

Universidade Federal de Campina Grande
Departamento de Sistemas e Computação
Pós-Graduação em Ciência da Computação

Inteligência Artificial

Aprendizagem – Parte II **(Árvores de Decisão)**

Prof.^a Joseana Macêdo Fachine Régis de Araújo
joseana@computacao.ufcg.edu.br

<https://www.pinterest.com/carlymundo/decision-tree-infographics/>



Aprendizagem

Tópico

- Aprendizagem
 - Árvores de Decisão



Árvores de Decisão

- ❑ **Árvore de decisão:** formada por um conjunto de nós de decisão, perguntas, que permitem a classificação de cada caso.
- ❑ **Entrada:** um objeto ou situação descrito por um conjunto de propriedades ou atributos.
- ❑ **Saída:** uma decisão.
- ❑ **Indução mediante árvores de decisão** - uma das formas mais simples de algoritmos de aprendizagem.



Árvores de Decisão

- **Aprendizagem de uma árvore de decisão com exemplos:**
 - Situações em que se tem exemplos previamente classificados com regras desconhecidas que se pretendem **aprender** (descobrir).
 - Essas classificações podem ter sido feitas por um especialista que queremos imitar ou traduzir em medidas difíceis ou caras de se obter (por exemplo, medidas realizadas por testes destrutivos).
 - Pode-se “substituir” o especialista por uma árvore de decisão.



Árvores de Decisão

- ❑ **Aprendizagem, Estimação, Treinamento:** Partindo de exemplos, "adivinhar" a árvore.
- ❑ **Árvore de Decisão:** Conjunto de “nós de decisão” e “folhas” que implementam um modelo.
- ❑ A partir de um conjunto de propriedades, decide sim ou não.
- ❑ Representação de árvores de decisão
 - Cada nó interno testa um atributo
 - Cada ramo corresponde a um valor do atributo
 - Cada folha atribui uma classificação



Árvores de Decisão

- ❑ **Exemplo** - descrito pelos valores dos atributos e o valor do **predicado meta**.
- ❑ Valor do predicado meta: **classificação** do exemplo.
- ❑ Predicado meta verdadeiro: **exemplo positivo**, caso contrário **exemplo negativo**.
- ❑ Conjunto completo de exemplos: **conjunto de treinamento**.

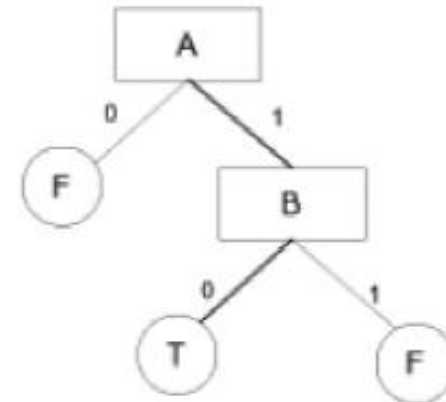


Árvores de Decisão

Exemplo: Qualquer função booleana pode ser escrita como uma árvore de decisão.

