas conj. treinamento)
vowel	$18,48 \pm 1,49$	$18,48 \pm 1,49$
vowel-t2	$20,51 \pm 1,35$	$66,26 \pm 1,69$
vowel-t1	$18,99 \pm 1,06$	$64,04 \pm 1,69$

Tabela 16: Taxa de erro do arredondamento completo e taxa de erro do arredondamento parcial vowel

Na Figura 15 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico para arredondamento parcial *versus* conjunto original. Em média, a taxa de erro aumentou de 18,48% (vowel), para 65,15% (vowel-t). Isso significa um aumento de 252,54% da taxa de erro. Como pode ser observado, a taxa de erro aumentou para todos os conjuntos utilizando arredondamento no conjunto de treinamento e deixando do conjunto de teste intacto, de forma significativa.

Na Figura 16 é mostrada a diferença absoluta em desvios padrões da taxa de erro no eixo vertical do gráfico para arredondamento parcial *versus* arredondamento completo. A taxa de

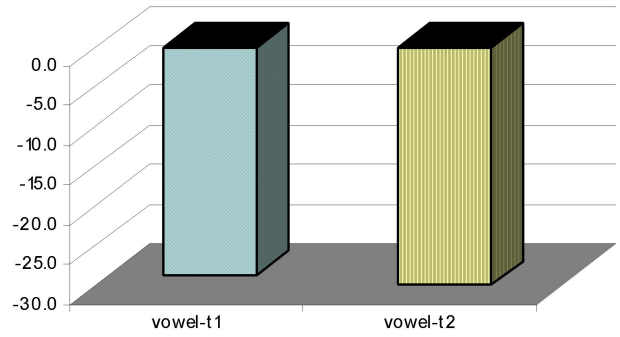

Figura 15: Diferença absoluta da taxa de erro (arredondamento parcial *versus* conjunto original) vowel

erro aumentou de 18,99% para 64,04% em (vowel-t1), e de 20,51% para 66,26% em (vowel-t2). Isso significa um aumento de 237,23% em (vowel-t1) e de 223,06% em (vowel-t2) da taxa de erro, ambos significativos.

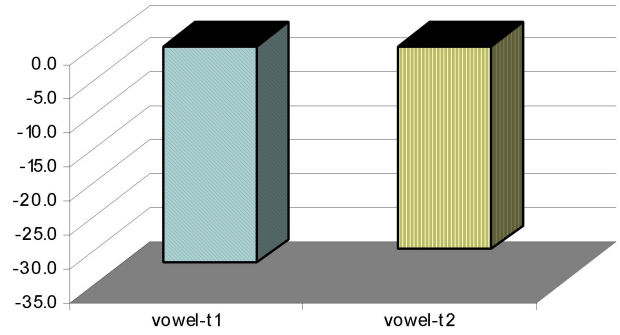

Figura 16: Diferença absoluta da taxa de erro (arredondamento parcial *versus* arredondamento completo) vowel

5 Algoritmo de Arredondamento

Além da noção básica de arredondamento científico, usando redução de casas decimais, neste trabalho também foi avaliada uma outra técnica de arredondamento proposta por Weiss & Indurkhya (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.

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

Na Tabela 17 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 17: 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.

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

Na Tabela 18 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 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}

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 18: Exemplo utilizando a Equação 2 na base 2

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 & Indurkhya (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 & 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.

Como estes algoritmos permitem que seja escolhida uma quantidade max de valores distintos para cada atributo, espera-se uma redução menos acentuada do que aquela simplesmente obtida utilizando arredondamento científico (usando redução de casas decimais), principalmente considerando o fato que ao invés de utilizar um valor absoluto para max, é possível utilizar um valor relativo, ou seja, em termos de porcentagem de valores diferentes.

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

6 Experimento 3

Neste experimento foram avaliados tempo de indução, taxa de erro, e 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 conjuntos de exemplos sonar, ionosphere, vowel e wine. Para o conjunto aml-all, seguindo a metodologia utilizada originalmente nesse conjunto por Golub (1999) e, posteriormente, também utilizada por Gamberger, Lavrac, Zelezny & Tolar (2004), foi utilizado *holdout*[4].

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. Esse processo foi repetido um total de 10 vezes, cada vez utilizando um *fold* diferente de teste (sem arredondamento) para os conjuntos de exemplos sonar, ionosphere, vowel e wine. Para os exemplos aml-all, no conjunto de treinamento contendo 38 exemplos foi aplicado arredondamento dos valores, deixando intacto o conjunto de teste que contém 34 exemplos.

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

Como já mencionado na Seção 5, 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% e 50%, 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% e sonar-50%, respectivamente. De forma análoga essa notação é utilizada para os demais conjuntos de exemplos.

Por exemplo, na Figura 17 é mostrado no número de valores distintos para sonar, sonar-t3, sonar-t2 e sonar-t1 nos quais foi utilizado arredondamento científico, descrito na Seção 3. Como é possível notar, há uma redução acentuada de um conjunto de exemplos em relação a outro.

Na Figura 18 é mostrado o número de valores distintos para sonar, sonar-90%, sonar-80%, sonar-70%, sonar-60%, sonar-50%, nos quais foi utilizado o arredondamento proposto por Weiss, descrito na Seção 5.

Adicionalmente ao parâmetro p, os Algoritmo 1 e 2 também possuem o parâmetro b, que corresponde a 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).

[4]Para uma revisão sobre métodos de amostragem e de avaliação de algoritmos vide Rezende (2003)[Cap. 4].

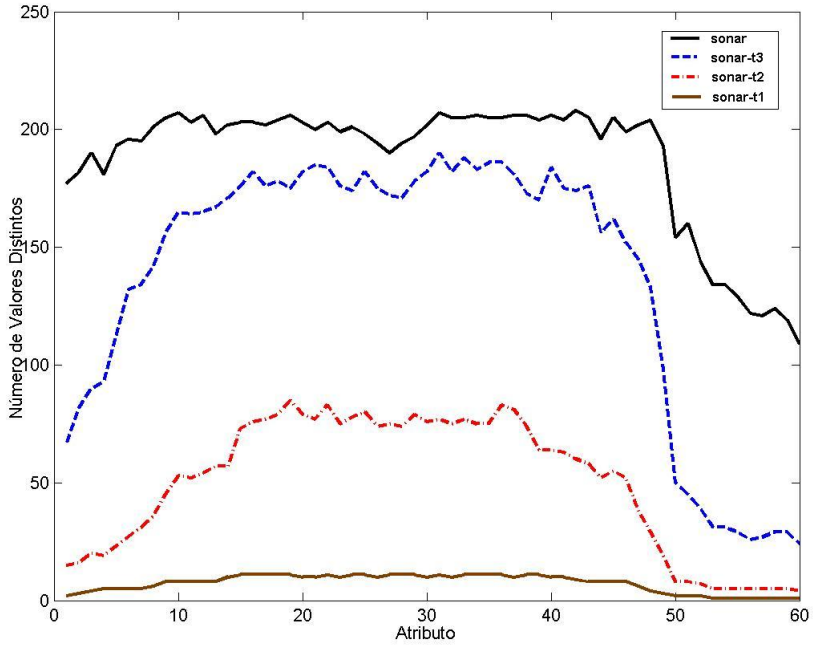

Figura 17: Número de valores distintos para sonar e seus conjuntos derivados pelo arredondamento científico

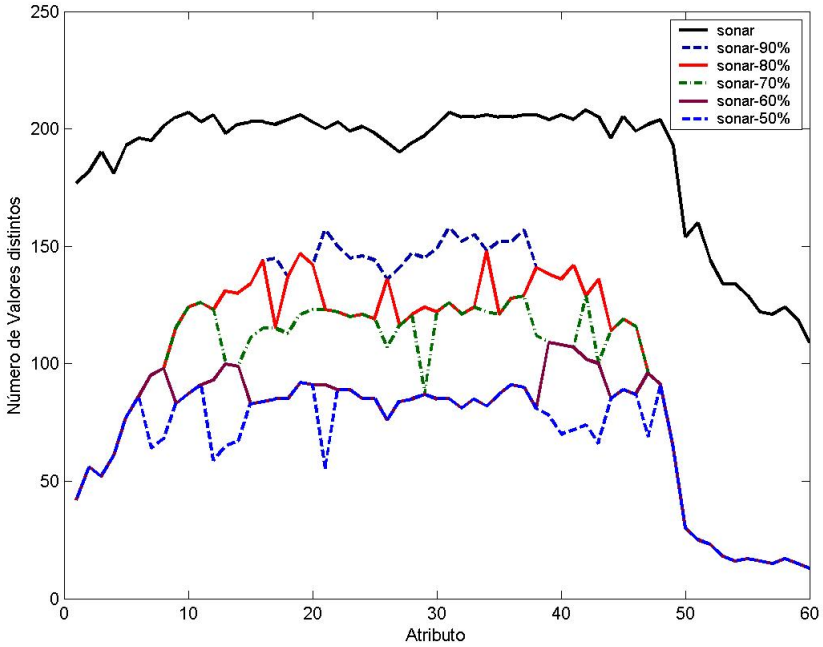

Figura 18: Número de valores distintos para sonar e seus conjuntos derivados pelo arredondamento proposto por Weiss

6.1 Resultados sonar

Nas Tabelas 19 e 20 são mostrados os atributos que aparecem nas árvores induzidas, utilizando todo o conjunto de exemplos número de atributos (#A) e porcentagem do total de atributos (%A), tanto para o conjunto original como para os derivados (utilizando o arredondamento proposto no Algoritmo 1) para sonar, utilizando bases 2 e 10, respectivamente.

