



Aprendizado de Máquina Conceitos e Definições

- Os diversos sistemas de AM possuem características particulares e comuns que possibilitam sua classificação quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizados
- Algumas dessas características, conceitos introdutórios e definições são introduzidos nesta aula, os quais são importantes para a compreensão das aulas subsequentes



Conteúdo

Introdução

- Hierarquia do Aprendizado
- Paradigmas de Aprendizado
- AM: Conceitos e Definições
- Linguagens de Descrição

Aprendizado de Máquina (AM)

- Aprendizado de Máquina é uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática
- Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem sucedida de problemas anteriores

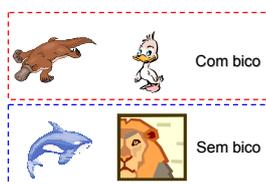
Motivação (1)

- Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



Motivação (1)

- Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



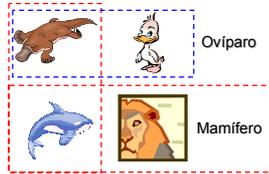
Motivação (1)

- Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



Motivação (1)

- Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



7

Motivação (2)

- Dados pares $(x, f(x))$, inferir $f(\cdot)$

x	f(x)
1	1
2	4
3	9
4	16
5	?

Dada uma amostra finita, é frequentemente impossível determinar a verdadeira função $f(\cdot)$

Abordagem: Encontre uma **hipótese (modelo)** nos exemplos de treinamento e assuma que a **hipótese** se repita para exemplos futuros também

8

Motivação (2)



Exemplo	X_1	X_2	X_3	X_4	Y
Z_1	0	1	1	0	0
Z_2	0	0	0	0	0
Z_3	0	0	1	1	1
Z_4	1	0	0	1	1
Z_5	0	1	1	0	0
Z_6	1	1	0	0	0
Z_7	0	1	0	1	0

- $f: X_1 \times X_2 \times X_3 \times X_4 \rightarrow Y$

9

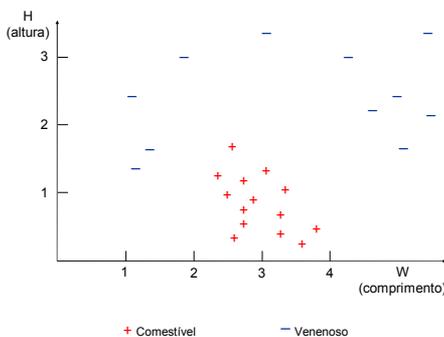
Exemplo: Cogumelos Comestíveis x Venenosos

- Um pesquisador foi a campo e coletou diversos cogumelos
- Ao chegar em seu laboratório, ele mediu o comprimento e altura de cada cogumelo
- Ele também classificou cada cogumelo coletado como **comestível** ou **venenoso**



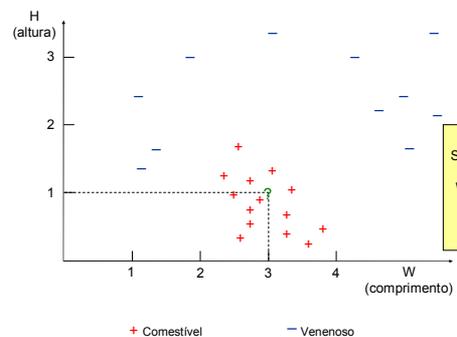
10

Exemplo: Cogumelos Comestíveis x Venenosos



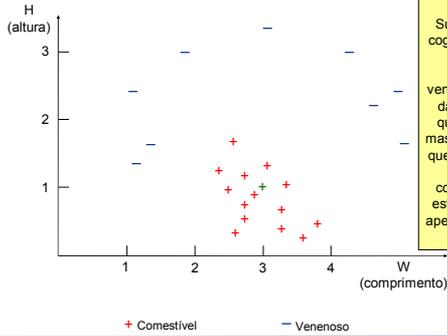
11

Exemplo: Cogumelos Comestíveis x Venenosos



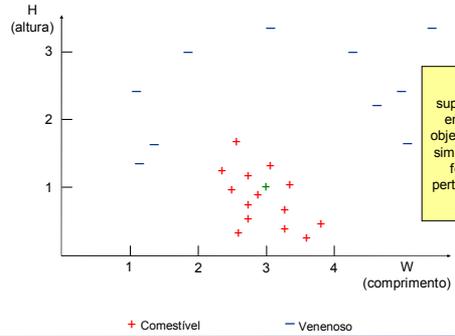
12

Exemplo: Cogumelos Comestíveis x Venenosos



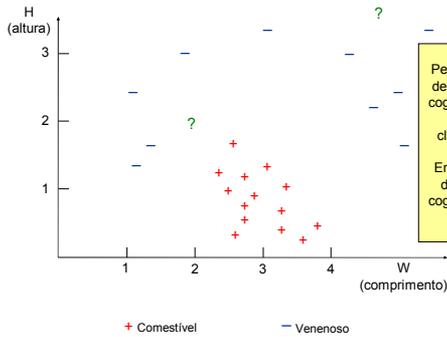
13

Exemplo: Cogumelos Comestíveis x Venenosos



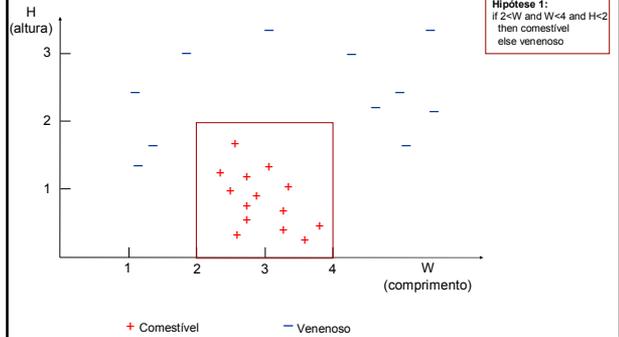
14

Exemplo: Cogumelos Comestíveis x Venenosos



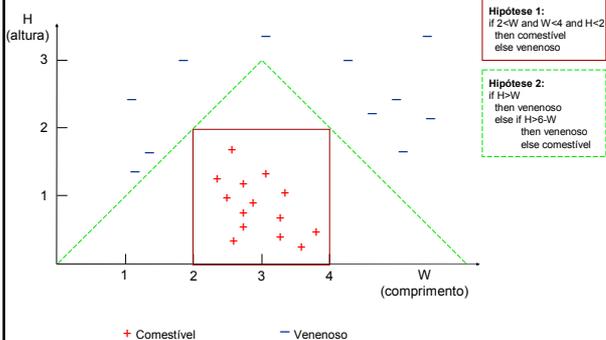
15

Exemplo: Cogumelos Comestíveis x Venenosos



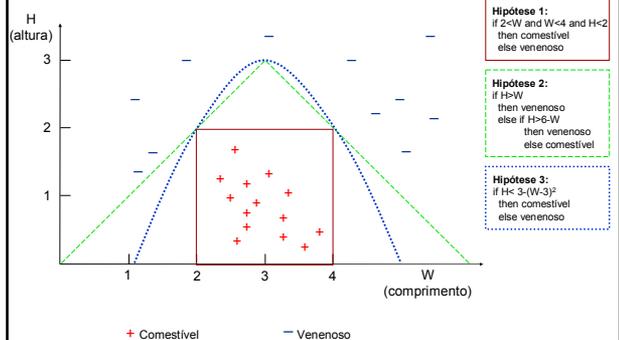
16

Exemplo: Cogumelos Comestíveis x Venenosos



17

Exemplo: Cogumelos Comestíveis x Venenosos



18

Aprendizado de Máquina

- (Simon, 1983):
"Aprender implica em alterações no sistema que são adaptativas, no sentido de elas capacitam o sistema a realizar a mesma tarefa, ou tarefas provenientes da mesma população, de forma mais eficiente e eficaz na próxima vez"
- (Weiss & Kulikowski, 1991)
"Um sistema de aprendizado [supervisionado] é um programa de computador que toma decisões baseadas na experiência contida em exemplos solucionados com sucesso"
- (Russel & Norvig 1995)
"...todo aprendizado pode ser visto como o aprendizado de uma função"

