

# Aprendizado Simbólico: Árvores de Decisão

---

SCC-230 – Inteligência Artificial

Thiago A. S. Pardo  
Daniel Honorato

1

## Aprendizado Simbólico

---

- **Conhecimento explicitado e interpretável** por humanos
  
- **Classificadores por Árvores de Decisão (ADs)**
  - CART (1977)
  - ID3 (1979)
  - ASSISTANT (1984)
  - C4.5 (1993), aprimoramento do ID3

---

2

## Árvores de Decisão (ADs)

- Um dos **métodos mais usados** e práticos para **inferência indutiva**
  - Conhecimento estrutural do domínio
  
- Indução a partir de um conjunto de **dados rotulados** (classificados)
  - Aprendizado supervisionado
  
- Algumas áreas de aplicação são **medicina** (diagnóstico médico) e **análise de risco de crédito** para novos clientes de banco

3

## Exemplo

<b>dia</b>	<b>aparência</b>	<b>temperatura</b>	<b>umidade</b>	<b>vento</b>	<b>jogar_tênis</b>
D1	ensolarado	quente	alta	fraco	<b>não</b>
D2	ensolarado	quente	alta	forte	<b>não</b>
D3	nublado	quente	alta	fraco	<b>sim</b>
D4	chuva	moderada	alta	fraco	<b>sim</b>
D5	chuva	fria	normal	fraco	<b>sim</b>
D6	chuva	fria	normal	forte	<b>não</b>
D7	nublado	fria	normal	forte	<b>sim</b>
D8	ensolarado	moderada	alta	fraco	<b>não</b>
D9	ensolarado	fria	normal	fraco	<b>sim</b>
D10	chuva	moderada	normal	fraco	<b>sim</b>
D11	ensolarado	moderada	normal	forte	<b>sim</b>
D12	nublado	moderada	alta	forte	<b>sim</b>
D13	nublado	quente	normal	fraco	<b>sim</b>
D14	chuva	moderada	alta	forte	<b>não</b>

4

## Exemplo

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4					
D5					
D6					
D7					
D8	e				
D9	e				
D10					
D11	e				
D12					
D13					
D14					

The decision tree starts with the root node 'aparência'. It branches into three paths: 'sol', 'nublado', and 'chuva'. The 'sol' path leads to the 'umidade' node, which branches into 'alta' (leading to 'não') and 'normal' (leading to 'sim'). The 'nublado' path leads directly to 'sim'. The 'chuva' path leads to the 'vento' node, which branches into 'forte' (leading to 'não') and 'fraco' (leading to 'sim').

## Quando Usar ADs

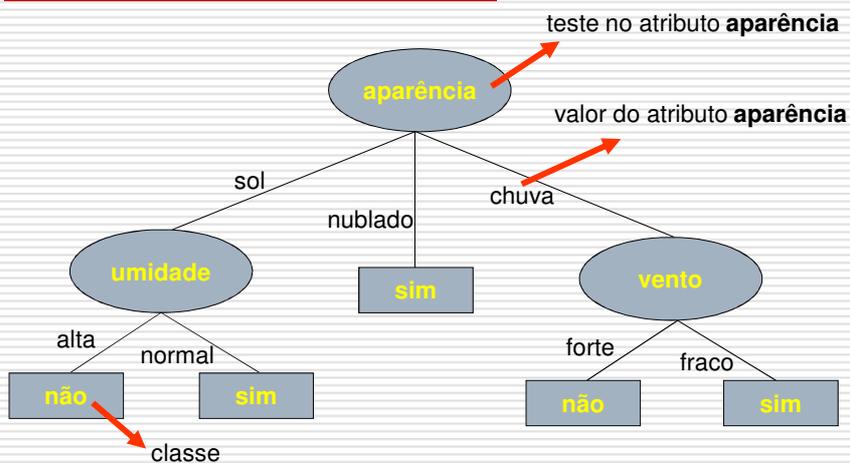
- ❑ Instâncias (exemplos) são representados por pares atributo-valor
- ❑ A hipótese/função objetivo (classe) tem valores discretos, preferencialmente
  - E se não fossem discretos?
- ❑ Descrições disjuntivas podem ser necessárias
- ❑ O conjunto de treinamento pode conter dados problemáticos: valores errados, incompletos ou inconsistentes

## Estrutura

- Uma árvore de decisão contém
  - Nós-folha que correspondem às classes
  - Nós de decisão que contêm testes sobre atributos
    - Para cada resultado de um teste, existe uma aresta para uma subárvore; cada subárvore tem a mesma estrutura que a árvore

7

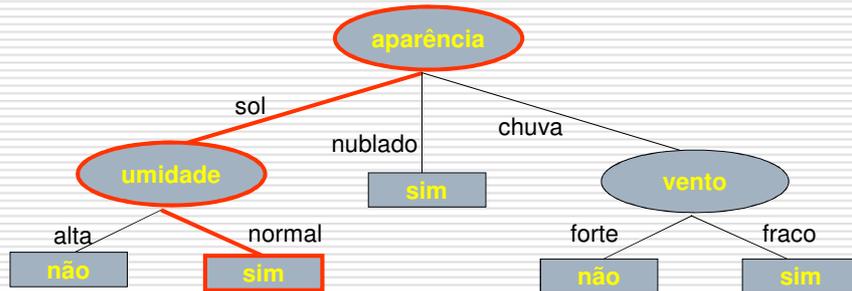
## Exemplo



8

## Exemplo

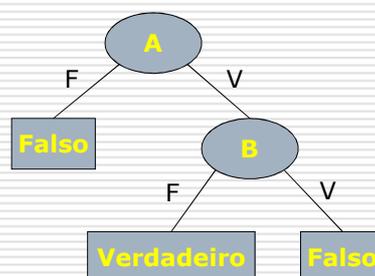
- Para classificar (dizer se vai chover), basta começar pela raiz, seguindo cada teste até que a folha seja alcançada.