Conjunto	Atributos	#A	%A
sonar	1, 2, 4, 8, 11, 18, 21, 23, 27, 28, 51, 53, 54	13	21,67%
sonar-90%	1, 2, 4, 8, 11, 18, 21, 23, 27, 28, 51, 53, 54	13	21,67%
sonar-80%	1, 2, 4, 8, 11, 18, 21, 23, 27, 28, 51, 53, 54	13	21,67%
sonar-70%	1, 2, 4, 8, 9, 11, 27, 28, 36, 39, 43, 45, 51, 54, 57	15	25,00%
sonar-60%	1, 2, 4, 7, 8, 11, 21, 27, 28, 34, 39, 43, 45, 51, 52, 58	16	26,70%
sonar-50%	4, 11, 17, 20, 23, 25, 36, 42, 45, 46, 50, 51, 54	13	21,67%

Tabela 19: Atributos que aparecem na árvore induzida sonar - arredondamento utilizando o Algoritmo 1 com base 2

Conjunto	Atributos	#A	%A
sonar	1, 2, 4, 8, 11, 18, 21, 23, 27, 28, 51, 53, 54	13	21,67%
sonar-90%	1, 2, 4, 8, 11, 21, 27, 28, 31, 36, 43, 50, 54	13	21,67%
sonar-80%	1, 2, 4, 7, 8, 11, 15, 21, 27, 28, 31, 33, 39, 43, 51, 52	16	26,70%
sonar-70%	1, 2, 4, 8, 11, 21, 27, 28, 29, 37, 43, 45, 55	13	21,67%
sonar-60%	1, 2, 4, 8, 11, 21, 27, 28, 29, 37, 43, 45, 55	13	21,67%
sonar-50%	1, 2, 4, 8, 11, 21, 27, 28, 29, 37, 43, 45, 55	13	21,67%

Tabela 20: Atributos que aparecem na árvore induzida sonar - arredondamento utilizando o Algoritmo 1 com base 10

Como pode ser notado, os atributos selecionados para sonar-90% e sonar-80% foram idênticos aos selecionados para sonar para a base 2. Além disso, há uma diminuição gradativa na intersecção entre os conjuntos de atributos sonar-70% e sonar, sonar-60% e sonar.

Na Tabela 21 são mostrados os resultados (média ± 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	$0,22 \pm 0,11$	$0,22 \pm 0,11$	$28,83 \pm 2,24$	$28,83 \pm 2,24$	$29,20 \pm 3,58$	$29,20 \pm 3,58$
sonar-90%	$0,17 \pm 0,02$	$0,16 \pm 0,02$	$27,92 \pm 3,89$	$28,81 \pm 2,55$	$26,20 \pm 2,53$	$32,20 \pm 2,53$
sonar-80%	$0,17 \pm 0,02$	$0,14 \pm 0,02$	$26,98 \pm 4,11$	$28,81 \pm 2,55$	$26,20 \pm 2,53$	$32,20 \pm 2,53$
sonar-70%	$0,16 \pm 0,02$	$0,13 \pm 0,01$	$26,98 \pm 4,11$	$28,81 \pm 2,55$	$26,20 \pm 2,53$	$32,20 \pm 2,53$
sonar-60%	$0,15 \pm 0,02$	$0,15 \pm 0,04$	$25,48 \pm 3,62$	$28,81 \pm 2,55$	$26,60 \pm 3,37$	$32,20 \pm 2,53$
sonar-50%	$0,13 \pm 0,01$	$0,13 \pm 0,02$	$25,95 \pm 3,19$	$28,81 \pm 2,55$	$26,40 \pm 3,53$	$32,20 \pm 2,53$

Tabela 21: Tempo de indução, taxa de erro e tamanho do classificador utilizando arredondamento com bases 2 e 10 sonar

Na Figura 19 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.

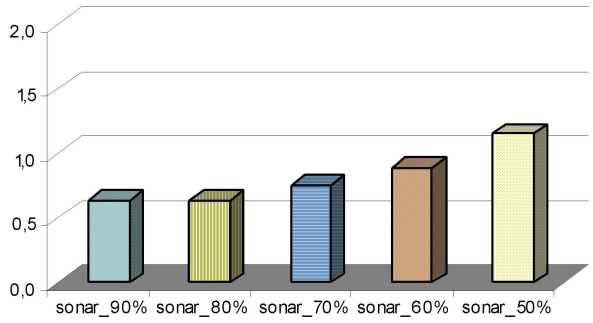

Figura 19: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 *versus* conjunto original) sonar

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 sonar e sonar-90%, entre sonar e sonar-80% e assim por diante, utilizando base 10.

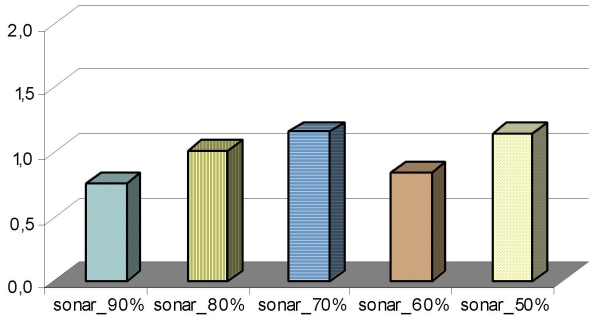

Figura 20: Diferença absoluta do tempo de indução (arredondamento utilizando base 10 *versus* conjunto original) sonar

Como esperado, 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 às Figuras 19 e 20, são mostrados os resultados para taxa de erro nas Figuras 21 e 22, e para tamanho do classificador nas Figuras 23 e 24.

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 diminuiu de 28,83% (sonar), para 26,66% utilizando o arredondamento com a base 2. Isso significa uma redução de 7,53% da taxa de erro.

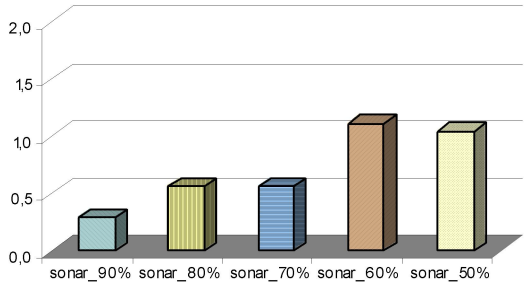

Figura 21: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 *versus* conjunto original) sonar

Na Figura 22 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 28,83% (sonar), para 28,81% dos conjuntos com base 10. Isso significa que a taxa de erro praticamente foi a mesma.

Como pode ser observado, a taxa de erro reduziu para todos os conjuntos utilizando arredondamento com ambas as bases, embora não de forma significativa.

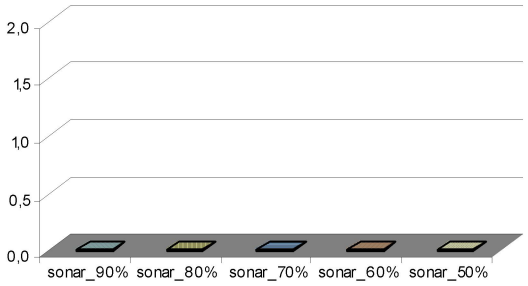

Figura 22: Diferença absoluta da taxa de erro (arredondamento utilizando base 10 *versus* conjunto original) sonar

A partir desse ponto, o termo sonar-base2 será utilizado para descrever todos os conjuntos derivados de sonar, ou seja, sonar-90%, sonar-80%, sonar-70%, sonar-60%, sonar-50%, que foram gerados utilizando a base 2. De forma similar, o termo sonar-base10 será utilizado para descrever todos os conjuntos derivados de sonar, ou seja, sonar-90%, sonar-80%, sonar-70%, sonar-60%, sonar-50%, que foram gerados utilizando a base 10. Analogamente para os demais conjuntos de exemplos.

