

Aprendizagem de Máquina

Francisco de A.T. de Carvalho

Árvores de Decisão 1/

- Arvores de decisão: inductive decision trees
 - ∠Lógica de ordem 0+ (atributo/valor)

 - ≥ aprendizagem não incremental
 - ≥ estatística (admite exceções)

Árvores de Decisão 2/

- Arvores de decisão: inductive decision trees
 - - ∠Um problema complexo é decomposto em subproblemas mais simples.

Árvores de Decisão 3/

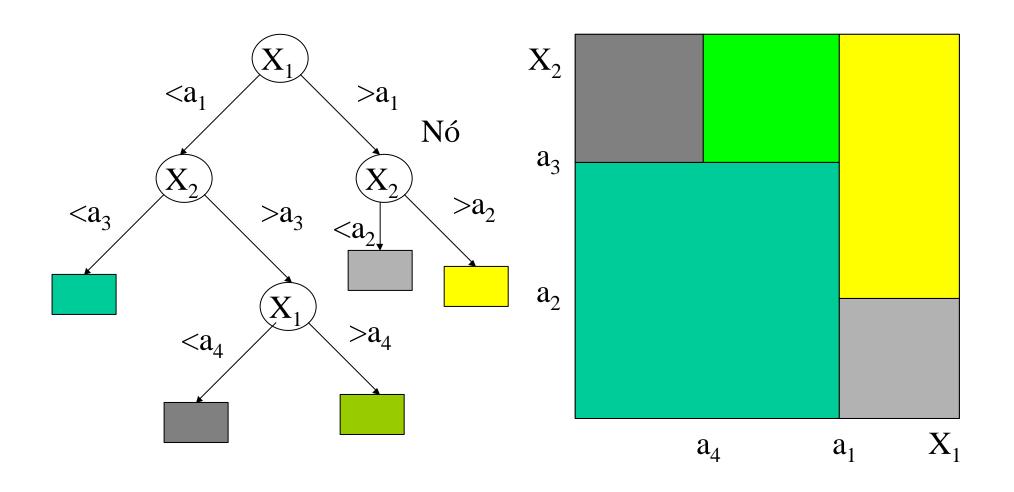
- Arvores de decisão: inductive decision trees
 - ≥ A capacidade de discriminação de uma arvore vem da:

 - ⊠A cada sub-espaço é associada uma classe.

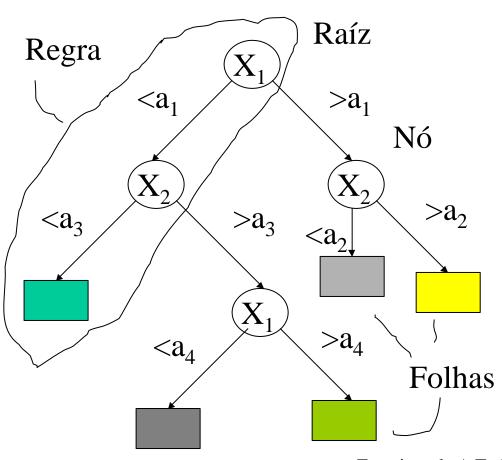
Árvores de Decisão 4/

- Arvores de decisão: inductive decision trees
 - **⊠**Crescente interesse
 - □ CART (Breiman, Friedman, et.al.)
 - ≥C4.5 (Quinlan)
 - S plus , Statistica, SPSS, SAS

 S plus , Statistica, SPSS, SAS



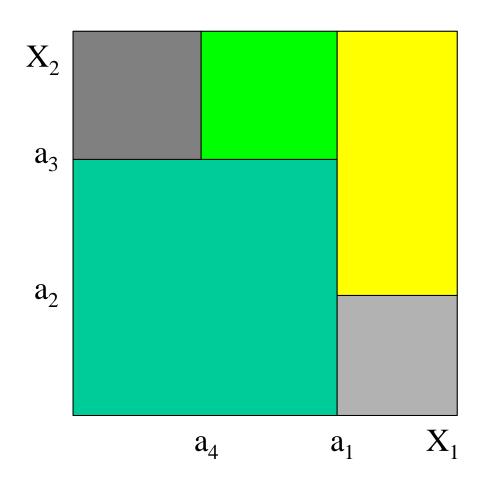
Árvores de Decisão 6/



- Representação por arvores de decisão:
 - Cada nó de decisão contem um teste num atributo.
 - Cada ramo descendente corresponde a um possível valor deste atributo.
 - Cada Folha está associada a uma classe.
 - Cada percurso na arvore (da raiz à folha) corresponde a uma regra de classificação.

