

Cap.6 – Árvores de Decisão

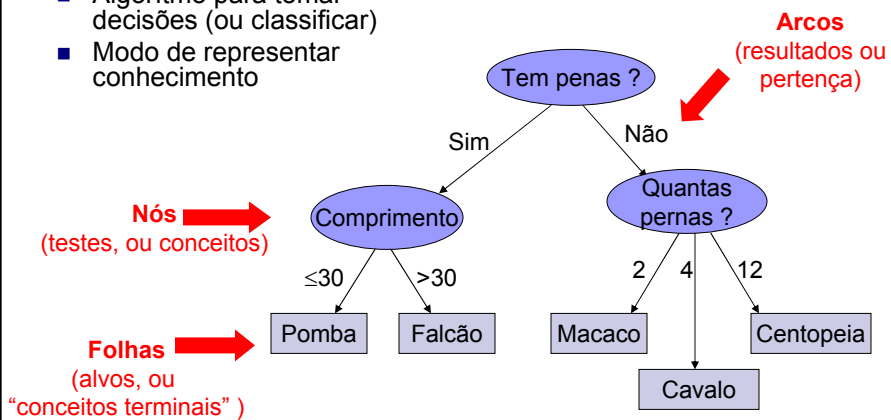
V 3.0, V.Lobo, EN/ISEGI, 2005

Árvores de decisão

Victor Lobo

O que é uma árvore de decisão ?

- Algoritmo para tomar decisões (ou classificar)
- Modo de representar conhecimento

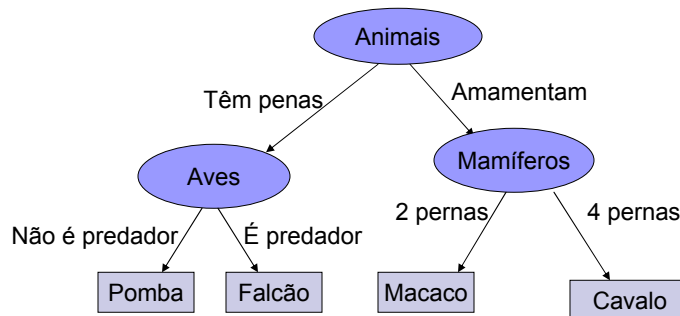
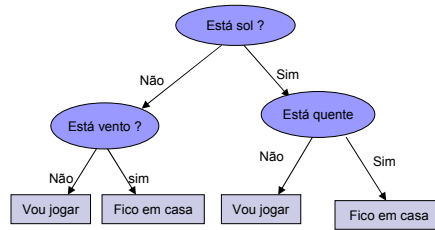


Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Outras árvores

- Sempre “divide and conquer” !



Vantagens das árvores

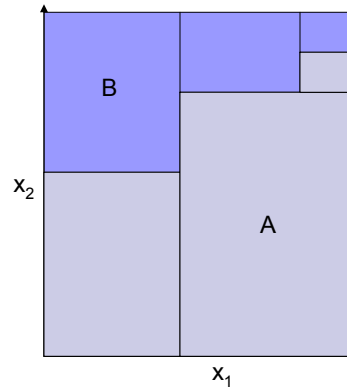
- Interpretação
 - Percebe-se a razão da decisão
- Facilidade em lidar com diversos tipos de informação
 - Real, nominal, ordinal, etc
- Facilidade em lidar com informação incompleta
- Escolha automática dos atributos mais relevantes em cada caso
- Adaptável também a problemas de regressão

Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

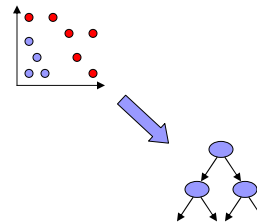
Desvantagens de árvores

- Fronteiras lineares e perpendiculares aos eixos
- Grande sensibilidade a pequenas perturbações no conjunto de treino (geram redes muito diferentes)



Indução de árvores de decisão

- A partir de um conjunto de treino, construir uma árvore
- Problemas:
 - Que pergunta fazer ?
 - Que variável interrogar ?
 - Qual o valor de corte ?
 - Qual o nó a “partir” ?
 - Quantos ramos pôr em cada nó ?
 - Quando parar ?



Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Algoritmo de indução de árvores de decisão

- Para cada nível divide o conjunto de exemplos em partições alternativas.
- Utilizando uma medida da *QUALIDADE* da partição selecciona a melhor partição.
- Para a partição seleccionada, volta a repetir o processo para cada um dos elementos da partição.

Algoritmos mais usados

- CART
 - Classification and regression trees (Breiman 84)
- ID3, C4.5 e C5
 - Iterative Dichotomizer 3
- Muitas outras variantes...
 - Em SAS: possibilidade de seleccionar os diferentes parâmetros para a construção da árvore

Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Algoritmo DDT (devisive decision tree - Hunt 62)

- Assume-se que existe uma atributo especial a “Classe” e que os exemplos foram previamente classificados.
- Cada nó especifica um único atributo, que é usado como teste, designado por atributo mais discriminante.
- N – o nó N
- ASET – Attribute Set – Conjunto de atributos
- ISET – Instance Set – Conjunto de exemplos

DDT(N , ASET, ISET)

Se o conjunto $ISET$ é vazio **então** o nó terminal N é da classe *desconhecida*
senão

Se todas os exemplos de $ISET$ são da mesma classe
então o nó terminal N tem o nome da classe

senão

Para cada atributo A do conjunto de atributos $ASET$

Avalia A de acordo com a capacidade de discriminar a classe

Selecciona o atributo B que tem o melhor valor discriminante

Para cada valor V do melhor atributo B

Cria um novo filho C do nó N

Coloca o par atributo valor (B, V) em C

Seja $JSET$ o conjunto de exemplos de $ISET$ com o valor V em B

Seja $KSET$ o conjunto de atributos de $ASET$ com B removido

DDT(C , $KSET$, $JSET$)

Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Pesquisa

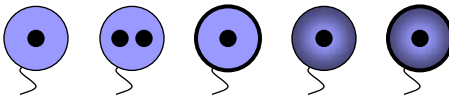
- É uma pesquisa “gulosa”.
- Não tem “backtracking”.
- Pode ficar presa num mínimo local.

O “bias” desta aproximação indutiva é que as árvores mais pequenas são preferíveis às árvores grandes.

(Occam’s razor: prefere a hipótese mais simples que justifica os dados - 1320)

Exemplo (Análise de células)

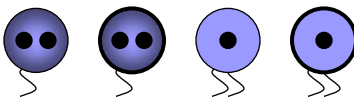
Lethargia



Burpoma



Saudável



Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Forma tabelar

# Núcleos	# Caudas	Cor	Membrana	Classe
1	1	Clara	Fina	Lethargia
2	1	Clara	Fina	Lethargia
1	1	Clara	Grossa	Lethargia
1	1	Escura	Fina	Lethargia
1	1	Escura	Grossa	Lethargia
2	2	Clara	Fina	Burpoma
2	2	Escura	Fina	Burpoma
2	2	Escura	Grossa	Burpoma
2	1	Escura	Fina	Saudável
2	1	Escura	Grossa	Saudável
1	2	Clara	Fina	Saudável
1	2	Clara	Grossa	Saudável

Métrica

- Seja a medida de discriminação do atributo

$$f(A) = \frac{1}{n} \sum_{i=1}^{|A|} C_i$$

em que n é o número total de exemplos e C_i é o número de exemplos correctamente classificados pela classe mais frequente.

É uma medida da “dominância” ou “pureza”

Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Se fizermos a partição pelo nº de núcleos...

Tabelas

# núcleos	1	2
Lethargia	4	1
Burpoma	0	3
Saudável	2	2

# Núcleos	# Caudas	Cor	Membrana	Classe
1	1	Clara	Fina	Lethargia
2	1	Clara	Fina	Lethargia
1	1	Clara	Grossa	Lethargia
1	1	Escura	Fina	Lethargia
1	1	Escura	Grossa	Lethargia
2	2	Clara	Fina	Burpoma
2	2	Escura	Fina	Burpoma
2	2	Escura	Grossa	Burpoma
2	1	Escura	Fina	Saudável
2	1	Escura	Grossa	Saudável
1	2	Clara	Fina	Saudável
1	2	Clara	Grossa	Saudável

Poder discriminante:
 $(4 + 3) / 12 = 0.58$

Se fizermos a partição pelo nº de caudas...