Nas Figuras 23 e 24 são mostradas as diferenças absolutas em desvios padrões do tamanho da árvore no eixo vertical do gráfico, utilizando arredondamento com base 2 (binária) e base 10 (decimal), respectivamente. Em média, o tamanho da árvore diminuiu de 29,20 (sonar) para 26,32 (sonar-base2) — média aritmética dos conjuntos arredondados utilizando base 2 — e aumentou para 32,20 (sonar-base10) — média aritmética dos conjuntos arredondados utilizando base 10. Isso significa uma redução de 9,86% para (sonar-base2) e um aumento de 10,27% para (sonar-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore diminuiu para todos os conjuntos de (sonar-base2), e aumentou para todos os conjuntos de (sonar-base10), embora não de forma significativa para ambas as bases.

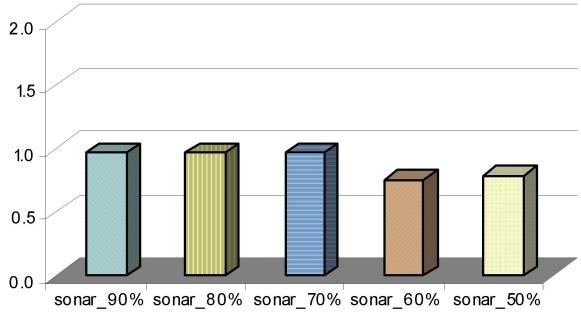

Figura 23: Diferença absoluta do tamanho do classificador (arredondamento utilizando base 2 *versus* conjunto original) sonar

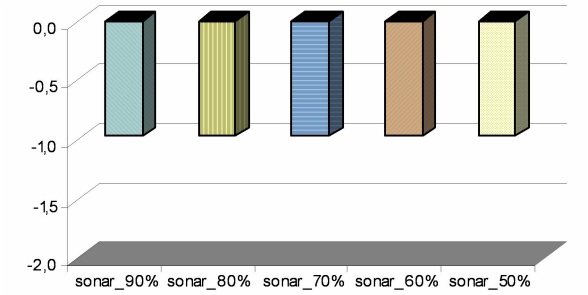

Figura 24: Diferença absoluta do tamanho do classificador (arredondamento utilizando base 10 *versus* conjunto original) sonar

6.2 Resultados ionosphere

Nas Tabelas 22 e 23 são mostrados os atributos que aparecem nas árvores induzidas, número de atributos (#A) e porcentagem do total de atributos (%A), tanto para o conjunto original como para os derivados (utilizando o arredondamento proposto no Algoritmo 1) para ionosphere, utilizando bases 2 e 10, respectivamente.

Conjunto	Atributos	#A	%A
ionosphere	1, 3, 4, 5, 6, 7, 8, 10, 16, 17, 19, 21, 27, 28	14	41,18%
ionosphere-90%	3, 4, 5, 6, 8, 15, 19, 25, 28, 30	10	29,41%
ionosphere-80%	3, 4, 5, 6, 8, 15, 16, 19, 25, 30, 32	11	32,35%
ionosphere-70%	3, 4, 5, 6, 7, 8, 14, 15, 19, 26	10	29,41%
ionosphere-60%	3, 4, 5, 6, 8, 9, 14, 16, 17, 32, 34	11	32,35%
ionosphere-50%	3, 4, 5, 6, 7, 8, 14, 15, 19, 26	10	29,41%

Tabela 22: Atributos que aparecem na árvore induzida ionosphere - arredondamento utilizando o Algoritmo 1 com base 2

Analogamente à Tabela 21, na Tabela 24 são mostrados os resultados (média ± 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

Conjunto	Atributos	#A	%A
ionosphere	1, 3, 4, 5, 6, 7, 8, 10, 16, 17, 19, 21, 27, 28	14	41,18%
ionosphere-90%	3, 5, 6, 7, 14, 18, 27, 30, 34	9	29,41%
ionosphere-80%	3, 4, 5, 6, 8, 11, 15, 19, 24, 31	10	32,35%
ionosphere-70%	3, 4, 5, 6, 8, 11, 15, 19, 24, 31, 45	11	29,41%
ionosphere-60%	3, 4, 5, 6, 8, 11, 15, 19, 24, 31	10	32,35%
ionosphere-50%	3, 4, 5, 6, 8, 11, 15, 19, 24, 31	10	29,41%

Tabela 23: Atributos que aparecem na árvore induzida ionosphere - arredondamento utilizando o Algoritmo 1 com base 10

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)	Erro (base 2)	Erro (base 10)	Tamanho (base 2)	Tamanho (base 10)
ionosphere	$0,21 \pm 0,02$	$0,21 \pm 0,02$	$8,54 \pm 1,03$	$8,54 \pm 1,03$	$27,40 \pm 3,75$	$27,40 \pm 3,75$
ionosphere-90%	$0,20 \pm 0,02$	$0,18 \pm 0,02$	$11,40 \pm 1,20$	$10,54 \pm 1,28$	$22,00 \pm 4,45$	$23,20 \pm 4,16$
ionosphere-80%	$0,18 \pm 0,02$	$0,17 \pm 0,05$	$11,40 \pm 1,20$	$9,13 \pm 1,19$	$22,20 \pm 4,92$	$23,40 \pm 5,23$
ionosphere-70%	$0,17 \pm 0,01$	$0,15 \pm 0,01$	$10,84 \pm 1,20$	$9,13 \pm 1,19$	$23,20 \pm 4,94$	$23,40 \pm 5,23$
ionosphere-60%	$0,15 \pm 0,02$	$0,15 \pm 0,01$	$11,13 \pm 1,38$	$9,13 \pm 1,19$	$24,40 \pm 3,78$	$23,40 \pm 5,23$
ionosphere-50%	$0,14 \pm 0,01$	$0,14 \pm 0,01$	$10,69 \pm 1,17$	$8,84 \pm 1,31$	$25,60 \pm 4,99$	$22,40 \pm 5,58$

Tabela 24: Tempo de indução, taxa de erro e tamanho do classificador utilizando arredondamento com bases 2 e 10 ionosphere

colunas representam os resultados do tamanho do classificador, utilizando a base binária e a base decimal, respectivamente.

Na Figura 25 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.

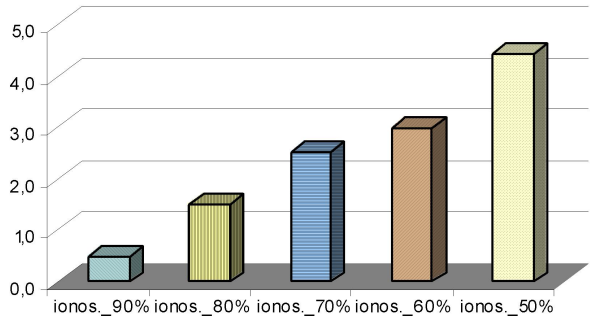

Figura 25: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 *versus* conjunto original) ionosphere

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 ionosphere e ionosphere-90%, entre ionosphere e ionosphere-80% e assim por diante, utilizando base 10.

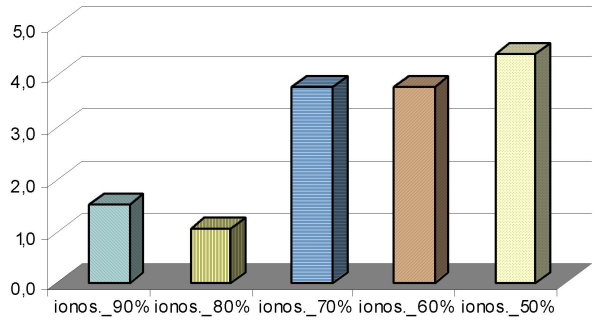

Figura 26: Diferença absoluta do tempo de indução (arredondamento utilizando base 10 *versus* conjunto original) ionosphere

Como esperado, o tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo de forma significativa (com grau de confiança de 95%), tanto para base 2 como para a base 10, exceto para os conjuntos ionosphere-90% e ionosphere-80% utilizando arredondamento em ambas as bases.

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 aumentou de 8,54% (ionosphere), para 11,29% dos conjuntos com base 2. Isso significa um aumento de 32,20% da taxa de erro.