Francisco de A.T. de Carvalho, CIn/UFPE

Árvores de Decisão 7/



- •No espaço definido pelos atributos:
 - Cada folha corresponde a uma região: Hiper-retângulo
 - A intersecção dos hiper retângulos é vazia
 - A união dos hiper-retângulos
 é o espaço completo

Árvores de Decisão 8/

- Arvores de decisão: inductive decision trees
 - ⊠A idéia *base*:
 - ≥ 1. Escolher um atributo.
 - ≥2. Estender a arvore adicionando um ramo para cada valor do atributo.
 - ≥3. Passar os exemplos para as folhas (tendo em conta o valor do atributo escolhido)

Árvores de Decisão 9/

- Arvores de decisão: inductive decision trees
 - ⊠A idéia *base*:
 - ≥4. Para cada folha
 - ≥1. Se todos os exemplos são da mesma classe, associar essa classe à folha
 - ≥2. Senão repetir os passos 1 a 4

Árvores de Decisão 10/

Arvores de decisão: inductive decision trees

⊠Exemplo: dados

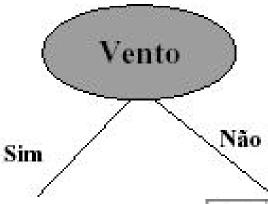
Tempo	Temperat	u Humidad	e vente	Joga
Sol	85	85	Não	Não
Sol	80	90	Sim	Não
Nublado	83	86	Não	Sim
Chuva	70	96	Não	Sim
Chuva	68	80	Não	Sim
Chuva	65	70	Sim	Não
Nublado	64	65	Sim	Sim
Sol	72	95	Não	Não
Sol	69	70	Não	Sim
Chuva	75	80	Não	Sim
Sol	75	70	Sim	Sim
Nublado	72	90	Sim	Sim
Nublado	81	75	Não	Sim
Chuva	71	91	Sim	Não

Árvores de Decisão 11/

Arvores de decisão: inductive decision trees

Exemplo: seleciona um atributo

Qual o melhor atributo?



Time.	Traperity.	Sheiter	(tests	200
58	10	- 98	Soc.	Slo
Chest	- 68	- 24	Sic	186
Skillah	- 44	-60	lin	No.
54	185	19	Ster	The .
9858	75	16	Gi .	The contract of
Class	10.7	- 21	this :	500

Terps	Toposta	Third lab	1980	Japa
100	80	10.0	200	700
Skillete	65.	196	1900	Ship
Own	76	- 64	Nin	Sec.
des	100	PF.	500	Mile
24	72	90	1996	7990
Sid	60.0	- 74	Sim	Stre
Own:	78.1	100	Na	State
Marian	- 84	20	900	24-16

Francisco de A.T. de Carvalho, CIn/UFPE

Árvores de Decisão 12/

- Arvores de decisão: inductive decision trees

 - Existem muitas medidas.
 - ⊠Todas concordam em dois pontos:

Árvores de Decisão 13/

Caracterização das medidas de partição



- Medida da diferença dada por uma função baseada nas proporções das classes entre o nó corrente e os nós descendentes.
- Valoriza a pureza das partições: Gini, entropia

