

Árvores de decisão

- Um dos métodos de aprendizado de máquina mais utilizados
- Método para aproximação de funções discretas que são robustas na presença de ruídos e capazes de expressar aprendizado disjuntivo
- Algoritmos mais estudados: ID3, ASSISTANT, C4.5, CART

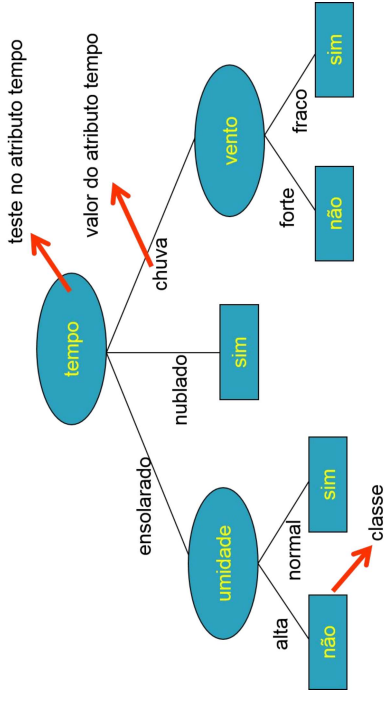
Árvores de decisão

- **Representação**
 - Classificação de instâncias percorrendo os nós-folha a partir da raiz da árvore
 - Cada nó da árvore especifica um teste de algum atributo da instância
 - Cada aresta a partir deste nó corresponde a um possível valor para o atributo

Exemplo

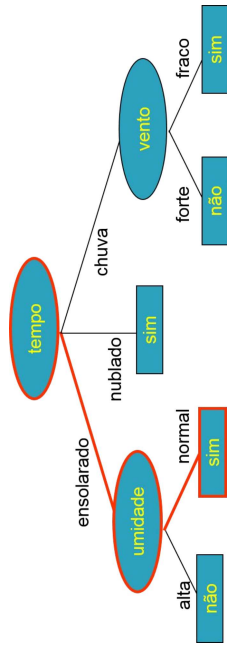
dia	tempo	temperatura	umidade	vento	jogar tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	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

Exemplo



Exemplo

- Para classificar (dizer se é possível jogar tênis), basta começar pela raiz, seguindo cada teste até que a folha seja alcançada.



Árvores de decisão

- Uma árvore de decisão corresponde à uma disjunção de conjuntos de restrições nos valores dos atributos
- Cada caminho corresponde a uma conjunção

(Tempo=ensolarado \wedge umidade = normal)

\vee (Tempo = nublado)

\vee (Tempo = chuva \wedge vento = fraco)

Árvores de decisão

- Quando utilizar árvores de decisão?

1. Instâncias representadas por pares atributo-valor

Ex: Temperatura = quente

- A situação mais fácil é quando tem-se um conjunto pequeno de valores para os atributos
- Extensões dos algoritmos básicos permitem utilizar valores reais para os atributos. Ex: Temperatura = 25

Árvores de decisão

- Quando utilizar árvores de decisão?

2. O problema apresenta descrições disjuntas

Ex: (Tempo = ensolarado \wedge umidade = normal)
 \vee (Tempo = nublado)

Árvores de decisão

- Quando utilizar árvores de decisão?

3. Exemplos com erros/ruídos e/ou atributos com valores ausentes

Ex: A umidade pode ser um fator conhecido somente para alguns exemplos

Algoritmo de aprendizado (AD)

- A maioria utiliza abordagem top-down e busca gulosa no espaço de possíveis árvores de decisão

- Começando a construir a árvore: qual atributo deve ser testado na raiz da árvore?

Algoritmo de aprendizado (AD)

- Cada atributo é individualmente testado para verificar quão bem ele consegue classificar sozinho o conjunto de exemplos

- O melhor atributo é escolhido como a raiz da árvore

Algoritmo ID3

- Cálculo do ganho de informação (*information gain*) de cada atributo
 - Medição de quão bem um dado atributo separa o conjunto de exemplos de acordo com o objetivo da classificação
 - Medida utilizada para selecionar entre os melhores atributos candidatos à medida que a árvore cresce

Algoritmo ID3

- *Information gain*
 - Calcular a entropia do atributo
 - Entropia: caracteriza a pureza ou homogeneidade do conjunto de exemplos
 - Quanto menor a entropia, mais previsível e organizado é o conjunto de dados

Algoritmo ID3

- *Entropia*
 - Propriedade da Termodinâmica usada para determinar a quantidade de energia útil de um sistema
 - Teoria da Informação para calcular o número de bits necessários para codificar uma mensagem
 - Quanto menor a entropia, menos bits são necessários para codificar a mensagem