19

Conteúdo

- Introdução
- Hierarquia do Aprendizado
- Paradigmas de Aprendizado
- AM: Conceitos e Definições
- Linguagens de Descrição

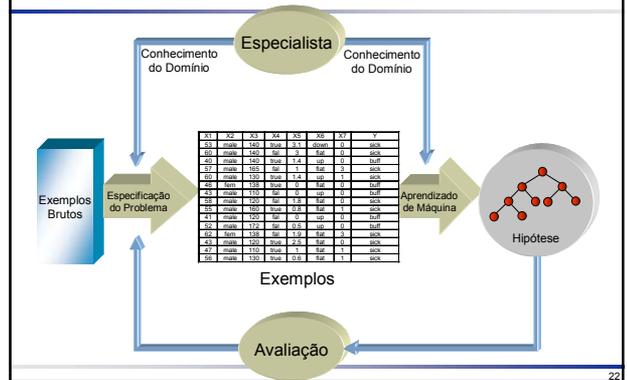
20

Hierarquia do Aprendizado



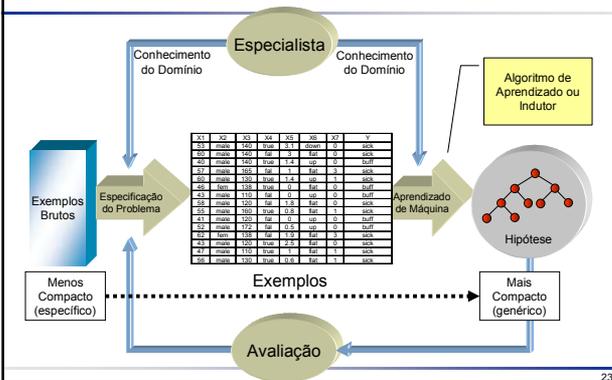
21

Aprendizado de Máquina



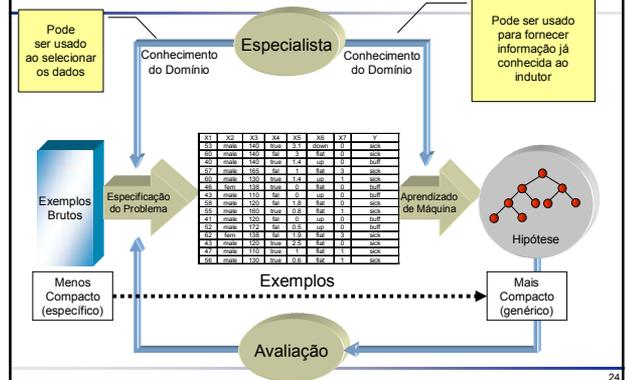
22

Aprendizado de Máquina



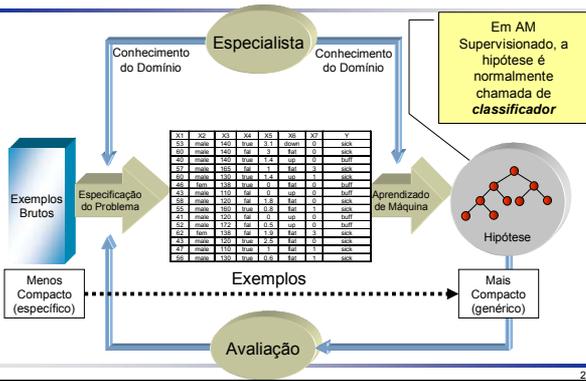
23

Aprendizado de Máquina



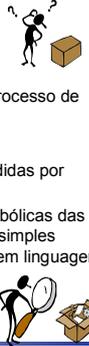
24

Aprendizado de Máquina



Categorias de Sistemas de Aprendizado

- ❑ Não Simbólico ou Caixa-preta
 - Não facilmente interpretado por humanos
 - Desenvolve sua própria representação de conceitos
 - Não fornece esclarecimento ou explicação sobre o processo de classificação
- ❑ Simbólico ou Orientado a conhecimento
 - Cria estruturas simbólicas que podem ser compreendidas por seres humanos
 - "Os resultados da indução devem ser descrições simbólicas das entidades dadas... devem ser compreensíveis como simples 'pedaços' de informação, diretamente interpretáveis em linguagem natural..." (Michalski 1983a)



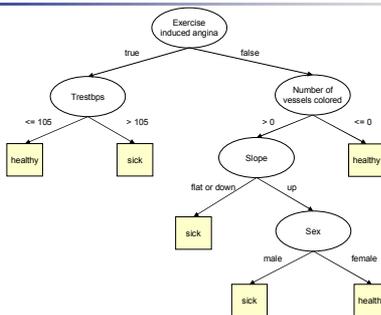
Categorias de Sistemas de Aprendizado

- ❑ A distinção entre essas duas categorias pode ser formulada em termos dos critérios:
 - **critério fraco:** o sistema utiliza exemplos para gerar subsídios para melhorar o desempenho com exemplos posteriores (ex: redes neurais, métodos estatísticos)
 - **critério forte:** o critério fraco é satisfeito e além disso o sistema é capaz de comunicar sua representação interna na forma simbólica explicitamente
 - **critério ultra-forte:** os critérios fraco e forte são satisfeitos; o sistema deve ser capaz de comunicar sua representação interna na forma simbólica explicitamente e esta pode ser usada por um humano sem a ajuda de um computador (apenas usando seu cérebro)

AS x AnS

- ❑ Aprendizado Supervisionado
 - Compreender o relacionamento entre os atributos e a classe
 - Predizer a classe de novos exemplos o melhor possível
- ❑ Aprendizado Não Supervisionado
 - Encontrar representações úteis dos exemplos, tais como:
 - ❖ Encontrar agrupamentos (clusters)
 - ❖ Redução da dimensão
 - ❖ Encontrar as causas ou as fontes ocultas dos exemplos
 - ❖ Modelar a densidade dos exemplos

Exemplo



Parte da árvore de decisão induzida por C4.5 para o conjunto de exemplos Cleveland heart disease

Conteúdo

- ❑ Introdução
- ❑ Hierarquia do Aprendizado
- ❑ Paradigmas de Aprendizado
- ❑ AM: Conceitos e Definições
- ❑ Linguagens de Descrição

Paradigmas de AM

- ❑ Simbólico
- ❑ Protótipo ou Memorização (Instance-Based)
- ❑ Conexionista
- ❑ Genético
- ❑ Estatístico

31

Paradigma Simbólico

- ❑ Os sistemas de aprendizado simbólico buscam aprender construindo representações simbólicas de um conceito através da análise de exemplos e contra-exemplos desse conceito
- ❑ As representações simbólicas estão tipicamente na forma de alguma expressão lógica tais como árvores de decisão, regras ou redes semânticas

32

Paradigma *Instance-Based*

- ❑ Uma forma de classificar um exemplo é lembrar de outro similar cuja classe é conhecida e assumir que o novo exemplo terá a mesma classe
- ❑ Essa filosofia exemplifica os sistemas baseados em exemplos, que classificam exemplos nunca vistos através de exemplos similares conhecidos
- ❑ Esse tipo de sistema de aprendizado é denominado **preguiçoso** (*lazy*)
- ❑ Sistemas *lazy* necessitam manter os exemplos na memória para classificar novos exemplos, em oposição aos sistemas **gulosos** (*eager*), que utilizam os exemplos para induzir o modelo, descartando-os logo após
- ❑ Assim, saber quais exemplos de treinamento devem ser memorizados por um indutor *lazy* é muito importante
- ❑ O ideal é reter apenas aqueles mais representativos do problema
- ❑ Os algoritmos mais conhecidos neste paradigma são os de Vizinhos mais Próximos (*Nearest Neighbours*) e Raciocínio Baseado em Casos (*Case Based Reasoning*)