Árvores de Decisão 14/

⊠Entropia

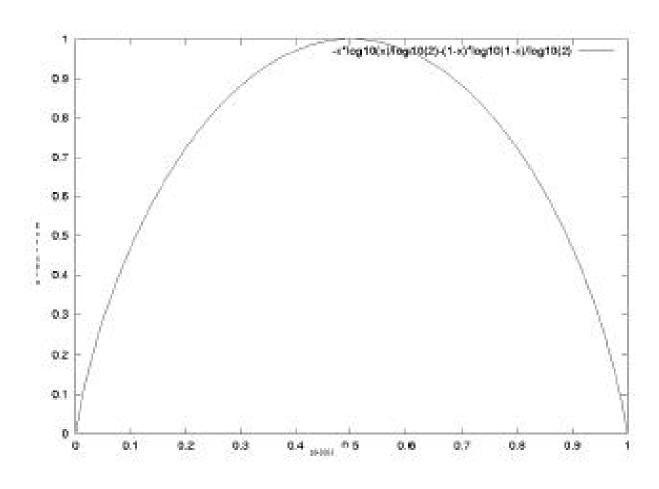
- Entropia é uma medida da aleatoriedade de uma variável.
- \bowtie A entropia de uma variável nominal X que pode tomar i valores:

$$entropia(X) = -\sum_{i} p_{i} \log_{2} p_{i}$$

- \bowtie A entropia tem máximo (log2 i) se $p_i = p_j$ para qualquer i \neq j
- \bowtie A entropia(x) = 0 se existe um *i* tal que p_i = 1
- \boxtimes É assumido que 0 * log2 0 = 0

Árvores de Decisão 15/

⊠Entropia



Francisco de A.T. de Carvalho, CIn/UFPE

Árvores de Decisão 16/

⊠Ganho de Informação

- No contexto das arvores de decisão a entropia é usada para estimar a aleatoriedade da variável a prever (classe)
- - ☑O ganho de informação mede a redução da entropia causada pela partição dos exemplos de acordo com os valores do atributo.

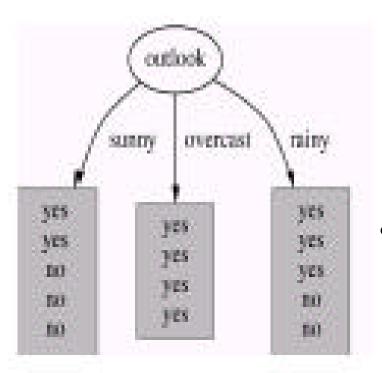
Árvores de Decisão 17/

⊠Ganho de Informação

$$ganho(Exs, Atri) = entropia(Exs) - \sum_{n} \frac{\#Exs_{n}}{\#Exs} entropia(Exs_{n})$$

A construção de uma arvore de decisão é guiada pelo objetivo de diminuir a entropia ou seja a aleatoriedade - dificuldade de previsão- da variável que define as classes.

Árvores de Decisão 18/



• Informação da Classe:

$$- p(sim) = 9/14$$

$$- p(n\tilde{a}o) = 5/14$$

$$- Info(joga) = -9/14 log2 9/14 -5/14 log2 5/14 = 0.940$$

• Informação nas partições:

$$-p(\sin|\text{tempo}=\text{sol}) = 2/5$$

$$-p(n\tilde{a}o|tempo=sol) = 3/5$$

Árvores de Decisão 19/

- Informação nas partições:
 - Info(joga|tempo=sol)

$$- = -2/5 \log_2 2/5 - 3/5 \log_2 3/5 = 0.971$$

- Info(joga|tempo=nublado) = 0.0
- Info(joga|tempo=chuva) = 0.971

<pre>- Info(tempo) =</pre>	5/14*0.971	+4/14*0+5/14*	60.971 = 0.693
----------------------------	------------	---------------	----------------

• Ganho de Informação obtida neste atributo:

$$-$$
 Ganho(tempo) = $0.940 - 0.693 = 0.247$

Árvores de Decisão 20/

- ⊠Cálculo do ganho para atributos numéricos
 - - ⊠Exemplos onde valor_do_atributo < ponto_referência
 - Exemplos onde valor_do_atributo > ponto_referência

 □

 Exemplos onde valor_do_atributo > ponto_referência

 □

 Exemplos onde valor_do_atributo > ponto_referência

Árvores de Decisão 21/

- ⊠Cálculo do ganho para atributos numéricos
 - Escolha do ponto de referência:

 - Qualquer ponto intermediário entre dois valores diferentes e consecutivos dos valores observados no conjunto de treinamento pode ser utilizado como possível ponto de referência.