$$A \wedge \neg B$$

A	B	$\neg B$	$A \wedge \neg B$
0	0	1	0
0	1	0	0
1	0	1	1
1	1	0	0



$$R : (A = 1 \wedge B = 0)$$

Árvores de Decisão

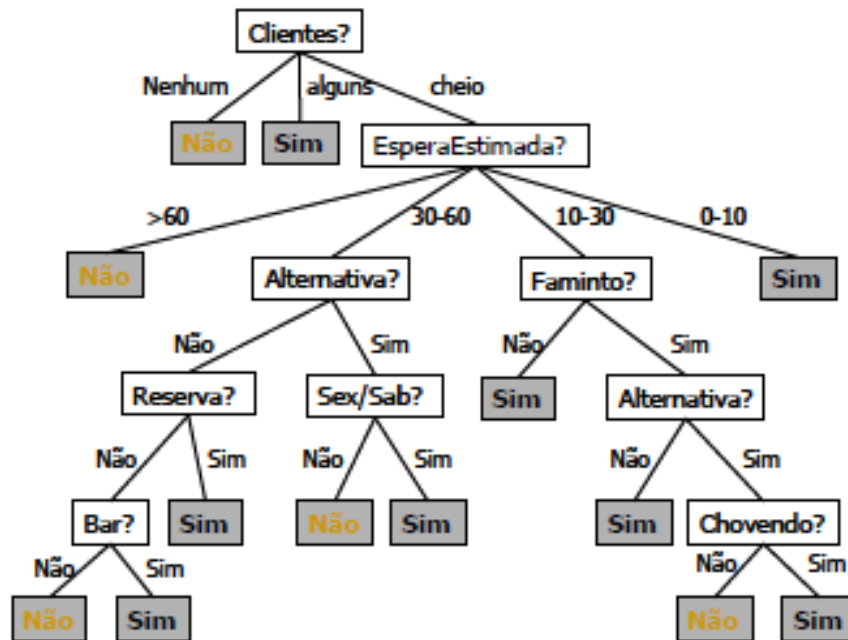
- ❑ Algoritmo de aprendizagem em árvore de decisão (**Ad**):
- ❑ Princípio: testar o **atributo mais importante**
 - aquele que faz a maior diferença para a classificação de um exemplo.
- ❑ Obter a classificação correta com pequeno n° de testes = caminhos curtos e árvore pequena.
- ❑ Árvores de decisão servem para alguns tipos de funções e não são boas para outros.
- ❑ Infelizmente, não existe uma espécie de representação que seja eficiente para todos os tipos de funções.

Limitação: Basicamente, a árvore memoriza as observações. Ela não extrai qualquer padrão dos exemplos e, assim, não podemos esperar que ela esteja apta a extrapolar para exemplos não vistos antes.

Árvores de Decisão

EXEMPLO no domínio do restaurante:

- Objetivo: **aprender** uma definição para o predicado meta **Esperará**, em que a definição é expressa como uma árvore de decisão.



Árvores de Decisão

Exemplo: Esperar ou não uma mesa em um restaurante?

Atributos disponíveis:

- **Alternativa:** há um outro restaurante apropriado por perto?
- **Bar:** o restaurante tem uma área de bar confortável para esperar?
- **Sex/Sab:** verdadeiro às sextas e aos sábados
- **Faminto:** Estamos com fome?
- **Cientes:** Quantas pessoas estão no restaurante?
- **Preço:** A faixa de preços do restaurante
- **Chovendo:** Está chovendo do lado de fora?
- **Reserva:** Fizemos uma reserva?
- **Tipo:** Qual o tipo do restaurante?
- **Espera estimada:** A espera estimada pelo gerente

Não é possível utilizar árvores de decisão para representar testes que se referem a dois ou mais objetos diferentes



Árvores de Decisão

- Qualquer hipótese de árvore de decisão específica para o predicado meta **VaiEsperar** pode ser vista como uma asserção da forma:

$$\forall s \text{ VaiEsperar}(s) \Leftrightarrow (P_1(s) \vee P_2(s) \vee \dots \vee P_n(s))$$

- Cada condição $P_i(s)$ é uma conjunção de testes que pode corresponder a um caminho da raiz até uma folha da árvore com resultado positivo.
- A árvore pode ser representada por uma conjunção de implicações individuais que correspondem aos caminhos que vão da **raiz** até o nó folha **Sim**.

$$\forall r \text{ Clientes}(r, \text{Cheio}) \wedge \text{TempoEsperaEstimado}(r, 10 - 30) \wedge \text{TerFome}(r, N) \\ \Rightarrow \text{Esperará}(r)$$