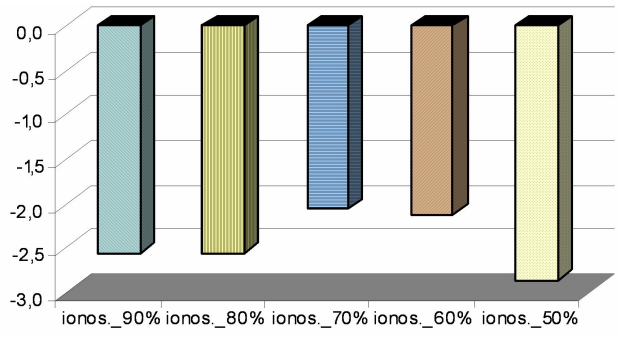

Figura 27: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 *versus* conjunto original) ionosphere

Na Figura 28 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 8,54% (ionosphere), para 9,35% dos conjuntos com base 10. Isso significa um aumento de 9,48% da taxa de erro.

Como pode ser observado, a taxa de erro aumentou para todos os conjuntos utilizando arredondamento com base 2, de forma significativa para todos os conjuntos derivados. E aumentou para todos os conjuntos utilizando arredondamento com base 10, embora de forma não significativa.

Nas Figuras 29 e 30 são mostradas as diferenças absolutas em desvios padrões do tamanho da árvore no eixo vertical do gráfico, utilizando arredondamento com base 2 e base 10, respectivamente. Em média, o tamanho da árvore diminuiu de 27,40 (ionosphere) para 23,48 (ionosphere-base2) e para 23,16 (ionosphere-base10). Isso significa uma redução de 14,31% para (ionosphere-base2) e de 15,47% para (ionosphere-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore diminuiu para todos os conjuntos (ionosphere-base2) e (ionosphere-base10), embora não de forma significativa.

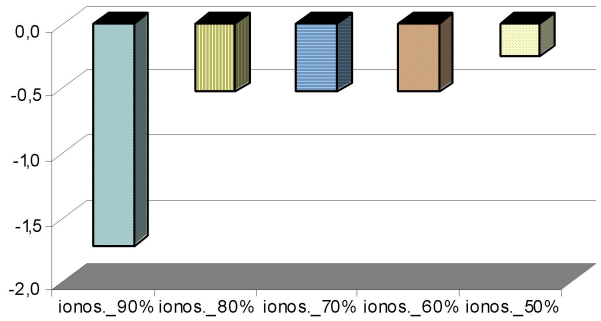

Figura 28: Diferença absoluta da taxa de erro (arredondamento utilizando base 10 *versus* conjunto original) ionosphere

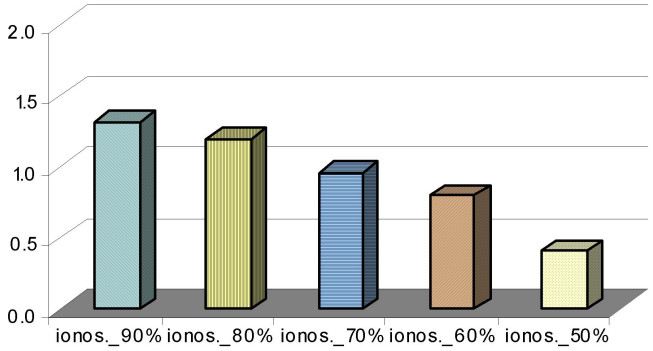

Figura 29: Diferença absoluta do tamanho do classificador (arredondamento utilizando base 2 *versus* conjunto original) ionosphere

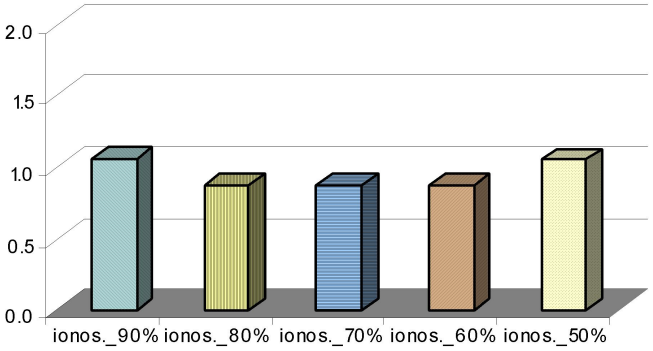

Figura 30: Diferença absoluta do tamanho do classificador (arredondamento utilizando base 10 *versus* conjunto original) ionosphere

6.3 Resultados vowel

Nas Tabelas 25 e 26 são mostrados os atributos que aparecem nas árvores induzidas, número de atributos (#A) e porcentagem do total de atributos (%A), tanto para o conjunto original como para os derivados (utilizando o arredondamento proposto no Algoritmo 1) para vowel,

utilizando bases 2 e 10, respectivamente. Nota-se que todos os atributos foram selecionados em ambas as bases.

Conjunto	Atributos	#A	%A
vowel	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-90%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-80%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-70%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-60%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-50%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%

Tabela 25: Atributos que aparecem na árvore induzida vowel - arredondamento utilizando o Algoritmo 1 com base 2

Conjunto	Atributos	#A	%A
vowel	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-90%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-80%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-70%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-60%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%
vowel-50%	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	13	100,00%

Tabela 26: Atributos que aparecem na árvore induzida vowel - arredondamento utilizando o Algoritmo 1 com base 10

Analogamente à Tabela 21, na Tabela 27 são mostrados os resultados (média ± 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	$0,50 \pm 0,07$	$0,50 \pm 0,07$	$18,48 \pm 1,49$	$18,48 \pm 1,49$	$213,40 \pm 17,54$	$213,40 \pm 17,54$
vowel-90%	$0,49 \pm 0,05$	$0,40 \pm 0,05$	$19,39 \pm 0,87$	$18,89 \pm 0,85$	$224,00 \pm 22,15$	$223,40 \pm 21,05$
vowel-80%	$0,50 \pm 0,05$	$0,42 \pm 0,06$	$19,70 \pm 0,89$	$18,89 \pm 0,85$	$214,72 \pm 22,39$	$223,40 \pm 21,05$
vowel-70%	$0,45 \pm 0,05$	$0,43 \pm 0,07$	$19,29 \pm 1,06$	$18,89 \pm 0,85$	$218,20 \pm 23,71$	$223,40 \pm 21,05$
vowel-60%	$0,46 \pm 0,07$	$0,40 \pm 0,06$	$19,29 \pm 0,70$	$18,89 \pm 0,85$	$216,40 \pm 23,23$	$223,40 \pm 21,05$
vowel-50%	$0,41 \pm 0,05$	$0,40 \pm 0,05$	$19,60 \pm 0,71$	$19,09 \pm 0,91$	$218,80 \pm 20,84$	$224,00 \pm 20,41$

Tabela 27: Tempo de indução, taxa de erro e tamanho do classificador utilizando arredondamento com bases 2 e 10 vowel

Na Figura 31 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.

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 vowel e vowel-90%, entre vowel e vowel-80% e assim por diante, utilizando base 10.

Como esperado, 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

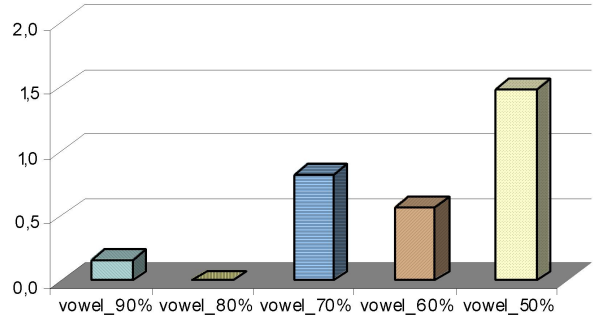

Figura 31: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 *versus* conjunto original) vowel

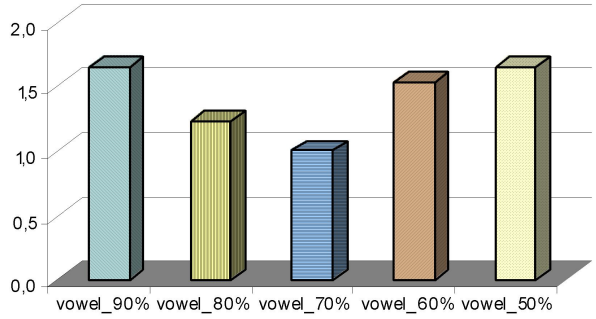

Figura 32: Diferença absoluta do tempo de indução (arredondamento utilizando base 10 *versus* conjunto original) vowel

como para a base 10.

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 18,48% (vowel), para 19,45% dos conjuntos com base 2. Isso significa um aumento de 5,25% da taxa de erro.