33

Paradigma Conexionista

- ❑ Redes Neurais são construções matemáticas simplificadas inspiradas no modelo biológico do sistema nervoso
- ❑ A representação de uma Rede Neural envolve unidades altamente interconectadas e, por esse motivo, o nome conexionismo é utilizado para descrever a área de estudo
- ❑ A metáfora biológica com as conexões neurais do sistema nervoso tem interessado muitos pesquisadores e tem fornecido muitas discussões sobre os méritos e as limitações dessa abordagem de aprendizado
- ❑ Em particular, as analogias com a biologia têm levado muitos pesquisadores a acreditar que as Redes Neurais possuem um grande potencial na resolução de problemas que requerem intenso processamento sensorial humano, tais como visão e reconhecimento de voz

34

Paradigma Genético

- ❑ Este paradigma de aprendizado é derivado do modelo evolucionário de aprendizado
- ❑ Um classificador genético consiste de uma população de elementos de classificação que competem para fazer a predição
- ❑ Elementos que possuem um desempenho ruim são descartados, enquanto os elementos mais fortes proliferam, produzindo variações de si mesmos
- ❑ Este paradigma possui uma analogia direta com a teoria de Darwin, na qual sobrevivem os mais bem adaptados ao ambiente
- ❑ Alguns operadores genéticos básicos que aplicados a população geram novos indivíduos são: Reprodução, Cruzamento, Mutação e Inversão
- ❑ Esses operadores atuam no controle da quantidade de cópias produzidas de um indivíduo, na troca de material genético, na preservação de uma espécie e na manutenção de uma certa diversidade na nova população

35

Paradigma Estatístico

- ❑ Pesquisadores em estatística têm criado diversos métodos de classificação, muitos deles semelhantes aos métodos posteriormente desenvolvidos pela comunidade de Aprendizado de Máquina
- ❑ A idéia geral consiste em utilizar modelos estatísticos para encontrar uma boa aproximação do conceito induzido
- ❑ Vários desses métodos são paramétricos, assumindo alguma forma de modelo, e então encontrando valores apropriados para os parâmetros do modelo a partir dos exemplos
- ❑ Por exemplo, um classificador linear assume que as classes podem ser expressas como combinação linear dos valores dos atributos, e então procura uma combinação linear particular que fornece a melhor aproximação sobre o conjunto de exemplos
- ❑ Dentre os métodos estatísticos, destacam-se os de aprendizado Bayesiano, que utilizam um modelo probabilístico baseado no conhecimento prévio do problema, o qual é combinado com os exemplos de treinamento para determinar a probabilidade final de uma hipótese

36

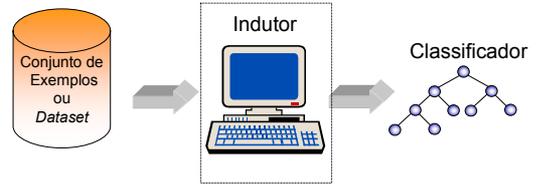
Conteúdo

- ❑ Introdução
- ❑ Hierarquia do Aprendizado
- ❑ Paradigmas de Aprendizado
- ❑ AM: Conceitos e Definições
- ❑ Linguagens de Descrição

37

Indutor

- ❑ Programa que gera uma hipótese (classificador) a partir de um conjunto de exemplos



38

Indutor

- ❑ Informalmente, o objetivo de um **indutor** (ou algoritmo de aprendizado ou algoritmo de indução) consiste em extrair um *bom* classificador a partir de um conjunto de exemplos rotulados
- ❑ A saída do indutor, o classificador, pode então ser usada para classificar exemplos novos (ainda não rotulados) com a meta de prever corretamente o rótulo de cada um
- ❑ Após isso, o classificador pode ser avaliado considerando sua precisão, compreensibilidade ou grau de interesse, velocidade de aprendizado, requisitos de armazenamento, grau de compactação ou qualquer outra propriedade desejável que determine quão bom e apropriado ele é para a tarefa em questão

39

Argumentos Dedutivos x Indutivos

Argumentos Dedutivos	Argumentos Indutivos
Se todas as premissas são verdadeiras então a conclusão é verdadeira	Se todas as premissas são verdadeiras, a conclusão é <i>provavelmente</i> verdadeira, mas não <i>necessariamente</i> verdadeira (à exceção dos argumentos matemáticos indutivos)
Toda a informação do conteúdo factual da conclusão já está, pelo menos implicitamente, nas premissas	A conclusão contém informação que não está implicitamente ou explicitamente nas premissas

Argumentos indutivos preservam falsidade

40

Argumentos Dedutivos x Indutivos

- ❑ Um exemplo de dedução:
 - Alguém diz a você “Todas as maçãs são vermelhas”. A seguir, uma pessoa lhe dá uma maçã. Você infere que ela é vermelha
- ❑ Um exemplo de indução:
 - Você vê 5 maçãs vermelhas. Você conclui “Todas as maçãs são vermelhas”
 - Suponha que você olhe com mais cuidado as maçãs e perceba que uma delas é de cor laranja. Isto falsifica sua conclusão (hipótese)

41

Exemplo, Atributo & Classe

- ❑ Exemplo
 - Exemplo, caso ou registro (*instance*)
 - É um conjunto fixo de atributos
 - Um exemplo descreve o objeto de interesse, tal como um paciente, exemplos médicos sobre uma determinada doença ou histórico de clientes de uma dada companhia
- ❑ Atributo
 - Atributo ou campo (*feature*)
 - Uma única característica de um exemplo
- ❑ Classe
 - Atributo especial que descreve o fenômeno de interesse (somente no Aprendizado Supervisionado)

42

Atributo

□ O domínio (conjunto de valores que um atributo pode assumir) do atributo X_i é indicado por $\text{dom}(X_i)$

□ Exemplo

- Atributo sexo; $\text{dom}(\text{sexo}) = \{m, f\}$
- Atributo cor; $\text{dom}(\text{cor}) = \{\text{verde, vermelho}, \dots\}$
- Atributo temperatura; $\text{dom}(\text{temperatura}) = \{\text{baixa, média, alta}\}$
- Atributo peso; $\text{dom}(\text{peso}) = \{\forall w : w \in \mathbb{R}^+\}$

43

Tipos de Atributos

□ **Nominal** (ou **discreto** ou **categorico**), quando o atributo assume valores em um conjunto finito, sendo que alguns indutores podem também aceitar uma subdivisão entre os atributos nominais:

- **Ordenado**: o domínio é ordenado, mas a diferença absoluta dos valores é desconhecida (e.g. escala de temperatura: baixa, média, alta ou severidade de um machucado)
- **Não-ordenado**: não existe uma ordem entre os valores (e.g., cor: vermelho, verde, azul; ocupação; estado civil, raça)

□ **Contínuo** (ou **numérico** ou **real**), quando o domínio é ordenado e pode ser representado por um valor real (e.g., peso $\in \mathbb{R}$, um número real)

44

Atributo

□ Para qualquer tipo de atributo, usualmente existe também um símbolo importante que significa **desconhecido**, ou seja, a ausência de um valor para aquele atributo

□ Este símbolo especial é bem diferente, por exemplo, do valor zero (às vezes usado para números) ou de cadeias de caracteres vazias

□ Na maioria dos indutores disponíveis, este valor é representado por um ponto de interrogação ?