9

## Exemplo: AD para função booleana

- $A \wedge \text{not } B$ 
  - Classes: verdadeiro, falso



10

## Exercício: AD para função booleana

---

□  $A \vee [B \wedge C]$

---

11

## Exercício: AD para função booleana

---

□  $A \text{ xor } B$

---

12

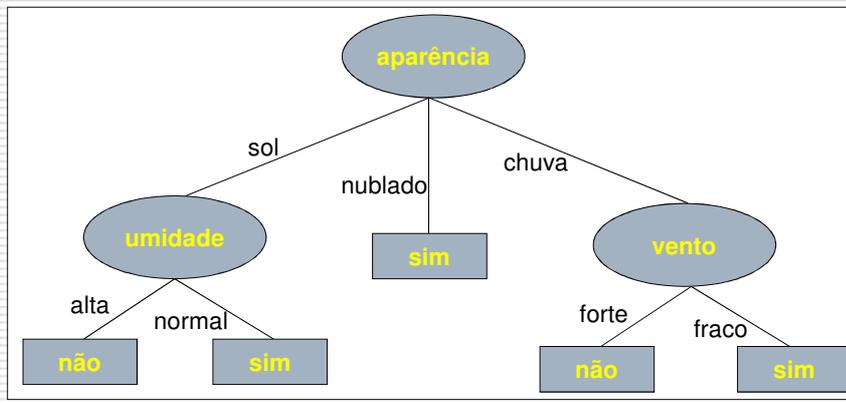
## Exercício: AD para função booleana

□  $[A \wedge B] \vee [C \wedge D]$

13

## Transformação em Regras

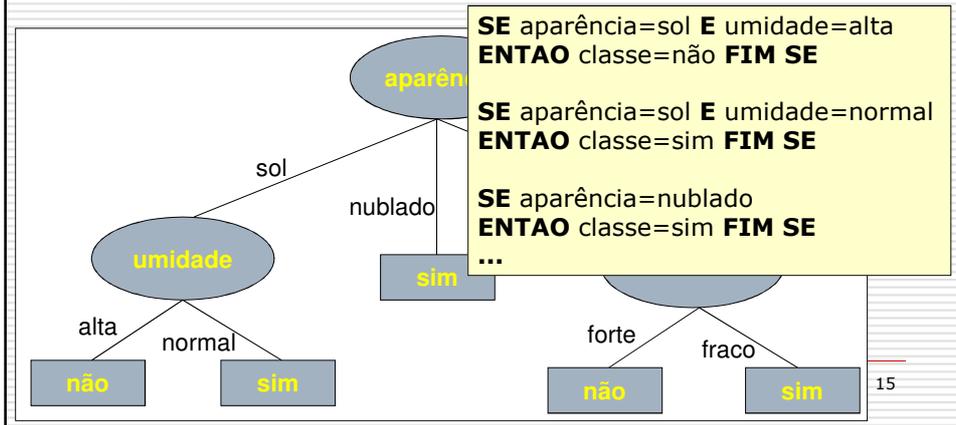
- Toda AD pode ser representada como um conjunto de regras de classificação, sendo que cada regra é gerada percorrendo-se a AD da raiz até as folhas



14

## Transformação em Regras

- Toda AD pode ser representada como um conjunto de regras de classificação, sendo que cada regra é gerada percorrendo-se a AD da raiz até as folhas



## Algoritmos de Aprendizado de AD

- **Muitos algoritmos** foram desenvolvidos para aprendizado de AD
- A maioria utiliza abordagem **top-down** e **busca gulosa** no espaço de possíveis árvores de decisão

## Construindo uma AD

1. O conjunto de treinamento  $T$  contém um ou mais exemplos, todos da mesma classe  $C_j$ : AD para  $T$  é um nó-folha identificando a classe  $C_j$
2.  $T$  não contém exemplos: AD é um nó-folha, mas a classe associada à folha é determinada por outras informações sobre  $T$
3.  $T$  contém exemplos de diversas classes: refinar  $T$  em subconjuntos de exemplos que são (ou possam vir a ser) conjunto de exemplos de uma única classe; a divisão em subconjuntos é feita em função de valores de atributos
4. Aplicar os passos 1, 2 e 3 recursivamente para cada subconjunto de exemplos de treinamento até que o  $i$ -ésimo ramo conduza a uma AD construída sobre o subconjunto  $T_i$  do conjunto de treinamento
5. Depois de construída a AD, utilizar técnicas de poda

## Atributo de Particionamento

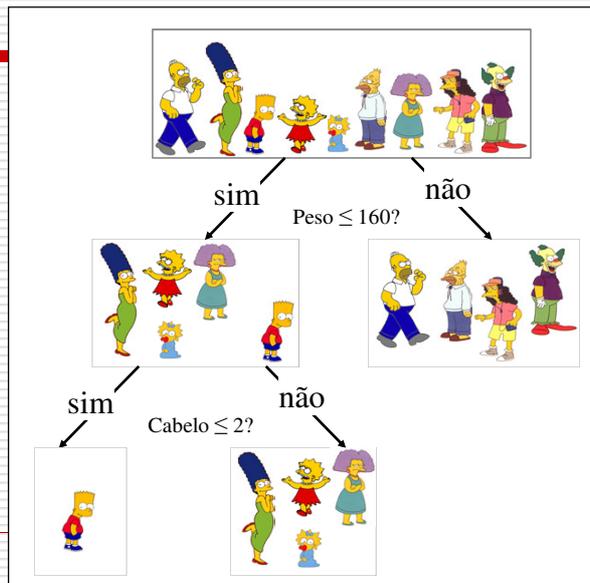
---

- A chave para o sucesso de um algoritmo de aprendizado de AD depende do critério utilizado para escolher o atributo que particiona (atributo de teste) o conjunto de exemplos em cada iteração
- Questão: **como escolher o atributo de particionamento?**