Árvores de Decisão 22/

⊠Cálculo do ganho para atributos numéricos

- É usual considerar o valor médio entre dois valores diferentes e consecutivos.
- □ Fayyard e Irani (1993) mostram que de todos os possíveis pontos de referência aqueles que maximizam o ganho de informação separam dois exemplos de classes diferentes.

Árvores de Decisão 23/

⊠Cálculo do ganho para atributos numéricos

peratu.	Joga
64	Sim
65	Não
68	Sim
69	Sim
70	Sim
71	Não
72	Não
72	Sim
75	Sim
75	Sim
80	Não
81	Sim
83	Sim
85	Não

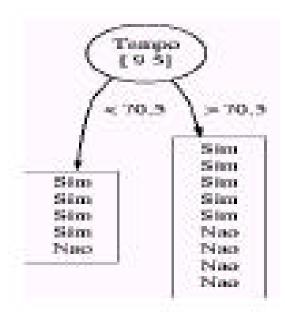
Considere o ponto de referência temperatura = 70.5

- Um teste usando este ponto de referência divide os exemplos em duas classes:
 - Exemplos onde temperatura < 70.5
 - Exemplos onde temperatura > 70.5
- Como medir o ganho de informação desta partição?

Árvores de Decisão 24/

⊠Cálculo do ganho para atributos numéricos

- Como medir o ganho de informação desta partição?
- Informação nas partições
 - $p(sim \mid temperatura < 70.5) = 4/5$
 - p(não | temperatura<70.5)=1/5
 - $p(sim \mid temperatura > 70.5) = 5/9$
 - $-p(n\tilde{a}o \mid temperatura > 70.5) = 4/9$



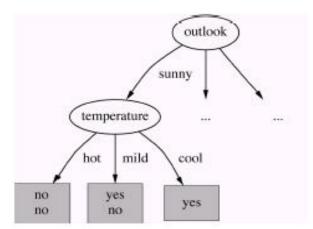
Árvores de Decisão 25/

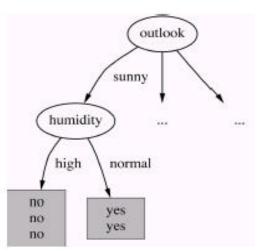
⊠Cálculo do ganho para atributos numéricos

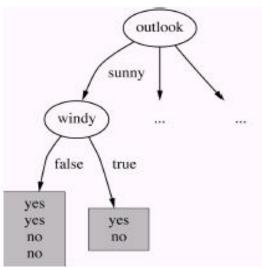
- Info(joga | temperatur<70.5) = -4/5 $\log_2 4/5 1/5$ $\log_2 1/5 = 0.721$
- Info(joga | temperatura >70.5) = -5/9 $\log_2 5/9 4/9$ $\log_2 4/9 = 0.991$
- Info(temperatura) = 5/14*0.721+9/14*0.991 = 0.895
- Ganho(temperatura) = 0.940 0.895 = 0.045 bits

Árvores de Decisão 26/

Repetir o processo...

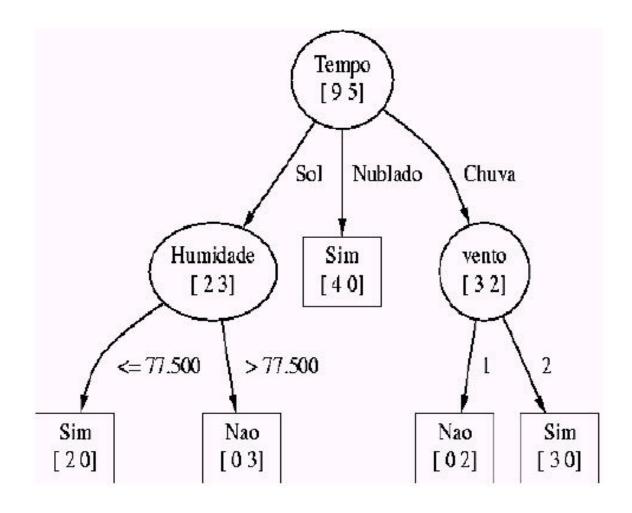






Árvores de Decisão 27/

✓Árvore de Decisão Final



Árvores de Decisão 28/

⊠Critérios de parada

- - ⊠Todos os exemplos pertencem a mesma classe.