Algoritmo ID3

- *Valores obtidos pelo cálculo da entropia:*
 - A entropia é 0 se todos elementos pertencem à mesma classe
 - A entropia é 1 quando a coleção contém número igual de exemplos da mesma classe (balanceamento)
 - Se a coleção contém número diferente de exemplos da mesma classe, a entropia varia entre 0 e 1

Algoritmo ID3

- Como calcular a entropia para um conjunto booleano (duas classes)

- Seja um conjunto de exemplo S
- Proporção de exemplos positivos $p_{(+)}$
- Proporção de exemplos negativos $p_{(-)}$
- A entropia dessa classificação booleana é:

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

Algoritmo ID3 - Exercício

- Calcule a entropia desse conjunto de exemplos de treinamento.

dia	tempo	temperatura	umidade	vento	jogar_tênis
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

Algoritmo ID3

- Como calcular o 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
 - O ganho na informação refere-se à redução esperada na entropia à medida que os exemplos são particionados

Algoritmo ID3

- Como calcular o ganho de informação de um atributo

- Seja um conjunto de exemplo S
- Seja o atributo A
- Seja $Valores(A)$ o conjunto de todos os valores possíveis de A
- Seja S_v o subconjunto de S para o qual o atributo A tem valor v

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

Algoritmo ID3

- Como calcular o ganho de informação de um atributo

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

Entropia da coleção S

Soma das entropias de cada valor presente no atributo A

Algoritmo ID3

- Como calcular o ganho de informação de um atributo

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

Valor esperado da entropia depois que o conjunto S é particionado usando A

Entropia da coleção S

Soma das entropias de cada valor presente no atributo A

Algoritmo ID3

- Como calcular o ganho de informação de um atributo

- Ex: Considere a tabela atributo-valor apresentada no slide 24. Qual o ganho de informação do atributo vento?

- Atributo vento, com valores forte e fraco
- Coleção S com 14 exemplos
- 9 exemplos são da classe sim (positivos);
- 5 são da classe não (negativos)
- ([9+, 5-])

Algoritmo ID3

- Como calcular o ganho de informação de um atributo
 - Ex: Considere a tabela atributo-valor apresentada no slide 24. Qual o ganho de informação do atributo vento? (cont.)

- Suponha que 6 dos exemplos positivos e dois exemplos negativos tem vento=fraco [6+, 2-] e o resto tem vento=forte ([3+, 3-]). Portanto, para distribuição de S = [9+, 5-] tem-se:
 - Distribuição de Sfraco = [6+, 2-]
 - Distribuição de Sforte = [3+, 3-]

Algoritmo ID3

- Como calcular o ganho de informação de um atributo (cont.)

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

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

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

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

Algoritmo ID3

- Como calcular o ganho de informação de um atributo

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

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

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

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

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

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

Exercício

- Qual é a entropia desse conjunto de treinamento?
- Qual o ganho de informação do atributo A2 em relação a esse conjunto de treinamento?

Lembre-se:

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

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

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

Construindo uma árvore de decisão

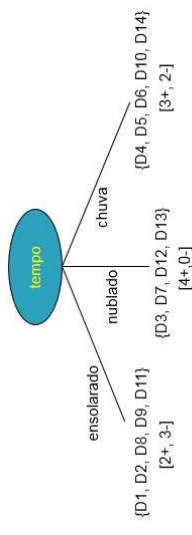
dia	tempo	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	não
D2	ensolarado	quente	alta	forte	não
D3	nublado	quente	alta	fraco	sim
D4	chuva	moderada	alta	fraco	sim
D5	chuva	fria	normal	fraco	sim
D6	chuva	fria	normal	forte	não
D7	nublado	fria	normal	forte	sim
D8	ensolarado	moderada	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 árvore de decisão

1. Qual deverá ser o nó raiz da árvore ?
2. Calcular o ganho de informação de cada atributo do conjunto de treinamento (menos o atributo classe)
 - o Para cada atributo
 - i. $\text{Ganho}(S, \text{tempo}) = 0.246$
 - ii. $\text{Ganho}(S, \text{umidade}) = 0.151$
 - iii. $\text{Ganho}(S, \text{vento}) = 0.048$
 - iv. $\text{Ganho}(S, \text{temperatura}) = 0.029$

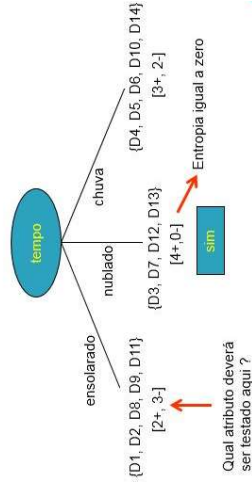
Construindo uma árvore de decisão

- De acordo com a medida de ganho de informação, o atributo tempo é o que melhor prediz o atributo classe, jogar_tênis, sobre o conjunto de treinamento