## Exemplo: como foi feito?

	Pessoa	Comprimento do Cabelo	Peso	Idade	Classe: Sexo
	Homer	0	250	36	<b>M</b>
	Marge	10	150	34	<b>F</b>
	Bart	2	90	10	<b>M</b>
	Lisa	6	78	8	<b>F</b>
	Maggie	4	20	1	<b>F</b>
	Abe	1	170	70	<b>M</b>
	Selma	8	160	41	<b>F</b>
	Otto	10	180	38	<b>M</b>
	Krusty	6	200	45	<b>M</b>

## Exemplo: como foi feito?



20

## Atributo de Particionamento

---

- Existem **várias possibilidades** para a **escolha do atributo**
  - **Aleatória:** seleciona um atributo aleatoriamente
  - **Menos valores:** escolhe o atributo com o menor número de valores possíveis
  - **Mais valores:** escolhe o atributo com o maior número de valores possíveis

---

21

## Atributo de Particionamento (cont.)

---

- Existem **várias possibilidades** para a **escolha do atributo**
  - **Ganho Máximo:** seleciona o atributo que possui o maior ganho de informação esperado, isto é, escolhe o atributo que resultará no menor tamanho para as subárvores "enraizadas" em seus filhos
  - **Índice Gini** (Breiman et al, 1984)
  - **Razão de ganho** (Quinlan, 1993)

---

22

## Algoritmo ID3

---

- Um dos mais famosos, proposto por Quinlan
  
  - Características
    - Top-down
    - Busca gulosa no espaço de possíveis AD
    - Medida para selecionar atributos de teste: ganho de informação
- 

23

## ID3

---

- O ID3 começa a construção da árvore com a pergunta
    - Qual atributo deveria ser testado como raiz da árvore ?
  
  - Para saber a resposta, usa a medida estatística **ganho de informação**, a qual mede quão bem um dado atributo separa o conjunto de treinamento de acordo com a classe
  
  - A medida de **ganho de informação** é usada para selecionar o atributo de teste, entre os atributos candidatos, em cada passo do crescimento da árvore
- 

24

## Ganho de Informação

---

- A medida de **ganho de informação** pode ser definida como a **redução esperada na entropia** causada pelo particionamento de exemplos de acordo com um determinado atributo

- **Entropia?**

---

25

## Ganho de Informação

---

- A medida de **ganho de informação** pode ser definida como a **redução esperada na entropia** causada pelo particionamento de exemplos de acordo com um determinado atributo

- **Entropia**: grau de desordem/surpresa de um conjunto de dados

- Quanto menor a entropia, mais previsível e organizado é o conjunto de dados

---

26

## Entropia

---

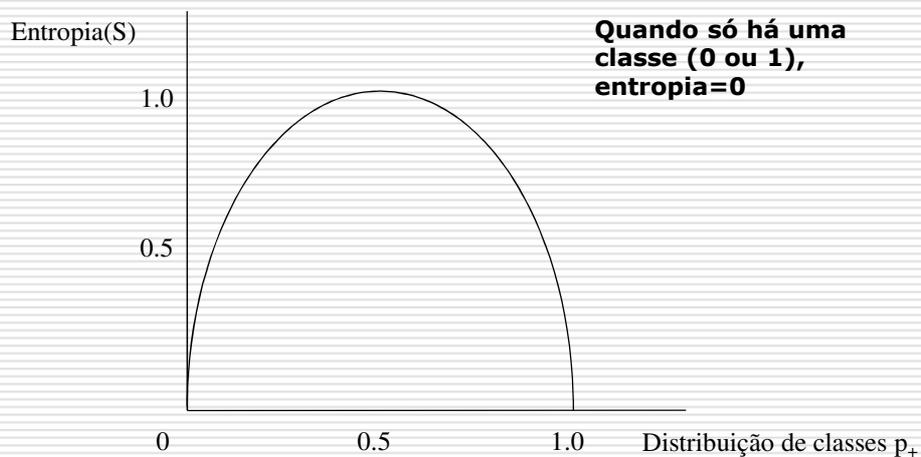
- Original da **Teoria da Informação**, para calcular o número de **bits** necessários para a codificação de uma mensagem
  - Quanto menor a entropia, menos bits são necessários para codificar a mensagem

---

27

## Entropia

---



28

## Entropia

---

- ❑ A entropia é 0 se todos os elementos pertencem a mesma classe
- ❑ A entropia é 1 quando a coleção contém número igual de exemplos positivos e negativos
- ❑ Se a coleção contém número diferente de exemplos positivos e negativos, a entropia varia entre 0 e 1

29

## Entropia

---

- ❑ Genericamente, para qualquer número de classes de um conjunto de dados  $S$ , a **entropia de  $S$**  é dada pela fórmula

$$Entropia(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

em que  $p_i$  é a proporção de instâncias (exemplos) de  $S$  pertencendo a classe  $i$  e  $C$  é o número total de classes

- ❑ Por que esse "menos"? Por que  $\log_2$ ?

30

## Entropia

---

- Fórmula para cálculo da entropia para uma coleção contendo duas classes

$$Entropia(S) = -p_{(+)} \log_2 p_{(+)} - p_{(-)} \log_2 p_{(-)}$$

- Exemplo
  - Dada uma coleção S com 14 exemplos, sendo que a classe é constituída por 9 positivos e 5 negativos, [9+,5-], a entropia de S é
  - Entropia(S) =  $-(9/14)\log_2 (9/14) - (5/14)\log_2 (5/14) = 0.940$

---

31

## Ganho de Informação

---

- O **ganho de informação** deve ser calculado para **cada atributo** do conjunto de atributos da coleção S
- O **atributo** que resultar no **maior ganho de informação** é selecionado como **atributo de teste**