Tabelas

# caudas	1	2
Lethargia	5	0
Burpoma	0	3
Saudável	2	2

# Núcleos	# Caudas	Cor	Membrana	Classe
1	1	Clara	Fina	Lethargia
2	1	Clara	Fina	Lethargia
1	1	Clara	Grossa	Lethargia
1	1	Escura	Fina	Lethargia
1	1	Escura	Grossa	Lethargia
2	2	Clara	Fina	Burpoma
2	2	Escura	Fina	Burpoma
2	2	Escura	Grossa	Burpoma
2	1	Escura	Fina	Saudável
2	1	Escura	Grossa	Saudável
1	2	Clara	Fina	Saudável
1	2	Clara	Grossa	Saudável

Poder discriminante:
 $(5 + 3) / 12 = 0.67$

Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Se fizermos a partição peça cor ...

Tabelas

Cor	Clara	Escura
Lethargia	3	2
Burpoma	1	2
Saudável	2	2

Poder discriminante:
 $(3 + 2) / 12 = 0.41$

# Núcleos	# Caudas	Cor	Membrana	Classe
1	1	Clara	Fina	Lethargia
2	1	Clara	Fina	Lethargia
1	1	Clara	Grossa	Lethargia
1	1	Escura	Fina	Lethargia
1	1	Escura	Grossa	Lethargia
2	2	Clara	Fina	Burpoma
2	2	Escura	Fina	Burpoma
2	2	Escura	Grossa	Burpoma
2	1	Escura	Fina	Saudável
2	1	Escura	Grossa	Saudável
1	2	Clara	Fina	Saudável
1	2	Clara	Grossa	Saudável

Se fizermos a partição pelo tipo de membrana...

Tabelas

Membrana	Fina	Grossa
Lethargia	3	2
Burpoma	2	1
Saudável	3	1

Poder discriminante:
 $(3 + 2) / 12 = 0.41$

# Núcleos	# Caudas	Cor	Membrana	Classe
1	1	Clara	Fina	Lethargia
2	1	Clara	Fina	Lethargia
1	1	Clara	Grossa	Lethargia
1	1	Escura	Fina	Lethargia
1	1	Escura	Grossa	Lethargia
2	2	Clara	Fina	Burpoma
2	2	Escura	Fina	Burpoma
2	2	Escura	Grossa	Burpoma
2	1	Escura	Fina	Saudável
2	1	Escura	Grossa	Saudável
1	2	Clara	Fina	Saudável
1	2	Clara	Grossa	Saudável

Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Tabelas

# núcleos	1	2
Lethargia	4	1
Burpoma	0	3
Saudável	2	2

0.58

Cor	Clara	Escura
Lethargia	3	2
Burpoma	1	2
Saudável	2	2

0.41

# caudas	1	2
Lethargia	5	0
Burpoma	0	3
Saudável	2	2

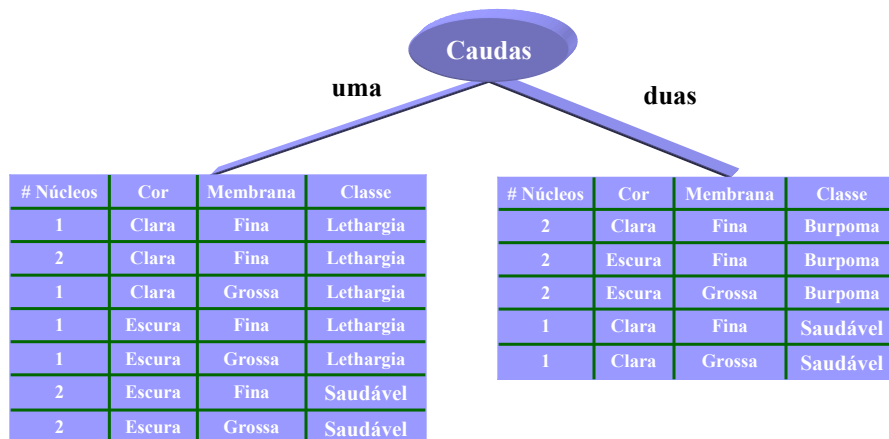
0.67

Membrana	Fina	Grossa
Lethargia	3	2
Burpoma	2	1
Saudável	3	1

0.41

Escolha: # Caudas

A árvore inicial



Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Tabelas sabendo #caudas = 1

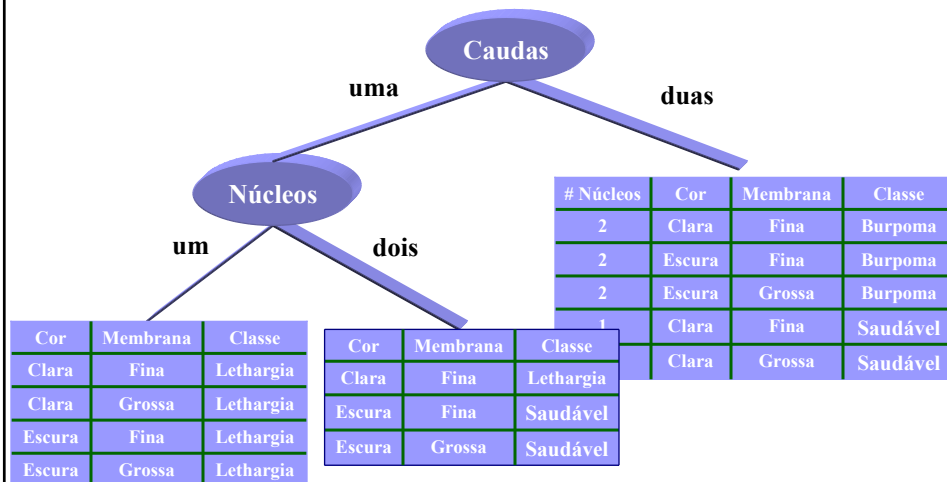
# núcleos	1	2
Lethargia	4	1
Burpoma	0	0
Saudável	0	2

# Núcleos	Cor	Membrana	Classe
1	Clara	Fina	Lethargia
2	Clara	Fina	Lethargia
1	Clara	Grossa	Lethargia
1	Escura	Fina	Lethargia
1	Escura	Grossa	Lethargia
2	Escura	Fina	Saudável
2	Escura	Grossa	Saudável

Cor	Clara	Escura
Lethargia	3	2
Burpoma	0	0
Saudável	0	2

Membrana	Fina	Grossa
Lethargia	3	2
Burpoma	0	0
Saudável	0	2

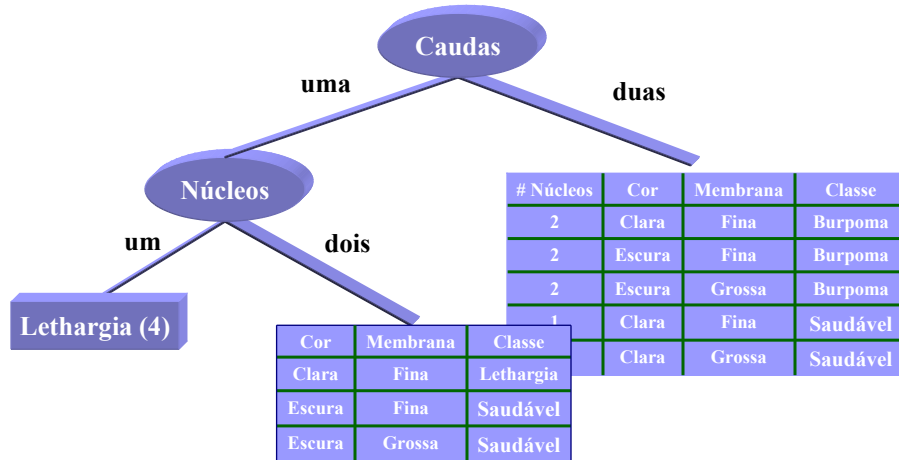
A árvore (continuação 1)