 - ⊠O número de exemplos é inferior a um certo limite.

Árvores de Decisão 29/

- ⊠Construção de uma árvore de decisão
 - ⊠Input: Um conjunto exemplos
 - ⊠Output: Uma arvore de decisão
 - - \boxtimes Se criterio_parada(Exs) = TRUE: retorna Folha
 - ⊠Escolhe o atributo que maximiza o critério_divisão(Exs)
 - \bowtie Para cada partição *i* dos exemplos baseada no atributo escolhido: Arvore_i = GeraArvore(Exsi)

 - **⊠**Fim

Árvores de Decisão 30/

- ⊠Construção de uma árvore de decisão
 - ⊠O problema de construir uma arvore de decisão:
 - ⊠Consistente com um conjunto de exemplos

 - ⊠É um problema NP completo.
 - **⊠**Dois problemas:

Árvores de Decisão 31/

- ⊠Construção de uma árvore de decisão
 - ⊠Os algoritmos mais populares:

 - Não reconsideram as opções tomadas.

Árvores de Decisão 32/

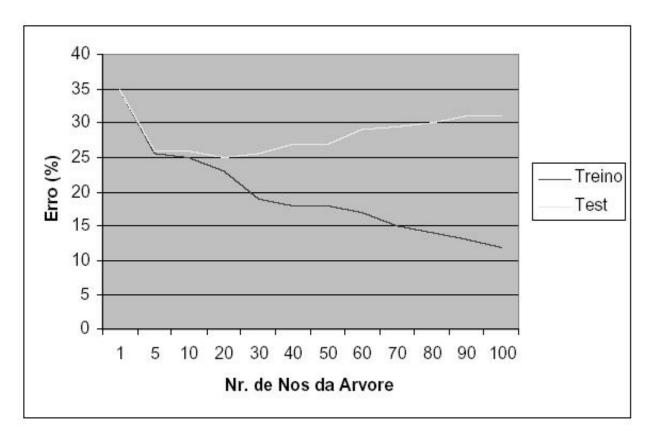
- ≥ Sobre-ajustamento (overfitting)
 - - Em domínios sem ruído o número de erros no conjunto de treinamento pode ser 0.

Árvores de Decisão 33/

- Sobre-ajustamento (overfitting)
 - Em problemas com *ruído* esta capacidade é problemática:
 - A partir de uma certa profundidade as decisões tomadas são baseadas em pequenos conjuntos de exemplos.
 - A capacidade de generalização para exemplos não utilizados no crescimento da arvore diminui.

Árvores de Decisão 34/

≥ Variação do erro com o número de nós



Árvores de Decisão 35/

- Sobre-ajustamento (overfitting)
 - ⊠Definição:
 - \boxtimes Uma arvore de decisão d faz sobre-ajustamento aos dados se existir uma arvore d tal que:
 - $\boxtimes d$ tem menor erro que d' no conjunto de treinamento mas d tem menor erro na população.
 - ⊠Como pode acontecer:
 - ⊠Ruído nos dados

Árvores de Decisão 36/

- Sobre-ajustamento (overfitting)
 - - ∠ Uma arvore de decisão pode obter um ajuste perfeito aos dados de treinamento.

Árvores de Decisão 37/

- Sobre-ajustamento (overfitting)
 - ⊠Occam's razor: preferência pela hipótese mais simples.
 - Existem menos hipóteses simples do que complexas.
 - ≥Se uma hipótese simples explica os dados é pouco provável que seja uma coincidência.

Árvores de Decisão 38/

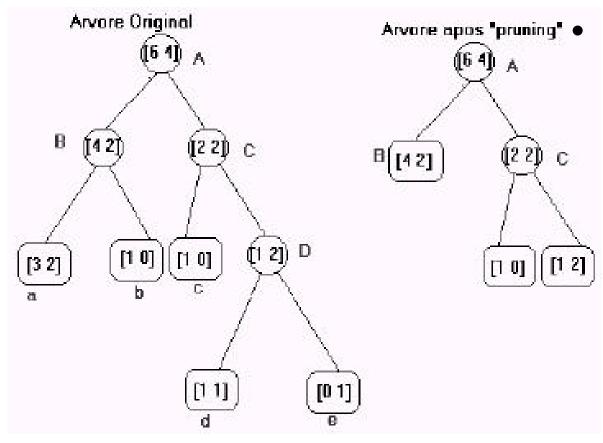
- **Simplificar** a árvore
 - **⊠**Duas possibilidades:
 - Parar o crescimento da arvore mais cedo (pre-pruning).