Árvores de Decisão

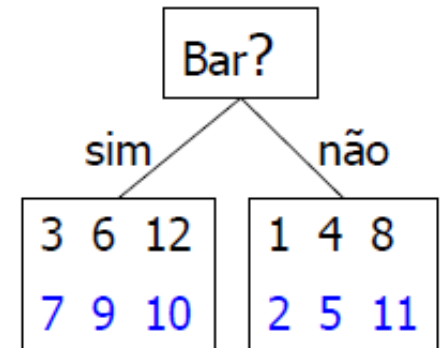
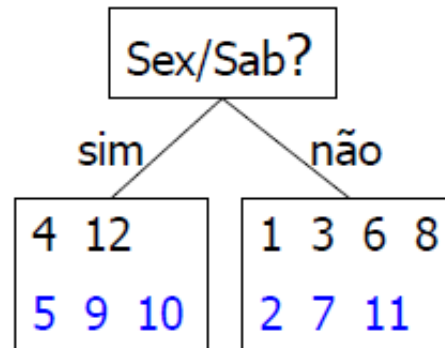
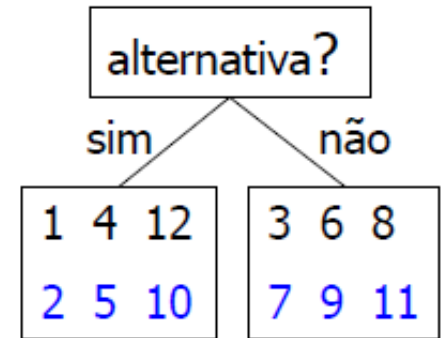
- Exemplo (Restaurante): Conjunto de Treinamento

Ex	Alt	Bar	Sex	Fam	Cli	Pre	Chu	Res	Tipo	Tem	Meta
X ₁	Sim	Não	Não	Sim	Alg	\$\$\$	Não	Sim	Fra	0-10	Sim
X ₂	Sim	Não	Não	Sim	Che	\$	Não	Não	Tai	30-60	Não
X ₃	Não	Sim	Não	Não	Alg	\$	Não	Não	Ham	0-10	Sim
X ₄	Sim	Não	Sim	Sim	Che	\$	Sim	Não	Tai	10-30	Sim
X ₅	Sim	Não	Sim	Não	Che	\$\$\$	Não	Sim	Fra	>60	Não
X ₆	Não	Sim	Não	Sim	Alg	\$\$	Sim	Sim	Ita	0-10	Sim
X ₇	Não	Sim	Não	Não	Ne	\$	Sim	Não	Ham	0-10	Não
X ₈	Não	Não	Não	Sim	Alg	\$\$	Sim	Sim	Tai	0-10	Sim
X ₉	Não	Sim	Sim	Não	Che	\$	Sim	Não	Ham	>60	Não
X ₁₀	Sim	Sim	Sim	Sim	Che	\$\$\$	Não	Sim	Ita	10-30	Não
X ₁₁	Não	Não	Não	Não	Ne	\$	Não	Não	Tai	0-10	Não
X ₁₂	Sim	Sim	Sim	Sim	Che	\$	Não	Não	Ham	30-60	Sim

Árvores de Decisão

Ex	Alt	Bar	Sex
X ₁	Sim	Não	Não
X ₂	Sim	Não	Não
X ₃	Não	Sim	Não
X ₄	Sim	Não	Sim
X ₅	Sim	Não	Sim
X ₆	Não	Sim	Não
X ₇	Não	Sim	Não
X ₈	Não	Não	Não
X ₉	Não	Sim	Sim
X ₁₀	Sim	Sim	Sim
X ₁₁	Não	Não	Não
X ₁₂	Sim	Sim	Sim