---

32

## Ganho de Informação

- O **ganho de informação** de um atributo A relativo à coleção S é definido como

$$Ganho(S, A) \equiv Entropia(S) - \sum_{v \in \text{Valores}(A)} \frac{|S_v|}{|S|} Entropia(S_v)$$

em que **Valores(A)** é o conjunto de todos os possíveis valores de A, **S<sub>v</sub>** é o subconjunto de S para o qual o atributo A tem valor **v**, o primeiro termo é a entropia da coleção S e o segundo termo é a soma das entropias de cada valor presente no atributo A

33

## Ganho de Informação

- Exemplo

- Atributo **vento**, com valores **forte** e **fraco**
- Coleção S com 14 exemplos, sendo que 9 são positivos e 5 são negativos ([9+,5-])
- Desses exemplos, suponha que 6 dos exemplos positivos e dois exemplos negativos tem **vento=fraco** [6+,2-] e o resto tem **vento=forte** ([3+,3-])
- Portanto
  - Valores(Vento) = fraco, forte
  - Distribuição de S = [9+,5-]
  - Distribuição de Sfraco = [6+,2-]
  - Distribuição de Sforte = [3+,3-]

34

## Ganho de Informação

$$Ganho(S, Vento) \equiv Entropia(S) - \sum_{v \in \{fraco, forte\}} \frac{|S_v|}{|S|} Entropia(S_v)$$

$$Ganho(S, Vento) \equiv Entropia(S) - \left(\frac{8}{14}\right) Entropia(S_{fraco}) - \left(\frac{6}{14}\right) Entropia(S_{forte})$$

$$Ganho(S, Vento) \equiv 0.940 - \left(\frac{8}{14}\right) 0.811 - \left(\frac{6}{14}\right) 1.00$$

$$Ganho(S, Vento) \equiv 0.048$$

35

## Ganho de Informação

$$Ganho(S, Vento) \equiv Entropia(S) - \sum_{v \in \{fraco, forte\}} \frac{|S_v|}{|S|} Entropia(S_v)$$

$$Ganho(S, Vento) \equiv Entropia(S) - \left(\frac{8}{14}\right) Entropia(S_{fraco}) - \left(\frac{6}{14}\right) Entropia(S_{forte})$$

$$Ganho(S, Vento) \equiv 0.940 - \left(\frac{8}{14}\right) 0.811 - \left(\frac{6}{14}\right) 1.00$$

$$Ganho(S, Vento) \equiv 0.048$$

$$-6/8 \log_2 6/8 - 2/8 \log_2 2/8$$

$$-3/6 \log_2 3/6 - 3/6 \log_2 3/6$$

## Exercício

Qual é a entropia desse conjunto de treinamento?

Qual o **ganho de informação** do atributo A2 em relação a esse conjunto de treinamento?

Instância	A1	A2	Classe
1	T	T	+
2	T	T	+
3	T	F	-
4	F	F	+
5	F	F	-
6	F	T	-

Lembre-se:

$$Entropia(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad Ganho(S, A) \equiv Entropia(S) - \sum_{v \in \text{Valores}(A)} \frac{|S_v|}{|S|} Entropia(S_v)$$

## Exercício

Dados

- 2 classes: 3+ e 3-

Entropia =

$$- 3/6 \log_2 3/6 - 3/6 \log_2 3/6 =$$

$$-1/2 (-1) - 1/2 (-1) =$$

1

→ Esse resultado é surpresa?

Instância	A1	A2	Classe
1	T	T	+
2	T	T	+
3	T	F	-
4	F	F	+
5	F	F	-
6	F	T	-

## Exercício

Dados

- Atributo A2: valores T e F
  - T: 2+ e 1-
  - F: 1+ e 2-

Ganho(S,A2) =

Entropia(S) -  $\sum_{v \in \{T,F\}} |S_v|/|S| \text{ Entropia}(S_v) =$

$1 - 3/6 \text{ Entropia}(S_T) - 3/6 \text{ Entropia}(S_F) =$

$1 - 3/6 (-2/3 \log_2 2/3 - 1/3 \log_2 1/3) - 3/6 (-1/3 \log_2 1/3 - 2/3 \log_2 2/3) \approx$

$1 - 0.45 - 0.45 \approx$

0.1

Instância	A1	A2	Classe
1	T	T	+
2	T	T	+
3	T	F	-
4	F	F	+
5	F	F	-
6	F	T	-

39

## Construindo uma AD

	dia	aparência	temperatura	umidade	vento	jogar tênis
Conjunto de treinamento	D1	ensolarado	quente	alta	fraco	não
	D2	ensolarado	quente	alta	forte	não
	D3	nublado	quente	alta	fraco	sim
	D4	chuva	moderada	alta	fraco	sim
	D5	chuva	fria	normal	fraco	sim
	D6	chuva	fria	normal	forte	não
	D7	nublado	fria	normal	forte	sim
	D8	ensolarado	moderada	alta	fraco	não
	D9	ensolarado	fria	normal	fraco	sim
	D10	chuva	moderada	normal	fraco	sim
	D11	ensolarado	moderada	normal	forte	sim
	D12	nublado	moderada	alta	forte	sim
	D13	nublado	quente	normal	fraco	sim
	D14	chuva	moderada	alta	forte	não

## Construindo uma AD

---

- Qual deverá ser o **nó raiz da árvore** ?
  - O nó raiz é identificado calculando o **ganho de informação** para cada atributo do conjunto de treinamento (menos o atributo classe)

---

41

## Construindo uma AD

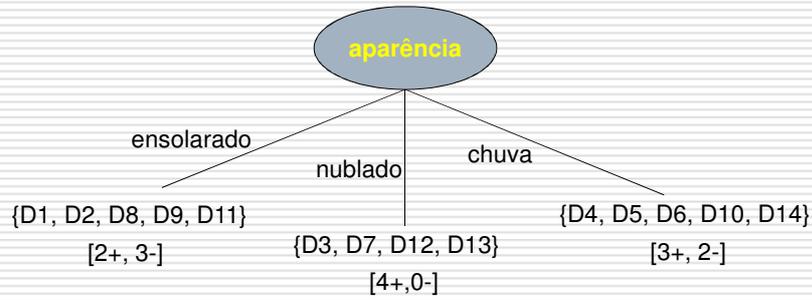
---

- Para cada atributo
  - $\text{Ganho}(S, \text{aparência}) = 0.246$
  - $\text{Ganho}(S, \text{umidade}) = 0.151$
  - $\text{Ganho}(S, \text{vento}) = 0.048$
  - $\text{Ganho}(S, \text{temperatura}) = 0.029$
  