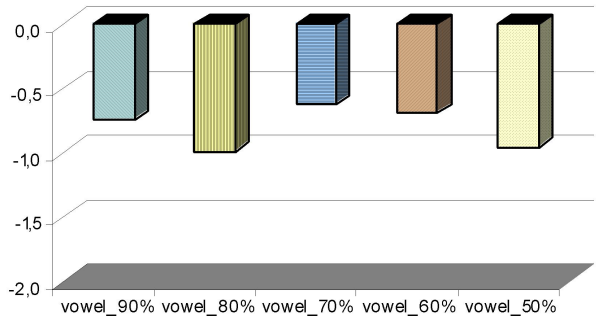

Figura 33: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 *versus* conjunto original) vowel

Na Figura 34 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 18,48% (vowel), para 18,93% dos conjuntos com base 10. Isso significa um aumento de 2,43% da taxa de erro.

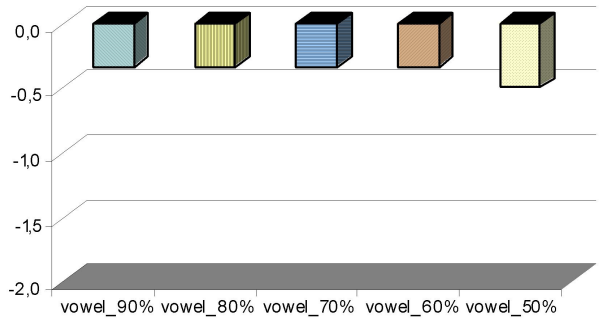

Figura 34: Diferença absoluta da taxa de erro (arredondamento utilizando base 10 *versus* conjunto original) vowel

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

Nas Figuras 35 e 36 são mostradas as diferenças absolutas em desvios padrões do tamanho da árvore no eixo vertical do gráfico, utilizando arredondamento com base 2 e base 10, respectivamente. Em média, o tamanho da árvore aumentou de 213,40 (vowel) para 218,42 (vowel-base2) e para 223,52 (vowel-base10). Isso significa um aumento de 2,35% para (vowel-base2) e de 4,74% para (vowel-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para todos os conjuntos de (vowel-base2), e de (vowel-base10), embora não de forma significativa.

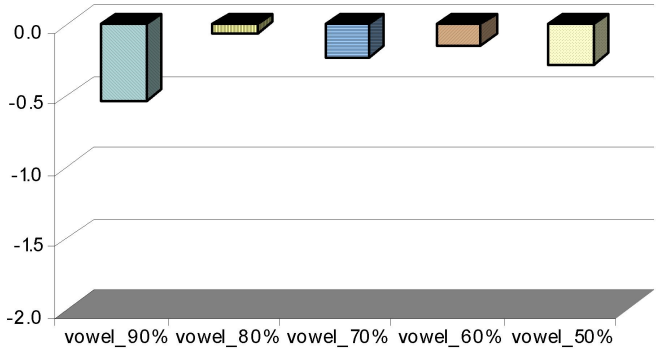

Figura 35: Diferença absoluta do tamanho do classificador (arredondamento utilizando base 2 *versus* conjunto original) vowel

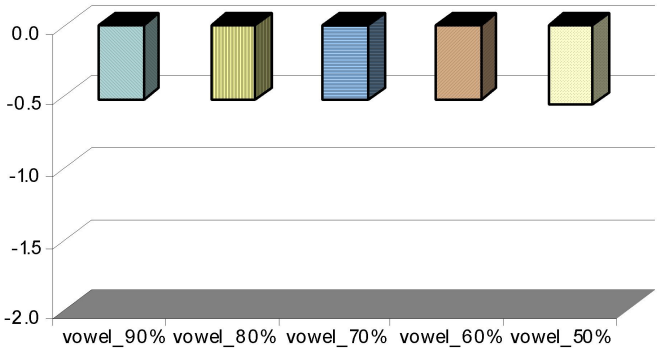

Figura 36: Diferença absoluta do tamanho do classificador (arredondamento utilizando base 10 *versus* conjunto original) vowel

6.4 Resultados wine

Nas Tabelas 28 e 29 são mostrados os atributos que aparecem nas árvores induzidas, número de atributos (#A) e porcentagem do total de atributos (%A), tanto para o conjunto original como para os derivados (utilizando o arredondamento proposto no Algoritmo 1) para wine, utilizando bases 2 e 10, respectivamente. Nota-se que os mesmos atributos foram selecionados pela árvore, exceto wine-50% base 10, que selecionou um atributo extra (#11).

Conjunto	Atributos	#A	%A
wine	7, 10, 13	3	23,08%
wine-90%	7, 10, 13	3	23,08%
wine-80%	7, 10, 13	3	23,08%
wine-70%	7, 10, 13	3	23,08%
wine-60%	7, 10, 13	3	23,08%
wine-50%	7, 10, 13	3	23,08%

Tabela 28: Atributos que aparecem na árvore induzida wine - arredondamento utilizando o Algoritmo 1 com base 2

Conjunto	Atributos	#A	%A
wine	7, 10, 13	3	23,08%
wine-90%	7, 10, 13	3	23,08%
wine-80%	7, 10, 13	3	23,08%
wine-70%	7, 10, 13	3	23,08%
wine-60%	7, 10, 13	3	23,08%
wine-50%	7, 10, 11, 13	4	30,77%

Tabela 29: Atributos que aparecem na árvore induzida wine - arredondamento utilizando o Algoritmo 1 com base 10

Analogamente à Tabela 21, na Tabela 30 são mostrados os resultados (média ± 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	$0,03 \pm 0,02$	$0,03 \pm 0,02$	$6,18 \pm 1,75$	$6,18 \pm 1,75$	$9,80 \pm 1,40$	$9,80 \pm 1,40$
wine-90%	$0,02 \pm 0,00$	$0,02 \pm 0,01$	$4,51 \pm 1,64$	$5,00 \pm 1,75$	$10,40 \pm 3,13$	$11,40 \pm 3,37$
wine-80%	$0,02 \pm 0,00$	$0,01 \pm 0,01$	$5,03 \pm 1,75$	$5,00 \pm 1,75$	$10,60 \pm 3,10$	$11,40 \pm 3,37$
wine-70%	$0,02 \pm 0,00$	$0,02 \pm 0,00$	$6,14 \pm 1,75$	$6,70 \pm 1,81$	$10,80 \pm 3,05$	$10,60 \pm 2,80$
wine-60%	$0,02 \pm 0,00$	$0,01 \pm 0,01$	$5,58 \pm 1,43$	$9,02 \pm 1,91$	$10,20 \pm 1,40$	$11,20 \pm 1,75$
wine-50%	$0,02 \pm 0,00$	$0,02 \pm 0,00$	$6,18 \pm 1,77$	$10,68 \pm 2,11$	$10,00 \pm 1,41$	$11,00 \pm 1,63$

Tabela 30: Tempo de indução, taxa de erro e tamanho do classificador utilizando arredondamento com bases 2 e 10 wine

Na Figura 37 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.

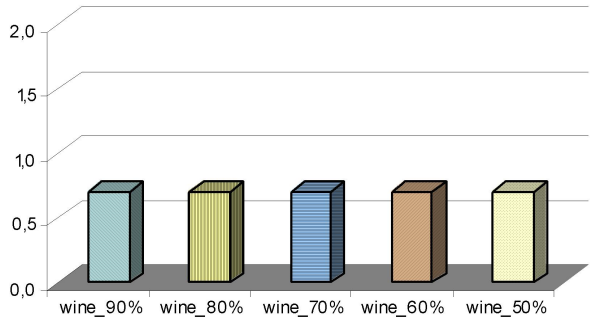

Figura 37: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 *versus* conjunto original) wine

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 wine e wine-90%, entre wine e wine-80% e assim por diante, utilizando base 10.

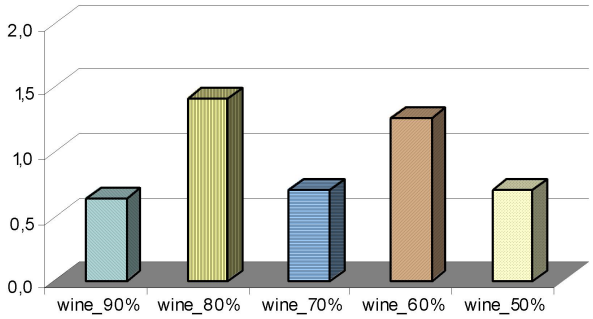

Figura 38: Diferença absoluta do tempo de indução (arredondamento utilizando base 10 *versus* conjunto original) wine

Como esperado, 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.