Árvores de Decisão 39/

- - ≥ Percorre a arvore em profundidade
 - ∠Para cada nó de decisão calcula:
 - ⊠Erro no nó
 - Soma dos erros nos nós descendentes
 - Se o erro no nó é menor ou igual à soma dos erros dos nós descendentes o nó é transformado em folha.

Árvores de Decisão 40/

⊠Um algoritmo básico de pruning



Francisco de A.T. de Carvalho, CIn/UFPE

Exemplo do nó B:

- Erro no nó = 2
- Soma dos erros nos nós descendentes:
 - 2+0
- Transforma o nó em folha
 - Elimina os nós descendentes.

Árvores de Decisão 41/

⊠Critérios

- Obter estimativas do erro a partir do conjunto de treinamento.
- Estimar o erro num conjunto de validação independente do utilizado para construir a arvore.
 - Reduz o volume de informação disponível para crescer a arvore.

Árvores de Decisão 42/

≥Estimativas do erro

- ⊠O "Cost complexity pruning", Breiman, 1984

 - ⊠Cart (Breiman et al.)
- **⊠O** "Error based pruning"

 - ⊠Assume uma distribuição Binomial para os exemplos de um nó.
 - ⊠Usado no C5.0

Árvores de Decisão 43/

- **∠**Valores ausentes
 - **⊠Pré-Processados**
 - - Atributos nominais: moda.
 - ≥Na construção do modelo

Árvores de Decisão 44/

∠Valores ausentes

- Na construção do modelo: as observações a serem classificadas não apresentam valores ausentes
 - ⊠usar apenas a informação disponível.
 - Exemplo: Nó N com n observações descritas por 3 atributos e apenas uma observação com o atributo x₃ ausente
 - \square Calcular a redução da impureza usando n pontos nos atributos x_1 e x_2 e n-1 quando se tratar do atributo x_3

Árvores de Decisão 45/

- ⊠Transformação de árvores em regras de decisão
 - ⊠Regras podem ser auto-interpretadas.
 - ⊠Uma transformação:
 - ⊠Cada ramo dá origem a uma regra
 - ⊠A regra prediz a classe associada á folha
 - A parte condicional da regra é obtida pela conjunção das condições de cada nó.

Árvores de Decisão 46/

- ⊠Transformação de árvores em regras de decisão
 - Em cada regra é testado a eliminação de condições. Uma condição é eliminada se:
 - **⊠**O erro não aumenta
 - MA estimativa de erro não aumenta

Árvores de Decisão 47/

⊠Transformação de árvores em regras de decisão

Árvores de Decisão 48/

⊠Transformação de árvores em regras de decisão

```
pruning rule for setosa
                                            pruning rule for virginica
                                                          Pess Absent condition
           Pess Absent condition
Err Used
                                                       4.5
           3.0% <base rule>
                                                       3 37.0% petal-width > 0.6
90 135 69.8% pet.width <= 0.6
                                                     3 37.0% petal-width <= 1.7
                                                 40 43 96.2% petal-length > 4.9
Rule 1:
                                                       6 55.7% petal-width <= 1.5
    petal-width <= 0.6
                                                eliminate test 2
    -> class setosa [97.0%]
                                                 Err Used
                                                          Pess Absent condition
                                                      3 37.0% <base rule>
                                                       3 37.0% petal-width > 0.6
pruning rule for versicolor
                                                40 43 96.2% petal-length > 4.9
                   Absent condition
 Err Haed
            Pess.
                                                          9.0% petal-width <= 1.5
          6.0% <base rule>
                                                eliminate test 3
                                                 Err Used Pess Absent condition
       88 56.5% petal-width > 0.6
                                                  2 42 9.0% <br/>
<br/>
*base rule>
       48 15.0% petal-width <= 1.7
                                                  2 42 9.0% petal-width > 0.6
       49 14.7% petal-length <= 4.9
                                                      90 54.2% petal-length > 4.9
                                                eliminate test 1
Rule 2:
                                                 Err Used Pess Absent condition
    petal-length <- 4.9
                                                          9.04 <base rules
                                                   2 42
                                                90 135 69.8% petal-length > 4.9
    petal-width > 0.6
    petal-width <= 1.7
                                            Rule 3:
    -> class versicolor
                          194.0%1
                                                     petal-length > 4.9
                                                -5 class virginica [91.0%]
```