Construindo uma árvore de decisão

- De acordo com a medida de ganho de informação, o atributo tempo é o que melhor prediz o atributo classe, jogar_tênis, sobre o conjunto de treinamento



Construindo uma árvore de decisão

- Todo exemplo que tem o atributo tempo=ensolarado 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 tempo=ensolarado e tempo=chuva ainda tem entropias diferentes de zero, portanto deverá ser criada uma nova árvore abaixo desses ramos

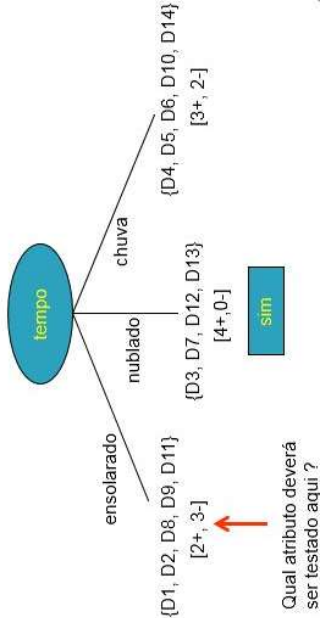
Construindo uma árvore de decisão

- 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

Construindo uma árvore de decisão

- O processo continua para cada novo nó-folha até que duas condições sejam satisfeitas:
 - Todos os atributos já foram incluídos na árvore
 - Os exemplos de treinamento em um nó-folha possuem a mesma classe (entropia igual a zero)

Construindo uma árvore de decisão



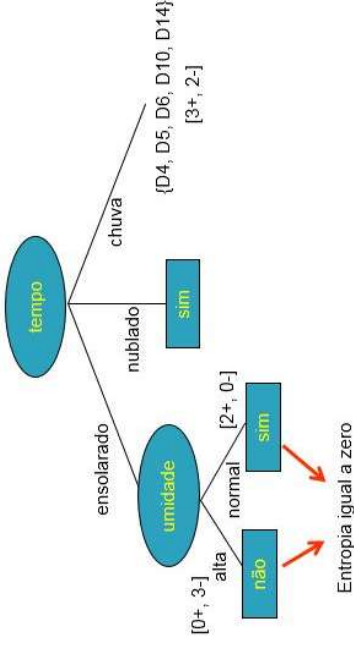
Construindo uma árvore de decisão

dia	tempo	temperatura	umidade	vento	jogar_tênis
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

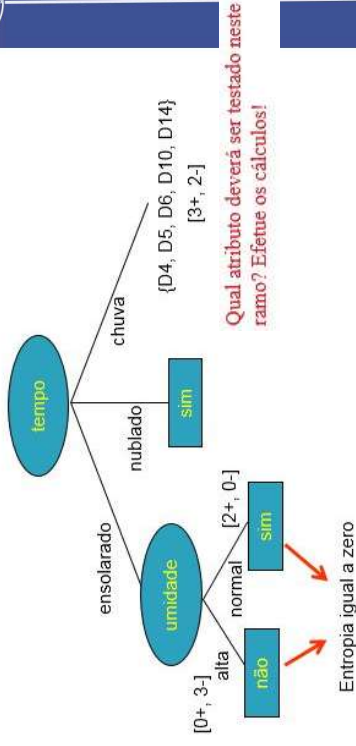
Construindo uma árvore de decisão

- Sensolarado = {D1, D2, D8, D9, D11}
 - $\text{Ganho}(\text{Sensolarado}, \text{umidade}) =$
 - $0.970 - (3/5)0.0 - (2/5)0.0 = 0.970$
 - $\text{Ganho}(\text{Sensolarado}, \text{temperatura}) =$
 - $0.970 - (2/5)1.0 - (2/5)0.0 - (1/5)0.0 = 0.570$
 - $\text{Ganho}(\text{Sensolarado}, \text{vento}) =$
 - $0.970 - (2/5)1.0 - (3/5)0.918 = 0.019$
- Nesse caso, o maior ganho de informação está no atributo umidade

Construindo uma árvore de decisão



Construindo uma árvore de decisão



Referências

- Faceli et al., Inteligência Artificial – Uma Abordagem de Aprendizado de Máquina, LTC, 2015.
- Smola, A. and Vishwanathan, S.V.N., Introduction to Machine Learning, Cambridge University Press, 2008
- Witten et al., Data Mining – Practical Machine Learning Tools and Techniques, 3d edition, Elsevier, 2011.
- J. Han; M. Kamber. Data Mining: Concepts and Techniques, Morgan Kaufmann, 2000