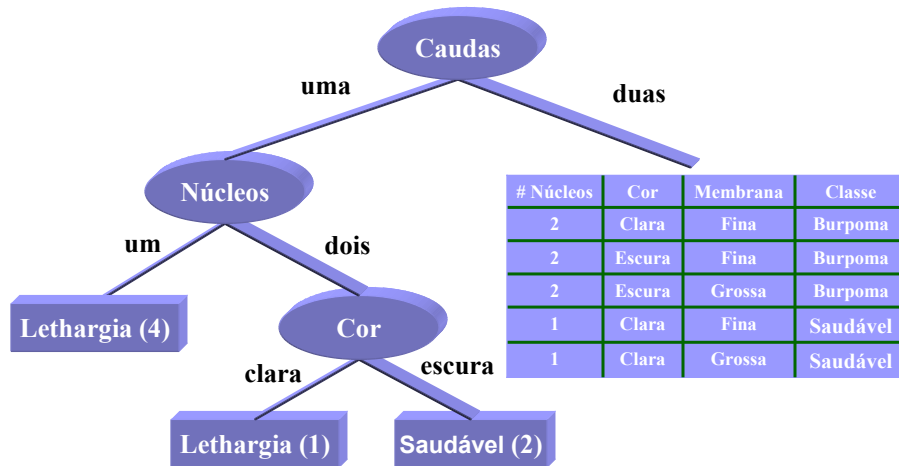
Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

A árvore (continuação 2)



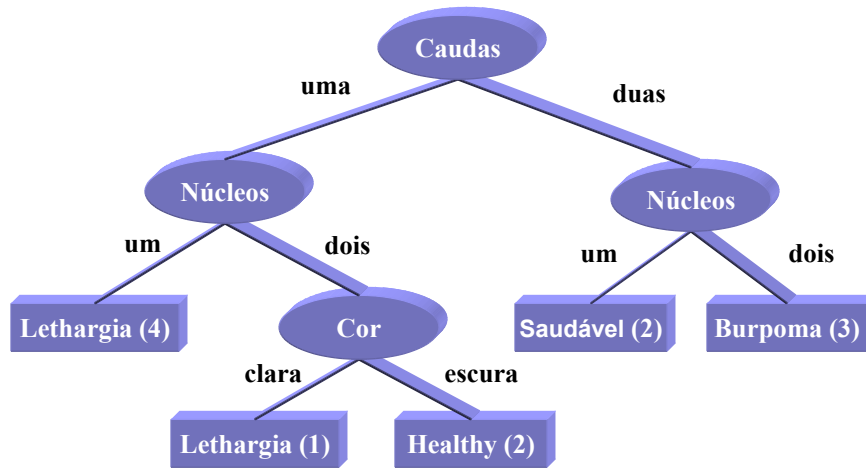
A árvore (continuação 3)



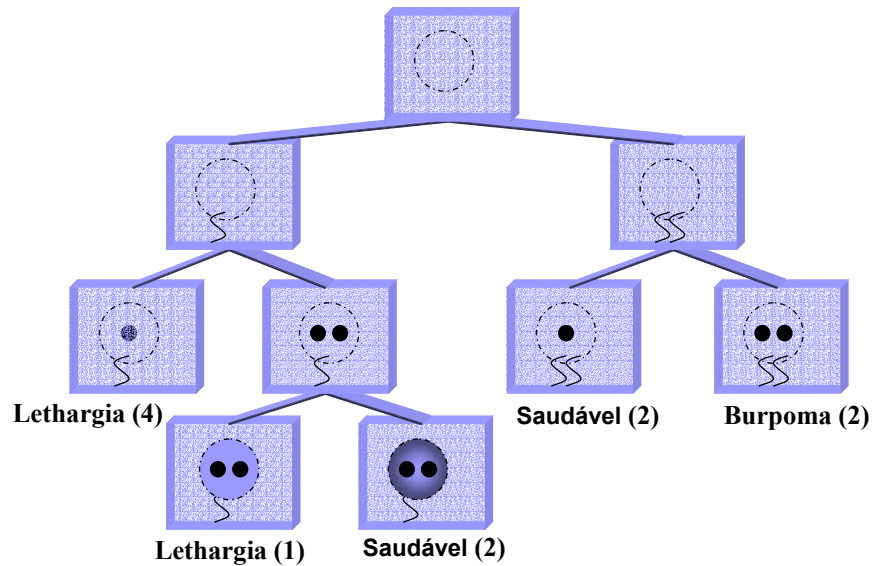
Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