Árvores de Decisão 49/

- ⊠Transformação de árvores em regras de decisão
 - **⊠**Porque regras

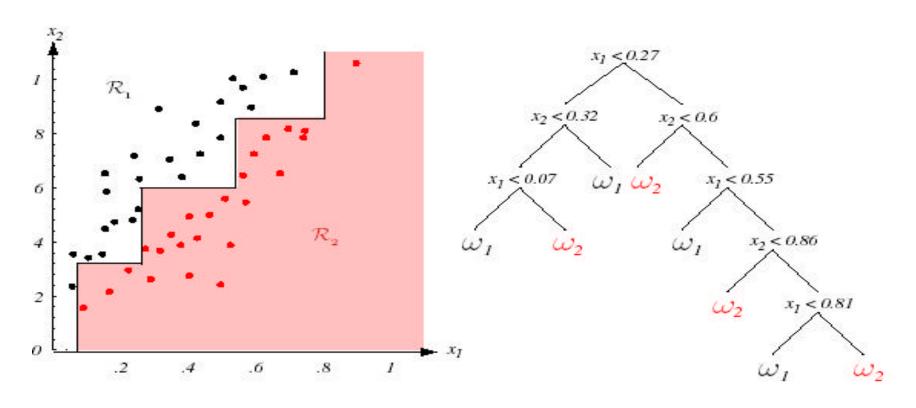
 - Permite eliminar um teste numa regra, mas pode reter o teste em outra regra.
 - Elimina a distinção entre testes perto da raiz e testes perto das folhas.

Árvores de Decisão 50/

- ≥ Pre-processamento
 - ⊠Motivação
 - Árvores de Decisão podem ser mais efetivas se técnicas de pré-processamento forem usadas (componentes principais, no caso de dados numéricos).

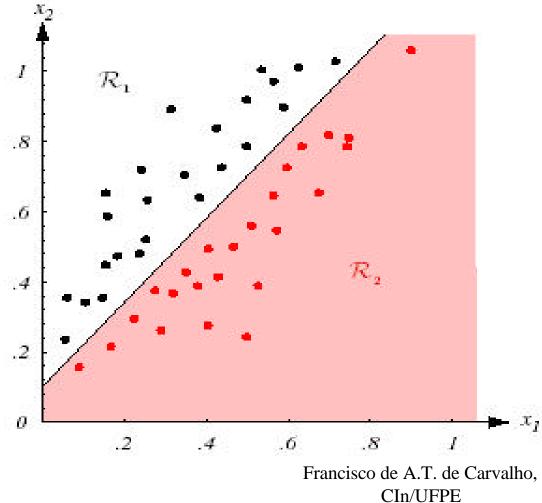
Árvores de Decisão 51/

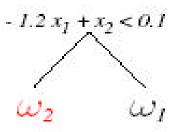
≥ Pre-processamento



Árvores de Decisão 52/

≥ Pre-processamento





Árvores de Decisão 53/

- - Método não-paramétrico
 - Não assume nenhuma distribuição particular para os dados.

Árvores de Decisão 54/

- - Elevado grau de interpretabilidade
 - ⊠É eficiente na construção de modelos:
 - ⊠Complexidade média O(n log n)
 - - Mecanismo de seleção de atributos.

Árvores de Decisão 55/

- - ⊠Instabilidade
 - Pequenas perturbações do conjunto de treinamento podem provocar grandes alterações no modelo aprendido.
 - ≥ Presença de valores desconhecidos
 - - ⊠Replicação de sub-arvores