45

Atributo

□ Um outro símbolo especial, mesmo não sendo reconhecido por vários indutores, é o **não-se-aplica**

□ Por exemplo, para o atributo *número de gestações*, pode ser utilizado o símbolo **não-se-aplica** caso o paciente seja do sexo masculino

□ Em geral, este símbolo é representado por um ponto de exclamação !

46

Atributo

□ Além disso, vários indutores assumem que os atributos originais que descrevem os exemplos são **relevantes** o suficiente para aprender a tarefa em questão

□ Entretanto, alguns atributos podem não ser diretamente relevantes e outros até **irrelevantes**

□ Um atributo é irrelevante se existe uma descrição **completa** e **consistente** das classes a serem aprendidas que não usa aquele atributo

47

Escolha de Atributos

□ Um ponto importante a ser considerado é a escolha de atributos com boa capacidade preditiva

□ Não importa qual método seja empregado, os conceitos que podem ser aprendidos estão à mercê dos exemplos e da qualidade dos atributos

□ Por exemplo, para a tarefa de determinar se uma pessoa está ou não com gripe, pode-se escolher atributos com baixo poder preditivo, tais como

- (cor-do-cabelo, cor-do-olho, modelo-do-carro, número-de-filhos)

□ ou atributos com alto poder preditivo, tais como

- (temperatura, resistência-da-pele, exame-do-pulmão)

□ Para esta tarefa específica, no segundo caso, melhores previsões em exemplos não-rotulados provavelmente ocorrerão do que com o primeiro conjunto de atributos

48

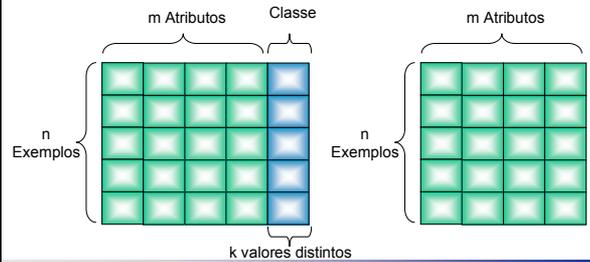
Classe

- ❑ No aprendizado supervisionado todo exemplo possui um atributo especial, o rótulo ou **classe**, que descreve o fenômeno de interesse, isto é, a meta que se deseja aprender e poder fazer previsões a respeito
- ❑ Um exemplo não-rotulado consiste do exemplo, exceto o rótulo, ou seja, um vetor de valores dos atributos
- ❑ Os rótulos são tipicamente pertencentes a um conjunto discreto (nominal) de classes $\{C_1, C_2, \dots, C_k\}$ no caso de *classificação* ou de valores reais no caso de *regressão*

49

Conjunto de Exemplos (*Dataset*)

- ❑ No Aprendizado Supervisionado, cada exemplo é rotulado segundo sua classe
- ❑ No Aprendizado Não Supervisionado, cada exemplo não possui classe associada



50

Exemplo de um Conjunto de Exemplos para Classificação

- ❑ Dez exemplos ($n=10$)
- ❑ Duas classes ($k=2$): Sim; Não
- ❑ Dois atributos ($m=2$):
 - Idade é ordenado
 - Tipo de Veículo é categórico (Esporte, Van, Caminhão)
- ❑ Rótulo da classe indica se a pessoa comprou o produto
- ❑ Atributo dependente (classe) é *categórico*

Idade	Veículo	Classe
20	V	Sim
30	V	Sim
25	C	Não
30	E	Sim
40	E	Sim
20	C	Não
30	V	Sim
25	V	Sim
40	V	Sim
20	E	Não

51

Exemplo de um Conjunto de Exemplos para Regressão

- ❑ Dez exemplos ($n=10$)
- ❑ Dois atributos ($m=2$): Idade e tipo de veículo (Esporte, Van, Caminhão)
- ❑ Despesa indica quanto a pessoa gastou durante uma visita recente à concessionária
- ❑ Atributo dependente (classe) é *numérico*

Idade	Veículo	Despesa
20	V	\$200
30	V	\$150
25	C	\$300
30	E	\$220
40	E	\$400
20	C	\$80
30	V	\$100
25	V	\$125
40	V	\$500
20	E	\$420

52

Exemplo de um Conjunto de Exemplos para Aprendizado Não Supervisionado

- ❑ Dez exemplos ($n=10$)
- ❑ Três atributos ($m=3$): Idade, tipo de veículo (Esporte, Van, Caminhão) e cor do veículo
- ❑ Neste caso, não há associação explícita de algum atributo com uma determinada classe

Idade	Veículo	Cor
20	V	preto
30	V	verde
25	C	azul
30	E	branco
40	E	azul
20	C	preto
30	V	branco
25	V	azul
40	V	Verde
20	E	azul

53

Conjuntos de Exemplos

- ❑ Em geral, um conjunto de exemplos é dividido em dois subconjuntos disjuntos:
 - **conjunto de treinamento** que é usado para o aprendizado do conceito e o
 - **conjunto de teste** que é usado para medir o grau de efetividade do conceito aprendido
- ❑ Os subconjuntos são disjuntos para assegurar que as medidas obtidas utilizando o conjunto de teste sejam de um conjunto diferente do usado para realizar o aprendizado, tornando a medida estatisticamente válida

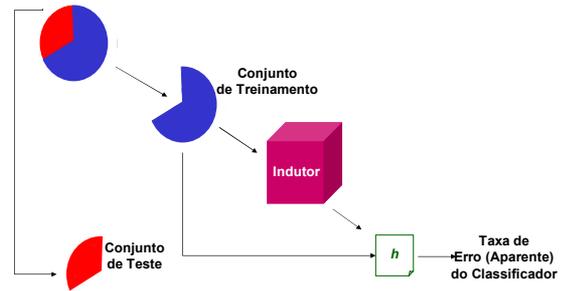
54

Conjuntos de Exemplos

- Após induzir uma hipótese, é possível avaliá-la no conjunto de treinamento bem como no conjunto de teste
- É usual denominar as medidas de desempenho de um classificador efetuadas sobre o conjunto de treinamento como **aparentes** (também conhecidas como medidas de **re-substituição**) e as medidas efetuadas sobre o conjunto de teste como medidas **reais** (ou **verdadeiras**)
- Por exemplo, caso a medida seja o **erro**, pode-se ter o **erro aparente** e o **erro verdadeiro**
- Para a maioria das hipóteses, a medida aparente é um estimador ruim do seu desempenho futuro, uma vez que ela tem a tendência de possuir um *bias* otimista
 - Em geral, o erro calculado sobre o conjunto de exemplos de treinamento (erro aparente) é menor que o erro calculado sobre o conjunto de exemplos de teste (erro verdadeiro)

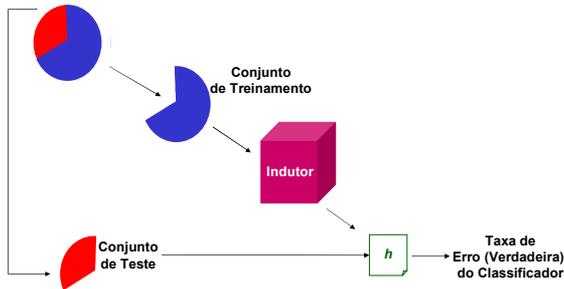
55

Erro Aparente



56

Erro Verdadeiro



57

Preparação de Dados

- Fase que antecede o processo de aprendizagem, para facilitar ou melhorar o processo.
- Exemplos:
 - remover exemplos incorretos
 - transformar o formato dos exemplos para que possam ser usados com um determinado indutor
 - selecionar um subconjunto de atributos relevantes (FSS – *Feature Subset Selection*)

58

Conhecimento do Domínio

- Background Knowledge*
- Informação sobre valores válidos de um atributo
- Crítérios para escolher atributos
- Crítérios para escolher hipóteses
- Restrições no relacionamento dos atributos
- Regras para geração de conceitos de nível mais alto
- Construção de novos atributos derivados dos atributos originais