A árvore final



Outra interpretação



Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Descrição das três classes

$$(caudas = uma) \wedge (núcleos = um)$$

∨

$$(caudas = uma) \wedge (núcleos = dois) \wedge (cor = clara) \rightarrow Lethargia$$

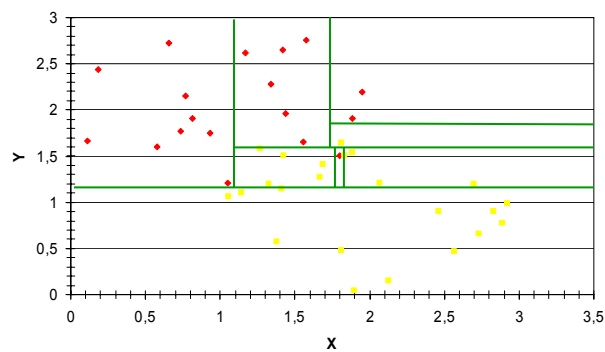
$$(caudas = duas) \wedge (núcleos = um)$$

∨

$$(caudas = uma) \wedge (núcleos = 2) \wedge (cor = escura) \rightarrow Saudável$$

$$(caudas = duas) \wedge (núcleos = dois) \rightarrow Burpoma$$

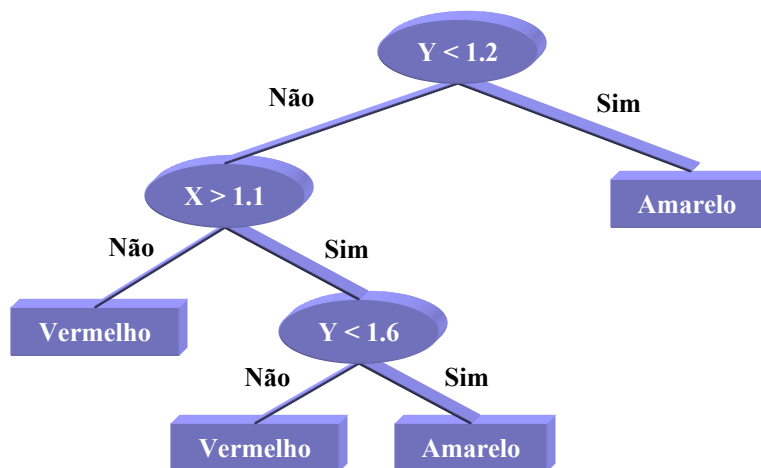
Outro exemplo



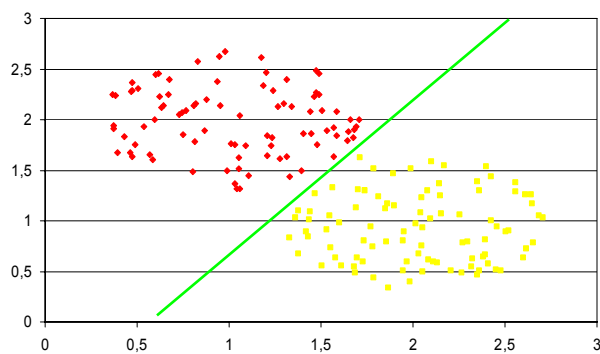
Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

A outra árvore



Qual a melhor solução?



Cap.6 – Árvores de Decisão

V 3.0, V.Lobo, EN/ISEGI, 2005

Crítérios para escolher as partições

- χ^2
 - Mede a significância da diferença entre os erros obtidos apenas com o “nó mãe” e com os “nós filhos”
- Entropia
 - Mede a “pureza” de um nó pela entropia que é definida como sendo
 - $E = p \log(p)$ onde p é a probabilidade dos exemplos terem uma dada classe
- Gini
 - Parecido com a entropia, mas evita o cálculo do logaritmo usando apenas $G = p(1-p)$