- De acordo com a medida de **ganho de informação**, o atributo **aparência** é o que melhor prediz o atributo classe, **jogar\_tênis**, sobre o conjunto de treinamento

---

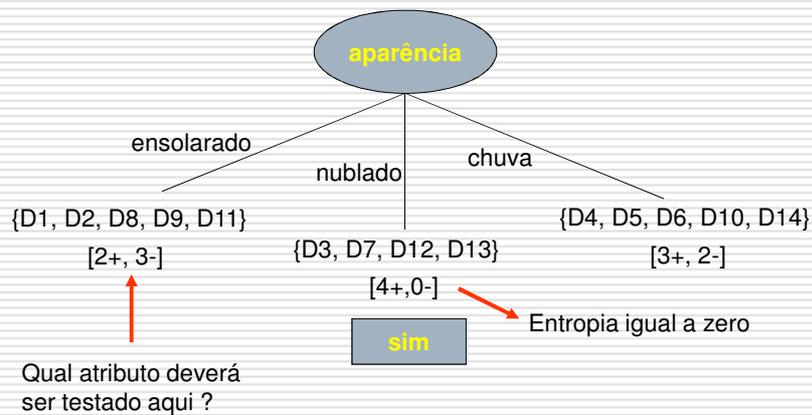
42

## Construindo uma AD



43

## Construindo uma AD



44

## Construindo uma AD

---

- Todo exemplo que tem o atributo **aparência=nublado** tem o atributo classe **jogar\_tênis=sim**; portanto esse nó da árvore se torna folha com a classificação **jogar\_tênis=sim**
  
- Entretanto, os ramos correspondentes a **aparência=ensolarado** e **aparência=chuva** ainda tem entropias diferentes de zero; portanto deverá ser criada uma nova árvore abaixo desses ramos

---

45

## Construindo uma AD

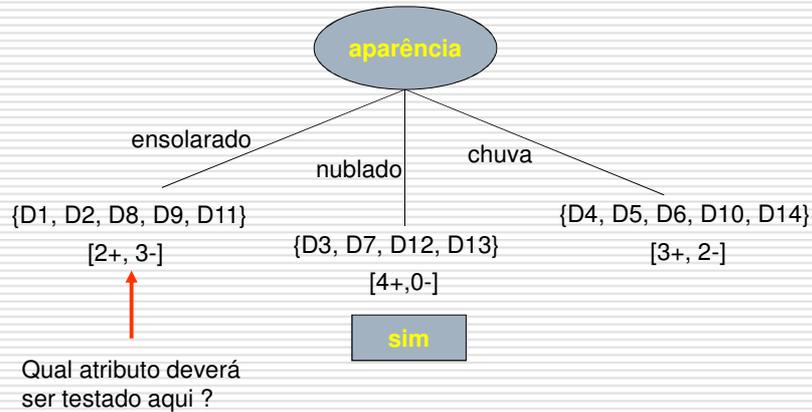
---

- O processo de selecionar um novo atributo e particionar o conjunto de treinamento é repetido para cada ramo, agora usando somente o conjunto de exemplos associado ao ramo
  - Se valores são nominais, atributo é usado uma única vez
  - Se valores podem ser subconjuntos/intervalos, atributo pode ser usado mais de uma vez na árvore
  
- O processo continua até que:
  - Todos os atributos já tenham sido usados
  - Os exemplos de treinamento nos nós-folha possuam a mesma classe (entropia igual a zero)

---

46

## Construindo uma AD



47

## Construindo uma AD

	dia	aparência	temperatura	umidade	vento	jogar tênis
Conjunto de treinamento	D1	ensolarado	quente	alta	fraco	não
	D2	ensolarado	quente	alta	forte	não
	D3	nublado	quente	alta	fraco	sim
	D4	chuva	moderada	alta	fraco	sim
	D5	chuva	fria	normal	fraco	sim
	D6	chuva	fria	normal	forte	não
	D7	nublado	fria	normal	forte	sim
	D8	ensolarado	moderada	alta	fraco	não
	D9	ensolarado	fria	normal	fraco	sim
	D10	chuva	moderada	normal	fraco	sim
	D11	ensolarado	moderada	normal	forte	sim
	D12	nublado	moderada	alta	forte	sim
	D13	nublado	quente	normal	fraco	sim
	D14	chuva	moderada	alta	forte	não

## Construindo uma AD

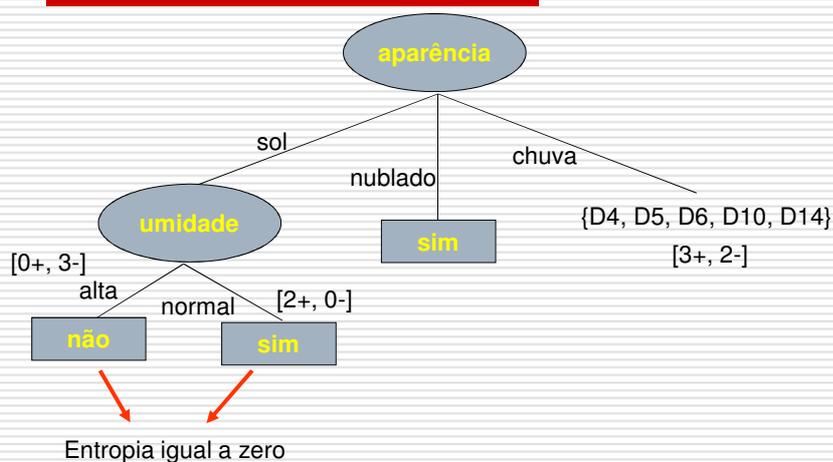
□  $S_{\text{ensolarado}} = \{D1, D2, D8, D9, D11\}$