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 diminuiu de 6,18% (wine) para 5,49% utilizando o arredondamento com a base 2. Isso significa uma redução de 11,16% da taxa de erro.

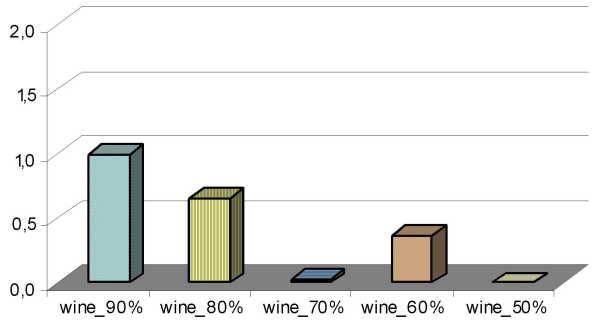

Figura 39: Diferença absoluta da taxa de erro (arredondamento utilizando base 2 *versus* conjunto original) wine

Na Figura 40 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 6,18% (wine), para 7,28% dos conjuntos com base 10. Isso significa um aumento de 17,80% da taxa de erro.

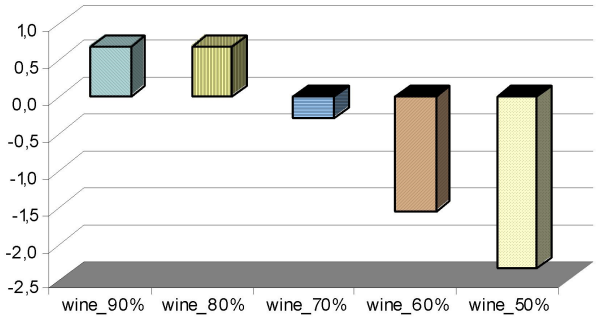

Figura 40: Diferença absoluta da taxa de erro (arredondamento utilizando base 10 *versus* conjunto original) wine

Como pode ser observado, a taxa de erro aumentou para os conjuntos (wine-70%), (wine-60%) e (wine-50%) utilizando arredondamento com a base 10, sendo de forma significativa apenas para o conjunto (wine-50%). Para todos os outros conjuntos a taxa de erro reduziu, embora de forma não significativa.

Nas Figuras 41 e 42 são mostradas as diferenças absolutas em desvios padrões do tamanho da árvore no eixo vertical do gráfico, utilizando arredondamento com base 2 e base 10, respectivamente. Em média, o tamanho da árvore aumentou de 9,80 (wine) para 10,40 (wine-base2) e para 11,12 (wine-base10). Isso significa um aumento de 6,12% para (wine-base2) e de 13,47% para (wine-base10) do tamanho da árvore. Como pode ser observado o tamanho da árvore aumentou para todos os conjuntos de ambas as bases, embora não de forma significativa.

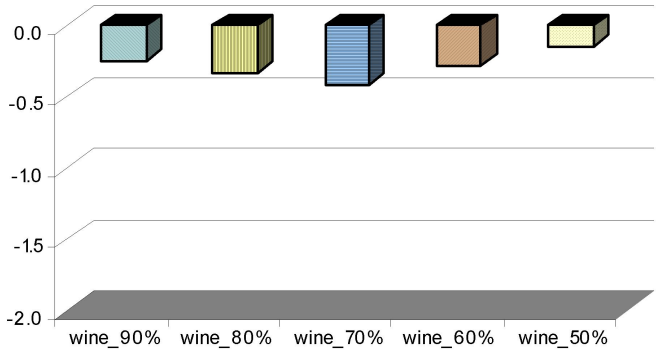

Figura 41: Diferença absoluta do tamanho do classificador (arredondamento utilizando base 2 *versus* conjunto original) wine

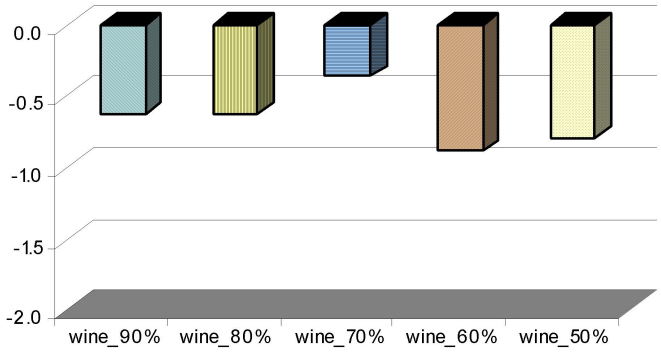

Figura 42: Diferença absoluta do tamanho do classificador (arredondamento utilizando base 10 *versus* conjunto original) wine

6.5 Resultados aml-all

É conveniente lembrar que este conjunto de exemplos consiste de 38 exemplos de treinamento e 34 exemplos de teste. Para avaliar o tempo de indução esses dois conjuntos foram unificados e foi utilizado *10-fold stratified cross-validation* sobre os 72 exemplos. Para permitir comparações de nossos resultados com aqueles publicados na literatura, o classificador foi induzido apenas sobre o conjunto de 38 exemplos de treinamento, enquanto a taxa de erro foi avaliada utilizando-se o conjunto independente de 34 exemplos de teste.

Nas árvores induzidas a partir do conjunto de treinamento aml-all, aparece um único atributo, *Zyxin*, tanto na árvore induzida a partir do conjunto original como aquelas induzidas a partir dos conjuntos derivados, como pode ser observado na Tabela 31. O tamanho da classificador é sempre constante para o conjunto original aml-all e todos seus derivados, e é igual a três.

A taxa de erro no conjunto de treinamento é sempre igual a zero para todos os conjuntos de exemplos, exceto para aml-all-60% e aml-all-50% utilizando base 10, cuja taxa de erro é de $1/38 = 2,63\%$. A taxa de erro no conjunto independente de teste é sempre igual a $3/34 = 8,82\%$ para todos os conjuntos de exemplos.

Na Tabela 32, são mostrados os resultados (média $\pm$ desvio padrão) obtidos em relação aos conjuntos de exemplos aml-all original e derivados. A segunda e terceira colunas representam os resultados do tempo de indução, utilizando a base binária e a base decimal, respectivamente.

Classificador	Árvore
aml-all	Zyxin <= 938: ALL (27.0) Zyxin > 938: AML (11.0)
aml-all-90%, base 2 aml-all-80%, base 2	Zyxin <= 960: ALL (27.0) Zyxin > 960: AML (11.0)
aml-all-70%, base 2 aml-all-60%, base 2 aml-all-50%, base 2	Zyxin <= 1024: ALL (27.0) Zyxin > 1024: AML (11.0)
aml-all-90%, base 10 aml-all-80%, base 10 aml-all-70%, base 10	Zyxin <= 900: ALL (27.0) Zyxin > 900: AML (11.0)
aml-all-60%, base 10 aml-all-50%, base 10	Zyxin <= 1000: ALL (28.0/1.0) Zyxin > 1000: AML (10.0)

Tabela 31: Classificador para o conjunto aml-all e derivados

Conjunto	Tempo(s) (base 2)	Tempo(s) (base 10)
aml-all	$3,00 \pm 0,55$	$3,00 \pm 0,55$
aml-all-90%	$2,69 \pm 0,40$	$2,19 \pm 0,30$
aml-all-80%	$2,40 \pm 0,31$	$2,09 \pm 0,26$
aml-all-70%	$2,33 \pm 0,45$	$1,95 \pm 0,32$
aml-all-60%	$2,15 \pm 0,38$	$1,97 \pm 0,39$
aml-all-50%	$1,94 \pm 0,38$	$1,69 \pm 0,17$

Tabela 32: Tempo de indução do classificador utilizando arredondamento com bases 2 e 10 aml-all

Na Figura 43 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 e aml-all-90%, entre aml-all e aml-all-80% e assim por diante, utilizando base 2.

Nota-se que o tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo de forma não significativa para todos os conjuntos, exceto aml-all-50% (com grau de confiança de 95%).

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 aml-all e aml-all-90%, entre aml-all e aml-all-80% e assim por diante, utilizando base 10.

O tempo de indução reduziu para todos os conjuntos utilizando arredondamento, sendo de forma não significativa apenas para o conjunto aml-all-90%; todos os outros tiveram uma redução significativa.

A seguir é efetuada uma comparação entre os resultados obtidos neste trabalho com aqueles obtidos por Golub (1999) e Gamberger, Lavrac, Zelezny & Tolar (2004), que abordam estratégias de voto ponderado e indução de regras, respectivamente. Nesta comparação, apenas os resultados obtidos a partir do conjunto independente de teste são considerados.