59

Ruído

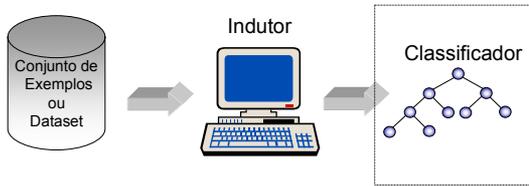
- Exemplos imperfeitos que podem ser derivados do processo de aquisição, transformação ou rotulação das classes
- Ex: exemplos com os mesmos atributos mas com classes diferentes

	X ₁	X ₂	X ₃	X ₄	Y
overcast	19	65	yes		dont_go
rain	19	70	yes		dont_go
rain	23	80	yes		dont_go
sunny	23	95	no		dont_go
sunny	28	91	yes		dont_go
sunny	30	85	no		dont_go
overcast	19	65	yes		go
rain	21	80	no		go
rain	22	95	no		go
sunny	22	70	no		go
overcast	23	90	yes		go
rain	25	81	no		go
sunny	25	72	yes		go
overcast	26	75	no		go
overcast	29	78	no		go

60

Classificador

- Dado um conjunto de exemplos, o classificador é a saída do indutor



61

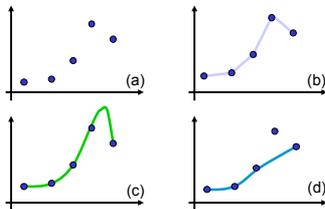
Classificador

- Dado um conjunto de treinamento, um indutor gera como saída um **classificador** (**hipótese** ou **descrição de conceito**) de forma que, dado um novo exemplo, ele possa prever precisamente sua classe
- Cada exemplo é um par $(x, f(x))$, onde
 - x é a entrada
 - $f(x)$ é a saída (f desconhecida!)
 - $y=f(x)$ assume valores discretos $y \in \{C_1, C_2, \dots, C_k\}$: classificação
 - $y=f(x)$ assume valores reais: regressão
- Indução ou inferência indutiva: dada uma coleção de exemplos de $f(\cdot)$, retornar uma função $h(\cdot)$ que aproxima $f(\cdot)$, ou seja, $h(x) \cong f(x)$
- $h(\cdot)$ é denominada uma **hipótese** sobre a função objetivo $f(\cdot)$

62

Exemplos de Hipóteses

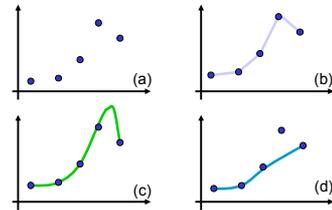
- (a) exemplos originais
- (b), (c), (d) possíveis hipóteses



63

Exemplos de Hipóteses

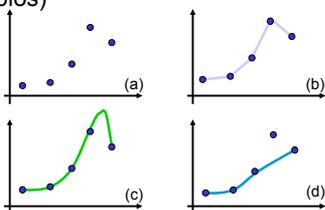
- Qual a melhor hipótese para os exemplos em (a)?



64

Bias

- Qualquer critério de preferência de uma hipótese sobre outra (além da consistência com os exemplos)



65

Variância

- Mede quanto as suposições do algoritmo da aprendizagem variam para diferentes conjuntos de treinamento
- Maiores detalhes serão vistos nas próximas aulas

66

Estabilidade

- Um indutor é instável se uma pequena perturbação (variação) no conjunto de treinamento pode causar modificação no classificador gerado
- Por exemplo:
 - FSS em regressão
 - árvores de decisão
 - redes neurais
- Indutores instáveis possuem alta variância

67

Estabilidade

- Um indutor é estável se o classificador gerado não muda muito caso os exemplos de treinamento se alterem
- Há sempre um *trade-off* entre *bias* e variância:
 - indutores instáveis, em geral, geram classificadores com alta variância mas com pequeno *bias*
 - indutores estáveis, em geral, geram classificadores com baixa variância mas com alto *bias*

68

Modo de Aprendizado

- Sempre que todo o conjunto de treinamento deva estar presente para o aprendizado, o modo de aprendizado de um algoritmo é **não-incremental**, também conhecido como modo *batch*
- Por outro lado, se o indutor não necessita construir a hipótese a partir do início, quando novos exemplos são adicionados ao conjunto de treinamento, o modo de aprendizado é **incremental**
- Portanto, no modo incremental o indutor apenas tenta atualizar a hipótese antiga sempre que novos exemplos são adicionados ao conjunto de treinamento
- Em geral, o aprendizado não-incremental deve fornecer resultados melhores, uma vez que é permitido, ao indutor, o acesso a todos os exemplos de treinamento de uma única vez, possibilitando que ele otimize suas decisões
- Entretanto, se o tempo computacional é um fator importante e novos exemplos são frequentemente adicionados ao conjunto de treinamento, o aprendizado incremental pode ser considerado para poupar tempo

69

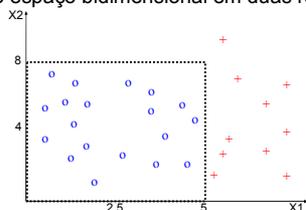
Espaço de Descrição

- m atributos podem ser vistos como um vetor
- Assim, cada atributo corresponde a uma coordenada em um espaço m -dimensional denominado **espaço de descrição**
- No Aprendizado Supervisionado, cada ponto no espaço de descrição pode ser rotulado com a classe associada

70

Espaço de Descrição

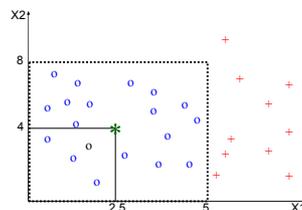
- Um indutor divide o espaço de descrição em regiões
- Cada região é rotulada com uma classe
- Exemplo: $m=2$ atributos (positivos) e seja o classificador:
`if $X_1 < 5$ and $X_2 < 8$ then classe=0 else classe=+`
divide o espaço bidimensional em duas regiões



71

Espaço de Descrição

- Para classificar um novo exemplo com $(X_1, X_2) = (2.5, 4)$, basta verificar em qual região ela se localiza e atribuir a classe associada àquela região (neste caso, classe 0)



72

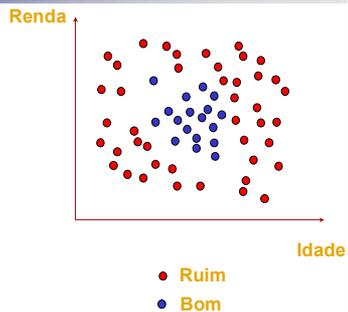
Espaço de Descrição: Exemplo

Assuma o seguinte conjunto de exemplos sobre exemplos de crédito bancário

Idade	Renda	Classe
20	2000	Ruim
30	5100	Bom
60	5000	Ruim
40	6000	Bom
...