- $\text{Ganho}(S_{\text{ensolarado}} \text{ umidade}) = 0.970 - (3/5)0.0 - (2/5)0.0 = 0.970$
- $\text{Ganho}(S_{\text{ensolarado}} \text{ temperatura}) = 0.970 - (2/5)1.0 - (2/5)0.0 - (1/5)0.0 = 0.570$
- $\text{Ganho}(S_{\text{ensolarado}} \text{ vento}) = 0.970 - (2/5)1.0 - (3/5)0.918 = 0.019$

□ Nesse caso, o maior **ganho de informação** está no atributo **umidade**

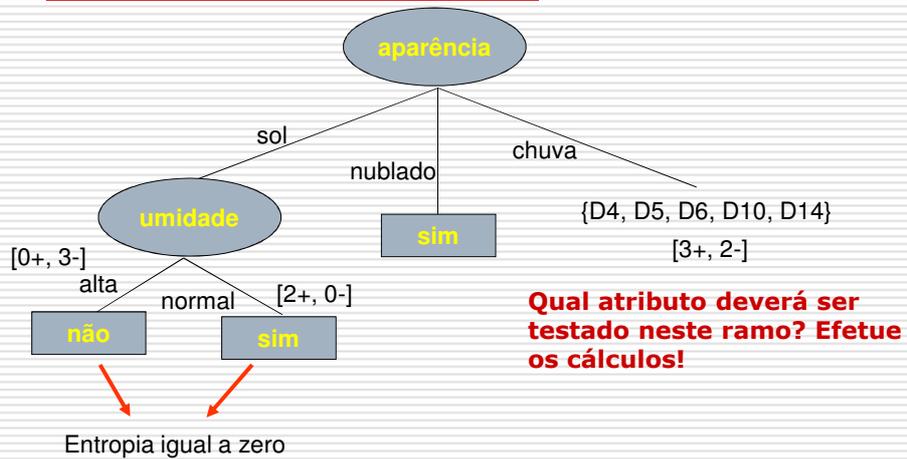
49

## Construindo uma AD



50

## Exercício para casa



51

## ID3

### □ Características

- Prefere **árvores menores**

□ Por quê?

- Atributos com maior ganho de informação mais próximos da raiz da árvore

— ■ Por que a busca é dita gulosa?

52

## O que acontece na AD desse caso?

dia	hora	aparência	temperatura	umidade	vento	jogar ténis
D1	6	ensolarado	quente	alta	fraco	<i>não</i>
D2	9	ensolarado	quente	alta	forte	<i>não</i>
D3	10	nublado	quente	alta	fraco	<i>sim</i>
D4	15	chuva	moderada	alta	fraco	<i>sim</i>
D5	7	chuva	fria	normal	fraco	<i>sim</i>
D6	8	chuva	fria	normal	forte	<i>não</i>
D7	16	nublado	fria	normal	forte	<i>sim</i>
D8	11	ensolarado	moderada	alta	fraco	<i>não</i>
D9	20	ensolarado	fria	normal	fraco	<i>sim</i>
D10	21	chuva	moderada	normal	fraco	<i>sim</i>
D11	13	ensolarado	moderada	normal	forte	<i>sim</i>
D12	12	nublado	moderada	alta	forte	<i>sim</i>
D13	19	nublado	quente	normal	fraco	<i>sim</i>
D14	18	chuva	moderada	alta	forte	<i>não</i>

## Missing values

- Como lidar com **atributos sem valores** (ausentes ou perdidos por alguma razão)?
  - As instâncias com valores ausentes são ignoradas
  - O valor "ausente" é usado como um valor comum, podendo ser um valor testado, inclusive
  - Diante de um valor "ausente", segue-se por um outro ramo, por exemplo, o que for mais popular (ou seja, que tenha mais instâncias associadas a ele)
  - Permitir que a instância siga por vários ramos, sendo que é dado um peso para cada ramo em função do número de instâncias associados a ele; os resultados são recombinaados (com os pesos) para se decidir pela classe

## Poda

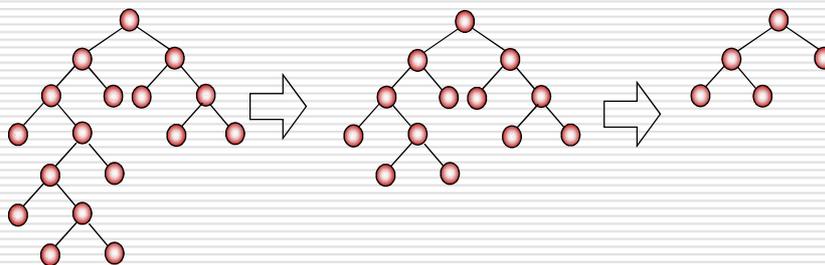
---

- **Maneira de tratar o ruído e o overfitting** em árvores de decisão
    - **Pré-poda:** durante a fase de geração da hipótese alguns exemplos de treinamento são deliberadamente ignorados
    - **Pós-poda:** após a geração da hipótese, esta é generalizada, e algumas partes suas são eliminadas (poda de nós-folha)
      - Mais usada, pois se tem mais informação; no caso anterior, não se sabe onde parar a poda
- 