O classificador obtido por Golub (1999) apresenta taxa de erro de $5/34 = 14,70\%$ no con-

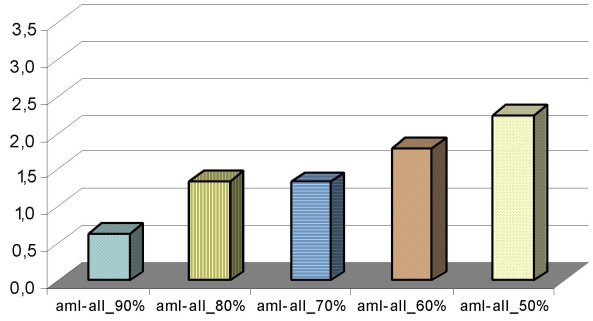

Figura 43: Diferença absoluta do tempo de indução (arredondamento utilizando base 2 *versus* conjunto original) aml-all

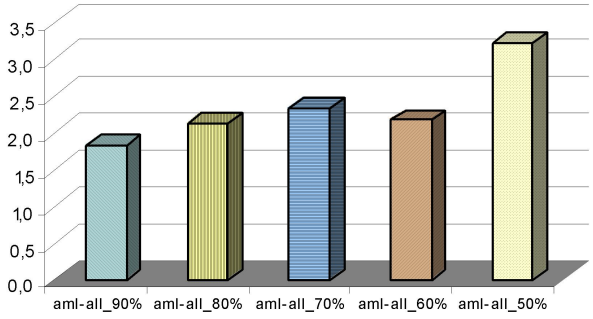

Figura 44: Diferença absoluta do tempo de indução (arredondamento utilizando base 10 *versus* conjunto original) aml-all

junto de teste.

O classificador obtido por Gamberger, Lavrac, Zelezny & Tolar (2004) que consiste de duas regras, ambas com duas condições cada, é similar em tamanho a uma árvore de decisão contendo 6 nós. Cada regra é considerada como um classificador separado pelos autores e, sendo assim, são reportadas duas taxas de erro no conjunto de teste de $7/34 = 20,59\%$ para a regra cuja conclusão é a classe AML e $2/34 = 5,88\%$ para a regra cuja conclusão é a classe ALL, ambas taxas calculadas sobre o conjunto de teste.

Comparados com os resultados relatados por (Golub 1999) e (Gamberger, Lavrac, Zelezny & Tolar 2004), a taxa de erro obtida em nossos resultados é de $3/34 = 8,82\%$ um pouco menor do que aquelas obtidas por abordagem de voto ponderado. A árvore de decisão obtida também é ligeiramente menor do que as regras induzidas podendo ser mais facilmente interpretada.

Outro ponto interessante é que a árvore de decisão identificou como importante para a separação das classes AML e ALL um atributo que também é reportado por Golub (1999) como sendo um "gene informativo", dentre outros.

É interessante notar que, embora abordagens mais sofisticadas tenham sido descritas na literatura para o conjunto de exemplos aml-all, a utilização de árvores de decisão parece ser promissora para a análise de dados de expressão gênica. Considerando a utilização de árvores de decisão com arredondamento também é possível notar que não houve alteração do atributo selecionado, assim com o valor escolhido para o teste do atributo teve uma pequena oscilação, mesmo para 50% de arredondamento de valores. É possível que esta técnica possa ser útil para melhor definir os valores de teste escolhidos pela árvore de decisão dos níveis de expressão gênica.

6.6 Discussão

Na Tabela 33 é mostrado um resumo dos experimentos realizados para o conjunto de exemplos sonar. Os resultados constituem uma comparação entre as árvores induzidas nos conjuntos arredondados quando comparadas com aquela induzida sobre o conjunto original de exemplos. Na segunda coluna encontra-se indicado se houve aumento ($\triangle$), igualdade ($=$), ou redução ($\triangledown$) do número total de atributos que apareceram na árvore. Na terceira coluna é mostrado a proporção de atributos em comum. Nas três últimas colunas é descrito se houve aumento ($\triangle$), igualdade ($=$), ou redução ($\triangledown$) no tempo de indução, taxa de erro e tamanho do classificador, respectivamente. Quando o aumento (redução) for significativo — com grau de confiança de 95% — isso é representado como ▲ (▼). A notação sonar-90%-b2 indica o conjunto de exemplos sonar-90% utilizando base 2; sonar-90%-b10 indica o conjunto de exemplos sonar-90% utilizando base 10, e assim sucessivamente. De forma análoga para as Tabelas 34, 35 e 36 para os demais conjuntos de exemplos.

Conjunto	Número de atributos	Proporção de atributos iguais	Tempo de indução	Taxa de erro	Tamanho do classificador
sonar-t3	$\triangle$	92,31%	$\triangledown$	$\triangle$	$\triangle$
sonar-t2	$\triangle$	46,15%	$\triangledown$	$\triangledown$	$\triangle$
sonar-t1	$\triangle$	30,77%	$\triangledown$	$\triangle$	$\triangle$
sonar-90%-b2	$=$	100,00%	$\triangledown$	$\triangledown$	$\triangledown$
sonar-80%-b2	$=$	100,00%	$\triangledown$	$\triangledown$	$\triangledown$
sonar-70%-b2	$\triangle$	69,23%	$\triangledown$	$\triangledown$	$\triangledown$
sonar-60%-b2	$\triangle$	69,23%	$\triangledown$	$\triangledown$	$\triangledown$
sonar-50%-b2	$\triangle$	38,46%	$\triangledown$	$\triangledown$	$\triangledown$
sonar-90%-b10	$=$	69,23%	$\triangledown$	$\triangledown$	$\triangle$
sonar-80%-b10	$\triangle$	69,23%	$\triangledown$	$\triangledown$	$\triangle$
sonar-70%-b10	$=$	61,54%	$\triangledown$	$\triangledown$	$\triangle$
sonar-60%-b10	$=$	61,54%	$\triangledown$	$\triangledown$	$\triangle$
sonar-50%-b10	$=$	61,54%	$\triangledown$	$\triangledown$	$\triangle$

Tabela 33: Resumo dos resultados sonar

Ainda considerando a Tabela 33, é possível notar que o arredondamento científico, usando redução de casas decimais, causou um aumento no número de atributos que aparecem no classificador. Considerando o Algoritmo 1, houve uma diminuição da proporção de atributos em comum exceto para sonar-90%-b2 e sonar-80%-b2. Todos os conjuntos arredondados diminuíram o tempo de indução; no geral a taxa de erro também diminui para os conjuntos arredondados, embora de forma não significativa. O tamanho do classificador aumentou para o arredondamento científico e aquele utilizando Algoritmo 1, base 10, entretanto houve diminuição do mesmo para a base 2.

Considerando a Tabela 34, é possível notar que houve uma redução do número de atributos que aparecem no classificador, exceto para ionosphere-t4, que permaneceu constante, tanto para arredondamento científico como para Algoritmo 1, no geral. No geral, houve uma diminuição da proporção de atributos em comum para cada experimento, exceto para ionosphere-90%-b10, que houve uma redução seguido por uma aumento em ionosphere-80%-b10, permanecendo constante até ionosphere-50%-b10. Tanto o tempo de indução e o tamanho do classificador diminuíram em todos os conjuntos arredondados, exceto para o arredondamento científico, sendo que para o tempo de indução houve diminuições significativas. A taxa de erro reduziu no arredondamento científico, sendo significativa para uma quantidade menor de casas decimais. Isto foi oposto do que ocorreu aplicando o Algoritmo 1, sendo significativo para base 2.

Conjunto	Número de atributos	Proporção de atributos iguais	Tempo de indução	Taxa de erro	Tamanho do classificador
ionosphere-t4	=	100,00%	▽	▽	▽
ionosphere-t3	▽	64,29%	▽	▽	△
ionosphere-t2	▽	64,29%	▼	▼	△
ionosphere-t1	▽	57,14%	▼	▼	=
ionosphere-90%-b2	▽	50,00%	▽	▲	▽
ionosphere-80%-b2	▽	50,00%	▽	▲	▽
ionosphere-70%-b2	▽	50,00%	▼	▲	▽
ionosphere-60%-b2	▽	50,00%	▼	▲	▽
ionosphere-50%-b2	▽	50,00%	▼	▲	▽
ionosphere-90%-b10	▽	35,71%	▽	△	▽
ionosphere-80%-b10	▽	42,86%	▽	△	▽
ionosphere-70%-b10	▽	42,86%	▼	△	▽
ionosphere-60%-b10	▽	42,86%	▼	△	▽
ionosphere-50%-b10	▽	42,86%	▼	△	▽

Tabela 34: Resumo dos resultados ionosphere

Na Tabela 35, é possível notar que os mesmos atributos apareceram em todos os classificadores. No geral, houve uma redução da tempo de indução, que foi significativa para vowel-t2. Ocorreu um aumento tanto da taxa de erro como do tamanho do classificador, sendo significativo apenas para taxa de erro utilizando arredondamento científico.