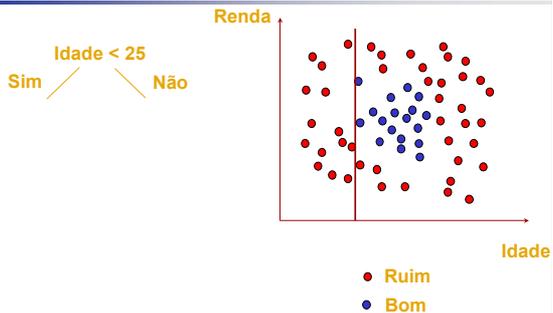
73

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



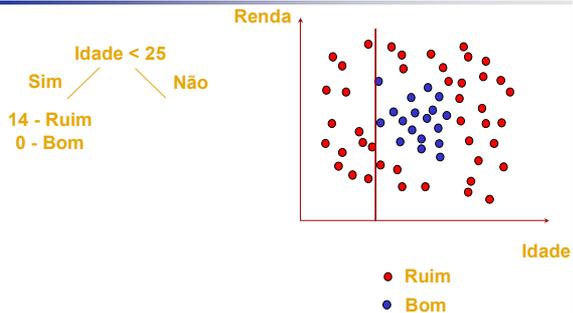
74

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



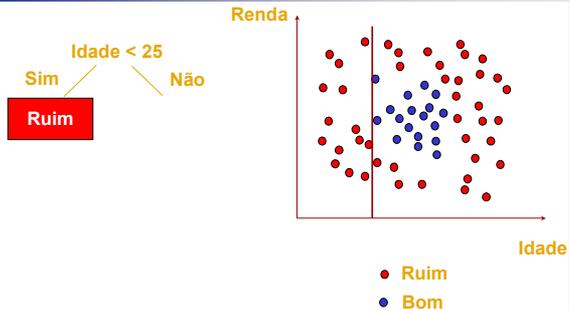
75

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



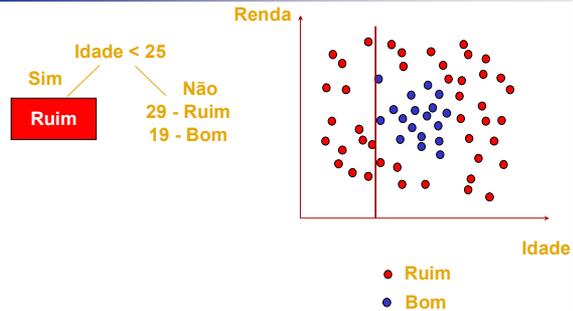
76

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



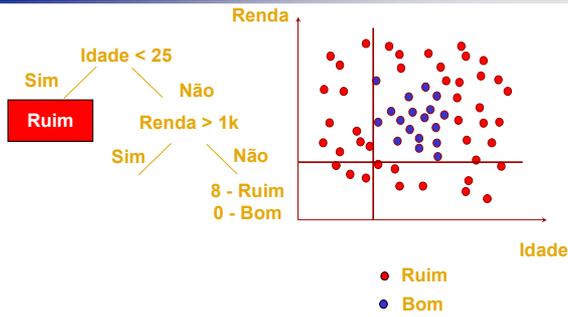
77

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



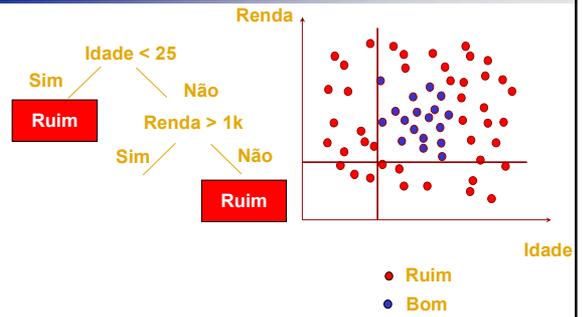
78

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



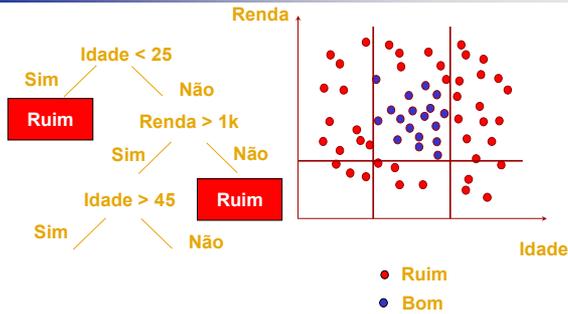
79

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



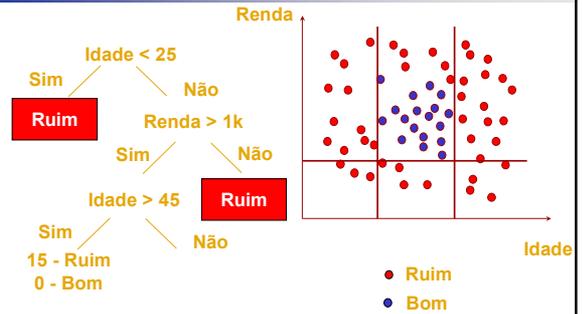
80

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



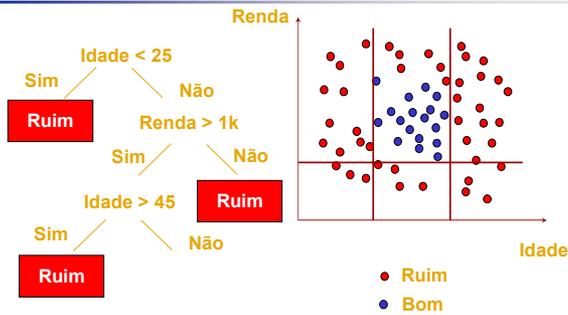
81

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



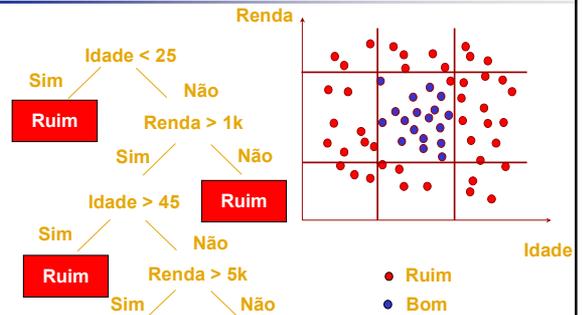
82

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



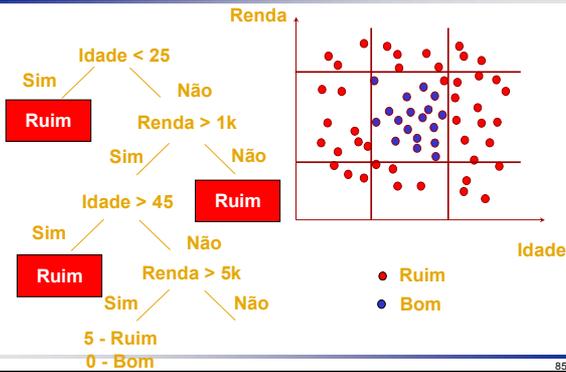
83

Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão

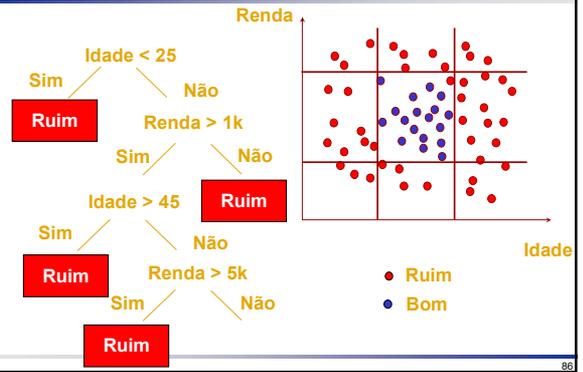


84

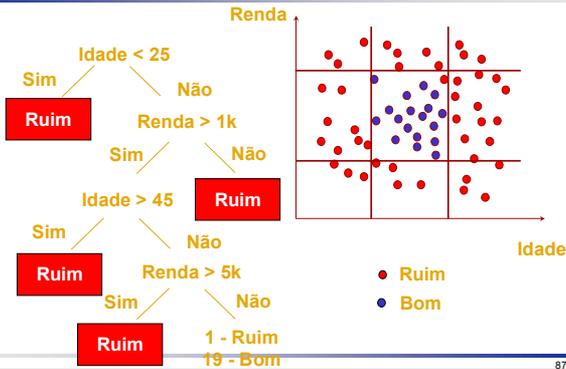
Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