55

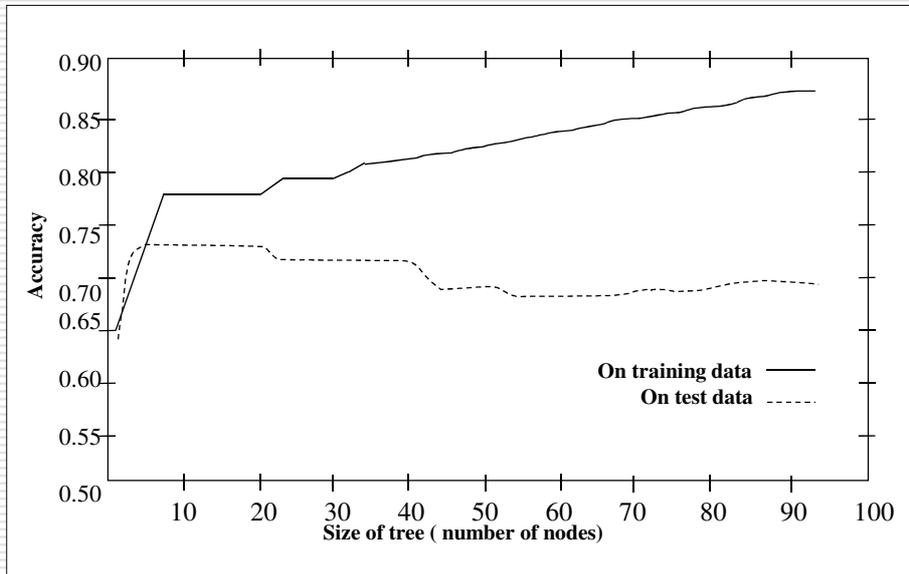
## Exemplo de pós-poda

---

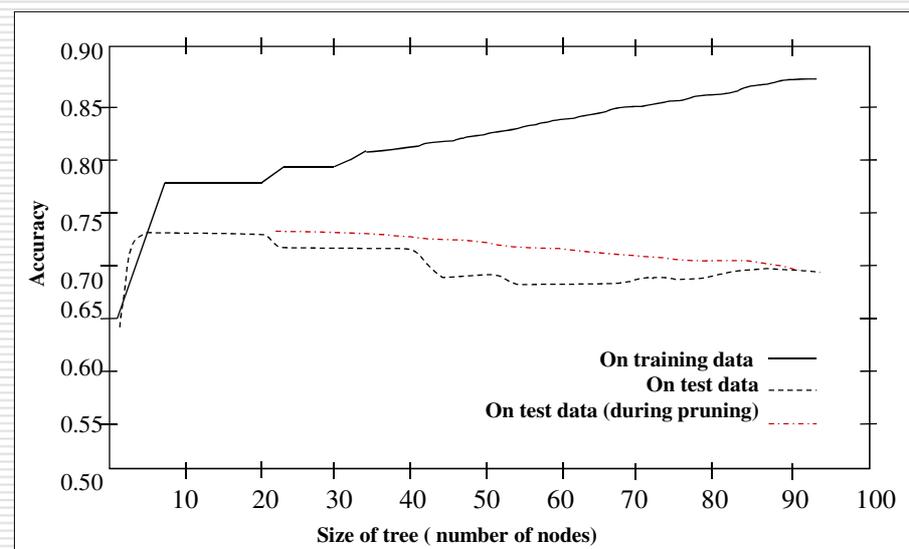


56

## Overfitting



## Overfitting após poda



## Pós-poda

---

- Cada nó da árvore é candidato a corte
- Cortar um nó significa remover a subárvore a partir daquele nó, tornando-o um nó-folha e designando-lhe a classificação mais comum
- Nós são cortados iterativamente, sempre se escolhendo o nó cuja remoção aumenta a precisão da árvore de decisão sobre um conjunto de validação

---

59

## Interpretação Geométrica

---

- Consideramos um exemplo como um vetor de **m** atributos
- Cada vetor corresponde a um ponto em um espaço **m**-dimensional
- A AD corresponde a uma divisão do espaço em regiões, com cada região rotulada por uma classe

---

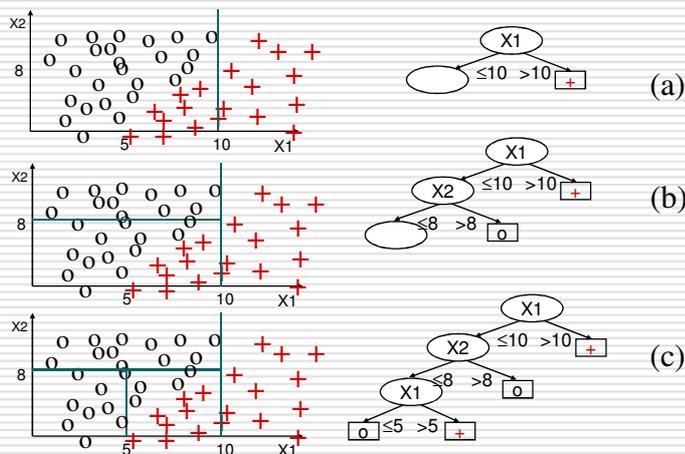
60

## Atributo-Valor

- Um teste para um atributo é da forma  $X_i \text{ op Valor}$ , em que  $X_i$  é um atributo, op um operador e valor é uma constante válida para o atributo
- O **espaço de descrição é particionado** em regiões retangulares, nomeadas hiperplanos, que são ortogonais aos eixos
- As regiões produzidas por uma AD são todas hiper-retângulos
- Enquanto a árvore está sendo formada, mais regiões são adicionadas ao espaço

61

## Interpretação Geométrica



62

## Combinação Linear de Atributos

- **Árvores de decisão oblíquas** (também chamadas multivariadas), com **hiperplanos não ortogonais**

- A representação para os testes são da forma

$$c_1 \times X_1 + c_2 \times X_2 + \dots + c_m \times X_m \text{ op Valor}$$

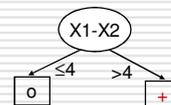
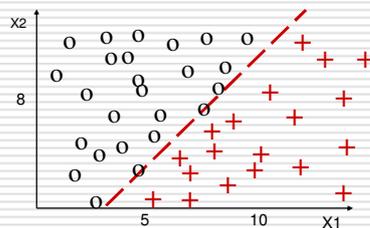
em que  $c_i$  é uma constante,  $X_i$  é um valor contínuo, op um operador e Valor uma constante

- O espaço de descrição é particionado em regiões não retangulares, denominadas hiperplanos que não são necessariamente ortogonais ao eixos

63

## Árvore de Decisão Oblíqua

\* Até agora, tínhamos visto as ADs "univariadas"



64

## Atributos Contínuos

---

- O que fazer com **atributos contínuos**, que não são discretos?