Conjunto	Número de atributos	Proporção de atributos iguais	Tempo de indução	Taxa de erro	Tamanho do classificador
vowel-t2	=	100,00%	▼	▲	△
vowel-t1	=	100,00%	▽	▲	△
vowel-90%-b2	=	100,00%	▽	△	△
vowel-80%-b2	=	100,00%	=	△	△
vowel-70%-b2	=	100,00%	▽	△	△
vowel-60%-b2	=	100,00%	▽	△	△
vowel-50%-b2	=	100,00%	▽	△	△
vowel-90%-b10	=	100,00%	▽	△	△
vowel-80%-b10	=	100,00%	▽	△	△
vowel-70%-b10	=	100,00%	▽	△	△
vowel-60%-b10	=	100,00%	▽	△	△
vowel-50%-b10	=	100,00%	▽	△	△

Tabela 35: Resumo dos resultados vowel

Na Tabela 36, novamente, é possível notar que os mesmos atributos apareceram em todos os classificadores, exceto para wine-50%-b10 em que houve acréscimo de um atributo no classificador. Houve uma redução da tempo de indução para todos os conjuntos; o oposto ocorreu como o tamanho do classificador. Para a base 2, no geral, ocorreu uma redução na taxa de erro; já para a base 10, acorreu um aumento significativo para sonar-50%-b10.

Conjunto	Número de atributos	Proporção de atributos iguais	Tempo de indução	Taxa de erro	Tamanho do classificador
wine-90%-b2	=	100,00%	▽	▽	△
wine-80%-b2	=	100,00%	▽	▽	△
wine-70%-b2	=	100,00%	▽	▽	△
wine-60%-b2	=	100,00%	▽	▽	△
wine-50%-b2	=	100,00%	▽	=	△
wine-90%-b10	=	100,00%	▽	▽	△
wine-80%-b10	=	100,00%	▽	▽	△
wine-70%-b10	=	100,00%	▽	△	△
wine-60%-b10	=	100,00%	▽	△	△
wine-50%-b10	△	100,00%	▽	▲	△

Tabela 36: Resumo dos resultados wine

Na Tabela 37, encontra-se uma resumo das Tabelas 33, 34, 35 e 36. Considerando o número de atributos que apareceram no classificador, podemos notar que para o arredondamento científico o número de aumentos, reduções e igualdades foram iguais, já para o arredondamento utilizando o Algoritmo 1, tanto para a base 2 quanto para a base 10, no geral, o número de atributos no classificador manteve-se constante.

Tipo de arredondamento	Número de atributos	Proporção média de atributos iguais	Tempo de indução	Taxa de erro	Tamanho do classificador
Arredondamento científico	3 = 3 △ 0 ▲ 3 ▽ 0 ▼	72,77%	0 = 0 △ 0 ▲ 6 ▽ 3 ▼	0 = 2 △ 2 ▲ 3 ▽ 2 ▼	1 = 7 △ 0 ▲ 1 ▽ 0 ▼
Arredondamento Algoritmo 1, base 2	12 = 3 △ 0 ▲ 5 ▽ 0 ▼	81,35%	1 = 0 △ 0 ▲ 16 ▽ 3 ▼	1 = 5 △ 5 ▲ 9 ▽ 0 ▼	0 = 10 △ 0 ▲ 10 ▽ 0 ▼
Arredondamento Algoritmo 1, base 10	13 = 2 △ 0 ▲ 5 ▽ 0 ▼	76,51%	0 = 0 △ 0 ▲ 17 ▽ 3 ▼	0 = 12 △ 1 ▲ 7 ▽ 0 ▼	0 = 15 △ 0 ▲ 5 ▽ 0 ▼

Tabela 37: Resumo dos resultados dos conjuntos de exemplos

A proporção média de atributos iguais foi menor para o arredondamento científico se comparado com o arredondamento utilizando o Algoritmo 1 com bases 2 e 10. Portanto, com base nesses resultados é possível concluir que o arredondamento científico preserva uma proporção menor do atributos que aparecem na árvore extraída a partir do conjunto original de exemplos. Das três abordagens, a que melhor preserva os atributos original é o Algoritmo 1 utilizando base 2.

No geral, todas as abordagens de arredondamento apresentaram uma redução no o tempo de indução, como era esperado.

Para a taxa de erro ocorre, no geral, aumentos e reduções não significativos em maior número do que os significativos para as três abordagens. Para o arredondamento utilizando Algoritmo 1, base 10, há uma tendência maior do aumento da taxa de erro, embora não significativa.

O número de vezes em que o tamanho do classificador aumenta ou diminui é o mesmo para a abordagem do Algoritmo 1, base 2. As abordagens de arredondamento científico e Algoritmo 1, base 10 apresentaram uma tendência de aumento no tamanho do classificador, embora não significativa.

7 Considerações Finais

Durante a pesquisa bibliográfica para este trabalho, foi possível encontrar alternativas de arredondamento, por exemplo em P. S. Miner & 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 um redução no tempo de indução de árvores de decisão quando o conjunto de exemplos tem seus valores arredondados. No geral, o Algoritmo 1, base 10, apresenta menor tempo de indução que a base 2. A proporção de atributos em comum que aparecem na árvore induzida a partir do conjunto original é maior para arredondamento utilizando o Algoritmo 1, base 2. Há uma tendência do tamanho do classificador ser menor no Algoritmo 1, base 2 do que na base 10. Assim, essa abordagem deve ser considerada inicialmente caso exista a necessidade de uma maior preservação dos atributos e/ou de obter um menor classificador. Dentre as três abordagens estudadas neste trabalho, o Algoritmo 1, base 10 foi o que apresentou uma maior tendência de aumento tanto da taxa de erro como do tamanho do classificador induzido.

Trabalhos futuros podem estender o aqui desenvolvido, aumentando o número de conjuntos de exemplos assim como utilizar indutores de diferentes paradigmas de aprendizado, por exemplo, indução de regras e redes neurais artificiais.

Referências

Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984). *Classification and Regression Trees*. Pacific Grove, CA: Wadsworth & Books.

Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. (2002). *Algoritmos: Teoria e Prática*. Campus. 2ª edição.

Deitel, H. M. & Deitel, P. J. (Eds.) (2005). *Java: Como Programar*. Prentice-Hall.

Dietterich, T. G. (1986). Learning at the knowledge level. *Machine Learning 1*(3), 287–315. Reprinted in Shavlik and Dietterich (eds.), 1990. Readings in Machine Learning, Morgan Kaufmann Publishers, Inc.

Forina, M. (1991). An extendible package for data exploration, classification and correlation.

Gamberger, D., Lavrac, N., Zelezny, F. & Tolar, J. (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. & Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks 1*, 75–89.

Kubat, M., Bratko, I. & Michalski, R. S. (1998). *A Review of Machine Learning Methods*, pp. 3–69. John Wiley & Sons Ltd., West Sussex, England.

Michalski, R. S. (1983a). A theory and methodology of inductive learning. *Artificial Intelligence 20*, 111–161.

Michalski, R. S. (1983b). *A Theory and Methodology of Inductive Learning*, pp. 83–134. Morgan Kaufmann, Los Altos, CA.

Moses, L. E. (Ed.) (1986). *Think and Explain with Statistics*. Addison–Wesley.

Newman, D. J., Hettich, S., Blake, C. & Merz, C. (1998). UCI repository of machine learning databases. `http://www.ics.uci.edu/~mlearn/MLRepository.html`.

P. S. Miner & J. F. Leathrum (1996). Verification of IEEE compliant subtractive division algorithms. In M. Srivas & A. Camilleri (Eds.), *First international conference on formal methods in computer-aided design*, Volume 1166, Palo Alto, CA, USA, pp. 64–78. Springer Verlag.

Pazzani, M. J. (2000). Knowledge discovery from data? *IEEE Intelligent Systems 13*, 10–12. March/April 2000.

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. & Indurkhya, N. (1998). *Predictive Data Mining: A Practical Guide*. San Francisco, CA: Morgan Kaufmann.

Weiss, S. M. & Kulikowski, C. A. (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. & Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Volume 1. Morgan Kaufmann.