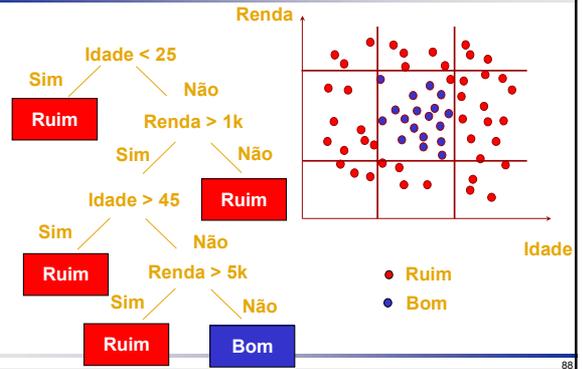
Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



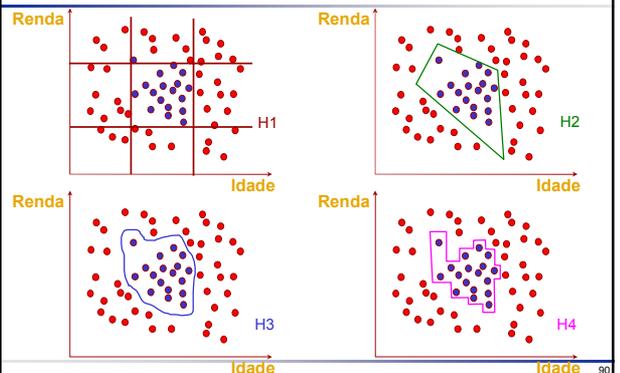
Espaço de Descrição: Exemplo Induzindo uma Árvore de Decisão



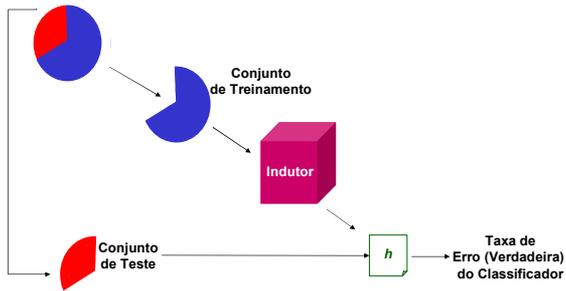
Erro e Precisão

- Principais fatores de erro:
 - Qualidade (representatividade) da informação dos atributos
 - Adaptação do algoritmo de aprendizado aos exemplos
 - Distribuição dos exemplos futuros
 - Quantidade de exemplos

Erro x Possíveis Hipóteses

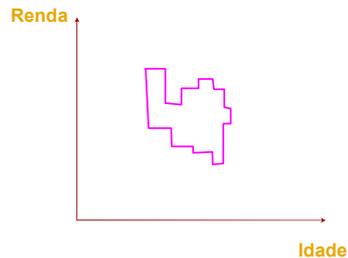


Qual a Melhor Hipótese? Não Esquecer o Erro...



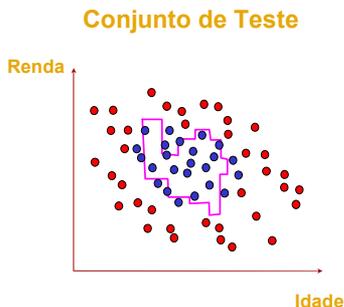
91

H4...



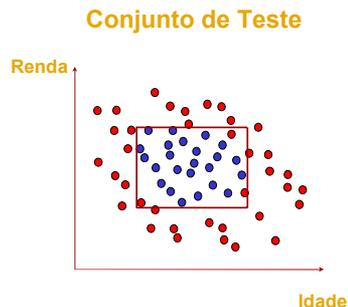
92

Erro de H4



93

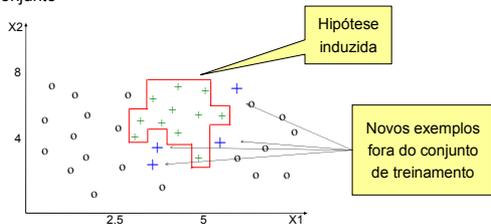
Erro de H1



94

Overfitting

- A hipótese extraída a partir dos exemplos é muito específica para o conjunto de treinamento
 - A hipótese apresenta um bom desempenho para o conjunto de treinamento, mas um desempenho ruim para os casos fora desse conjunto



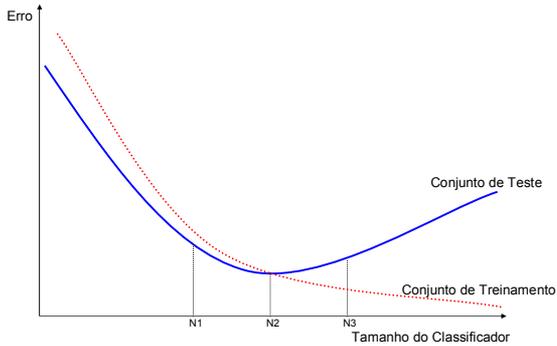
95

Underfitting

- A hipótese induzida apresenta um desempenho ruim tanto no conjunto de treinamento como de teste
 - poucos exemplos representativos foram dados ao sistema de aprendizado (e.g. algoritmos de árvores de decisão ou de indução de regras)
 - o usuário pré-definiu um tamanho muito pequeno para o classificador (e.g. insuficientes neurônios em uma rede neural ou um alto valor de poda para árvores de decisão)

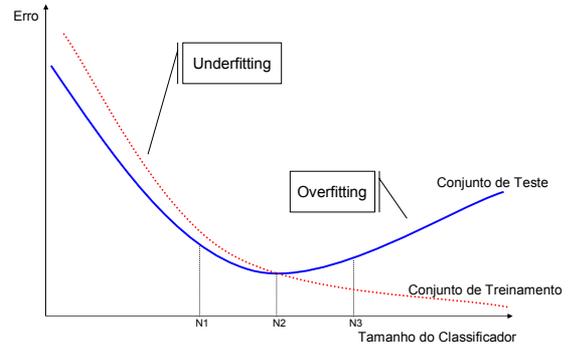
96

Relação entre o Tamanho do Classificador e o Erro



97

Relação entre o Tamanho do Classificador e o Erro



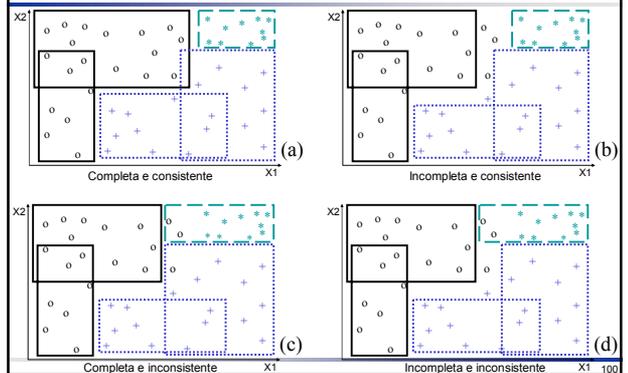
98

Consistência e Completude

- Depois de induzida, uma hipótese pode ser avaliada sobre
 - consistência, se classifica corretamente os exemplos
 - completude, se classifica todos os exemplos

99

Relação entre Completude e Consistência



100

Complexo

- É uma conjunção de disjunções dos atributos de teste, na forma:

$$X_i \text{ op valor}$$

onde X_i é um atributo, *op* é um operador relacional e **valor** é constante válida para o atributo X_i

- Exemplos

- Sexo = Masculino
- Idade >= 20
- Sexo = Feminino and Idade < 90

101

Regra

- Uma regra assume a forma **if L then R** que é equivalente a $L \rightarrow R \equiv R \leftarrow L \equiv R :- L$
- As partes esquerda L e direita R são complexos sem atributos comuns entre eles, ou seja
 - $\text{atributos}(L) \cap \text{atributos}(R) = \emptyset$
- A parte esquerda **L** é denominada **condição**, **premissa**, **antecedente**, **cauda** ou **corpo** da regra
- A parte direita **R** é denominada **conclusão** ou **cabeça** da regra