---

65

## Atributos Contínuos: ID3

---

- O que fazer com **atributos contínuos**, que não são discretos?
  - Valores dos atributos são **discretizados**, ou seja, divididos em intervalos
  - Novos **atributos booleanos são criados e testados** em função dos intervalos definidos

---

66

## Atributos Contínuos: ID3

---

### □ Exemplo

Temperature: 40 48 60 72 80 90  
PlayTennis: No No Yes Yes Yes No

### □ Escolhe-se o valor que divide o espaço melhor, de forma que haja o maior ganho de informação

■ Foi mostrado que esse valor fica entre os limites das mudanças de classes (entre 48 e 60; entre 80 e 90)

□  $(48+60)/2 \rightarrow \text{Temperature} > 54$

□  $(80+90)/2 \rightarrow \text{Temperature} > 85$

■  $\text{Temperature} > 54$  causa maior ganho de informação e é escolhido como atributo teste

---

67

## Indutores

---

### □ Softwares disponíveis na web

■ C4.5 (free)

■ See5.0 (limitado, para MS. Windows)

■ **Weka** (free)

---

68

## Árvores & tarefas reais

---

- Como modelar uma tarefa como um problema de AM?
    - Supondo que escolheu a representação em AD
      - Quais são os atributos? Qual o custo de conseguí-los?
      - Quais são as classes?
    - **Exercício em grupos de 2 alunos!**
- 

69

## Exercício

---

- Tarefa contratada por uma empresa
  - **Redução de sentenças**: aplicação em legendas, celulares, simplificação de textos, preenchimento de formulários, etc.

**Sentença original:** *O assessor de Relações Institucionais da Transierra, Hugo Muñoz, afirmou nesta quinta-feira que o envio de gás natural da Bolívia para o Brasil está quase totalmente recuperado.*

**Sentenças reduzida:** *O assessor afirmou que o envio de gás natural está quase recuperado.*

## Exercício

---

- **Modele a tarefa** como um problema de aprendizado de máquina

**Sentença original:** *O assessor de Relações Institucionais da Transierra, Hugo Muñoz, afirmou nesta quinta-feira que o envio de gás natural da Bolívia para o Brasil está quase totalmente recuperado.*

**Sentenças reduzida:** *O assessor afirmou que o envio de gás natural está quase recuperado.*

## Exercício: possível início de solução

---

- Modelagem da tarefa
  - A unidade de informação sobre a qual se tomará alguma decisão é a palavra
  - Cada palavra deve ser classificada como "fica" ou "cai fora"
  - Conjunto de atributos: classe gramatical da palavra, classes gramaticais das 2 palavras anteriores e das duas posteriores, função sintática à qual a palavra está associada, posição da palavra na sentença

## Exercício: possível início de solução

---

### Modelagem da tarefa

- A unidade de informação sobre a qual se tomará alguma decisão é a palavra
- Cada palavra deve ser classificada como "fica" ou "cai fora"
- Conjunto de atributos: classe gramatical da palavra, classes gramaticais das 2 palavras anteriores e das duas posteriores, função sintática à qual a palavra está associada, posição da palavra na sentença

**Desenvolva  
/instancie a  
idéia para a  
empresa**

## Exercício: possível início de solução

---

### Modelagem da tarefa

- A unidade de informação sobre a qual se tomará alguma decisão é a palavra

### Implicações

- Detecção de palavras
- O que é uma palavra? Pontuação conta?

## Exercício: possível início de solução

---

- Modelagem da tarefa
  - Cada palavra deve ser classificada como “fica” ou “cai fora”
  - Problema de 2 classes

---

75

## Exercício: possível início de solução

---

- Modelagem da tarefa
  - Conjunto de atributos: classe gramatical da palavra, classes gramaticais das 2 palavras anteriores e das duas posteriores, função sintática à qual a palavra está associada, posição da palavra na sentença
  - Como conseguir os dados para treino e teste?
  - Vale a pena pedir para humanos anotarem textos?
  - Qual a classe associada a cada instância? De onde ela vem?

---

76

## Exercício: uma solução real

---

- Extração de regras a partir da AD criada
    - Humanos indicaram em textos as partes das sentenças que podiam ser excluídas
      - Classe das palavras vêm junto
    - Ferramentas automáticas para anotação dos textos (classes de palavras, funções sintáticas, contagem de frequência, etc.)
    - Uso do Weka
- 

77

## Exercício: uma solução real

---

- Exemplos de regras obtidas

SE a palavra está no meio da sentença E é parte de um adjunto adverbial ENTÃO cai fora

SE a palavra está no começo da sentença E é um pronome ENTÃO fica

---

78

## Exercício: uma solução real

- Matriz de confusão para as sentenças de um texto de teste
  - Acurácia? Taxa de acerto por classe?

<i>Prevista</i>	<b>Fica</b>	<b>Cai fora</b>
<i>Real</i>		
<b>Fica</b>	342	82
<b>Cai fora</b>	3	35

79

## Exercício: uma solução real

- Matriz de confusão para as sentenças de um texto de teste
  - Acurácia? Taxa de acerto por classe?

<i>Prevista</i>	<b>Fica</b>	<b>Cai fora</b>
<i>Real</i>		
<b>Fica</b>	342	82
<b>Cai fora</b>	3	35

Acurácia =  $377/462 = 81,6\%$  → Taxa de erro geral = 18,4%

Taxa de acerto "fica" (+) =  $342/424 = 80,6\%$

Taxa de acerto "cai fora" (-) =  $35/38 = 92,1\%$

80