102

Regra de Classificação

- ❑ Uma regra de classificação assume a forma restrita de uma regra
 - if L then classe = C_i
- ❑ ou simplesmente
 - if L then C_i
- ❑ onde C_i pertence ao conjunto de k valores de classe $\{C_1, C_2, \dots, C_k\}$
- ❑ A parte esquerda L é um complexo

103

Regra de Associação

- ❑ Uma regra de associação assume que não existe uma definição explícita de classe e qualquer atributo (ou atributos) pode ser usado como parte da conclusão da regra
- ❑ Exemplo
 - if $X_3='S'$ and $X_5 > 2$ then $X_1='N'$ and $X_2 < 1$

104

Cobertura

- ❑ Seja regra $L \rightarrow R$
 - Exemplos que satisfazem a parte L da regra são **cobertos** pela regra (ou a regra **dispara** para esses exemplos)
 - Exemplos que satisfazem tanto a condição L como a conclusão R são **cobertos corretamente** pela regra
 - Exemplos satisfazendo a condição L mas não a conclusão R são **cobertos incorretamente** pela regra
 - Exemplos que não satisfazem a condição L **não são cobertos** pela regra

Exemplos satisfazendo...	são...
$\neg L$	Não cobertos pela regra
L	Cobertos pela regra
$L \wedge R$	Cobertos corretamente pela regra
$L \wedge \neg R$	Cobertos incorretamente pela regra

105

Conteúdo

- ❑ Introdução
- ❑ Hierarquia do Aprendizado
- ❑ Paradigmas de Aprendizado
- ❑ AM: Conceitos e Definições
- ❑ Linguagens de Descrição

106

Linguagens de Descrição

- ❑ Qualquer que seja o tipo de aprendizado, é necessário uma linguagem para descrever objetos (ou possíveis eventos) e uma linguagem para descrever conceitos em termos computacionais
- ❑ Três tipos de DL:
 - Linguagem Descrição de Exemplos (Instance Description Language - IDL)
 - Linguagem de Descrição de Hipóteses (Hypotheses Description Language - HDL)
 - Linguagem de Descrição de Conhecimento do Domínio (Background Knowledge Language - BDL)

107

Linguagens de Descrição

- ❑ Lógica de ordem zero ou Proposicional
 - O objeto é representado e descrito em termos de conjunções, disjunções e negações de constantes booleanas que representam um campo
 - Ex:
 $fêmea \wedge adulta \rightarrow pode_ter_filhos$
- ❑ Lógica de atributos
 - Notação equivalente à LP, mas os atributos são tratados como variáveis
 - Ex:
 $sexo=fêmea \wedge idade=adulta \rightarrow classe=pode_ter_filhos$
ou
 $sexo(fêmea) \wedge idade(adulta) \rightarrow classe(pode_ter_filhos)$

108

Linguagens de Descrição

- Lógica de 1ª ordem ou Relacional**
 - Pode representar objetos como predicados que especificam propriedades ou relações
 - Cláusulas de Horn são um exemplo
 - Ex:

$$\text{macho}(X) \wedge \text{progenitor}(Z,X) \wedge \text{progenitor}(Z,Y) \rightarrow \text{irmão}(X,Y)$$
 ou

$$\text{irmão}(X,Y) \leftarrow \text{macho}(X) \wedge \text{progenitor}(Z,X) \wedge \text{progenitor}(Z,Y)$$
 ou

$$\text{irmão}(X,Y) :- \text{macho}(X), \text{progenitor}(Z,X), \text{progenitor}(Z,Y)$$
- Lógica de 2ª ordem**
 - Extensão da lógica de primeira ordem, em que os predicados podem ser considerados como variáveis
 - Ex:

$$P_1(X,Y) :- P_2(X), P_3(Z,X), P_4(Z,Y)$$
 pode ser instanciado com:

$$\text{irmão}(X,Y) :- \text{macho}(X), \text{progenitor}(Z,X), \text{progenitor}(Z,Y)$$
- Funções Matemáticas**

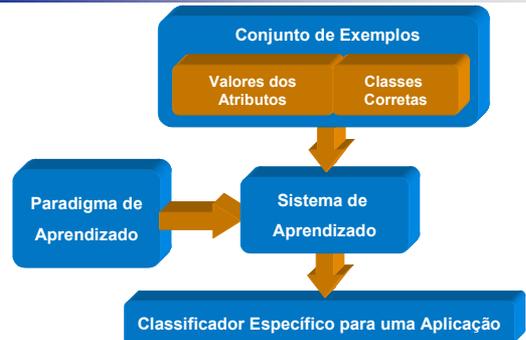
Linguagens de Descrição de Alguns Indutores

Indutor	IDL	HDL	BDL
C4.5	Atributo	Atributo	
CART	Atributo	Atributo	
CN2	Atributo	Atributo	
Ripper	Atributo	Atributo	
Foil	Atributo	Primeira Ordem	Atributo
Rede Neural	Atributo	Função Matemática	Primeira Ordem

Características Gerais dos Sistemas de AM

Modo de Aprendizado	Paradigmas de Aprendizado	Linguagens de Descrição	Formas de Aprendizado
- Supervisionado	- Simbólico	- Exemplos	- Incremental
- Não Supervisionado	- Estatístico	- Hipóteses	- Não Incremental
	- Instance-Based	- Conhecimento do Domínio	
	- Conexionista		
	- Genético		

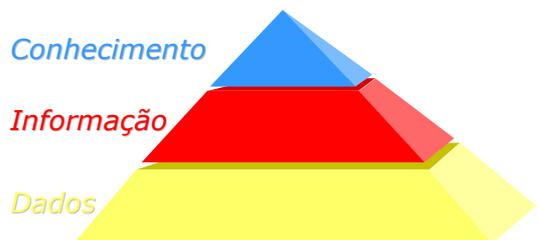
Representação da Classificação



Esforço em Cada Etapa



Estrutura



Dado, Informação, Conhecimento

- ❑ Dado: é a estrutura fundamental sobre a qual um sistema de informação é construído
- ❑ Informação: a transformação de dados em informação é freqüentemente realizada através da apresentação dos dados em uma forma compreensível ao usuário
- ❑ Conhecimento:
 - Fornece a capacidade de resolver problemas, inovar e aprender baseado em experiências prévias
 - Uma combinação de instintos, idéias, regras e procedimentos que guiam as ações e decisões

115

Importante Observar que...

- ❑ Dado não é Informação
- ❑ Informação não é Conhecimento
- ❑ Conhecimento não é Inteligência
- ❑ Inteligência não é Sabedoria

116

Resumo

- ❑ Nesta aula foram apresentados conceitos e definições de alguns termos amplamente utilizados em Aprendizado de Máquina, além de uma descrição sobre as principais linguagens de descrição
- ❑ A investigação de estruturas diferentes, que podem ser apropriadas para diferentes contextos, bem como o entendimento do seu poder e limitação são necessários para o uso com êxito de Aprendizado de Máquina
- ❑ Quanto maior a compreensão sobre as estruturas fundamentais usadas por classificadores, mais adequadamente pode-se aplicar ou alterá-las com base no conhecimento do domínio
- ❑ Além da compreensão dos algoritmos de AM, é igualmente importante poder avaliar seu desempenho, o que veremos nas próximas aulas

117

Slides baseados no Capítulo 4 do livro:

Rezende, S.O. (ed).
Sistemas Inteligentes, Manole, 2003,
ISBN 85-204-1683-7

Material elaborado por
José Augusto Baranauskas
Revisão 2007

118