Meta	
Sim	1 3 4 6 8 12
Não	2 5 7 9 10 11

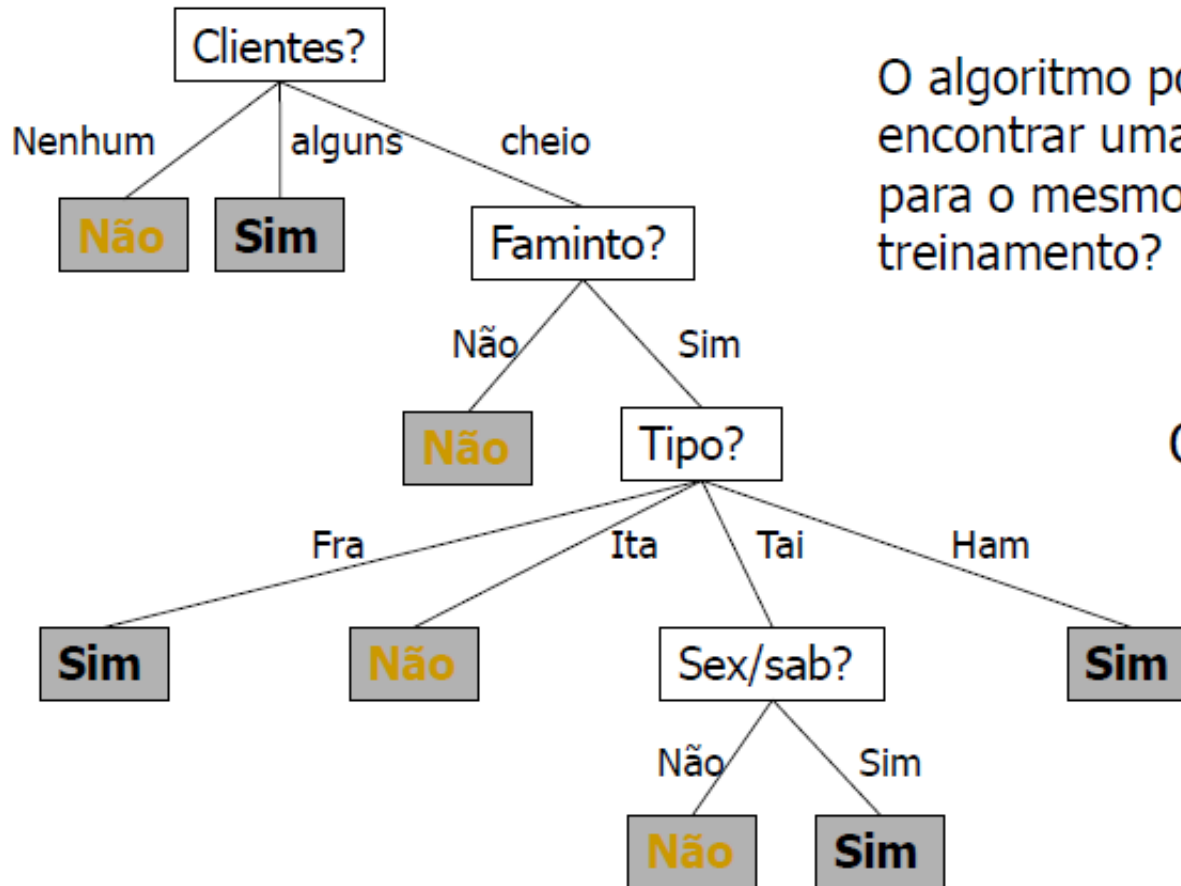


Árvores de Decisão

- ❑ Se existem alguns exemplos positivos e alguns negativos, escolha o melhor atributo para dividi-los.
- ❑ Se todos os exemplos restantes forem positivos (ou negativos), FIM.
- ❑ Se não resta nenhum exemplo, nenhum exemplo deste tipo foi observado, retornar à maioria do nó pai;
- ❑ Não resta nenhum atributo, mas há exemplos positivos e negativos – descrições iguais com classificações diferentes = **ruído** nos dados.



Árvores de Decisão



O algoritmo poderia encontrar uma Ad diferente para o mesmo conjunto de treinamento?

Qual seria?

Árvores de Decisão

Árvore Resultante

- É diferente da árvore original.
- Mas, a hipótese concorda com todos os exemplos.
- É consideravelmente mais simples do que a árvore original.

□ Simplificações

- *Chovendo* e *Reserva* ficaram de fora porque a árvore não necessita destes para classificar os exemplos.
- Mas, nunca viu um caso de espera de 0-10 min e o restaurante cheio.
- Para um caso no qual faminto é falso a **Ad** informa que não devemos esperar.



Árvores de Decisão

Esquema de aprendizagem da Ad

- Projetado para minimizar a profundidade da árvore final
- **Estratégia:** escolher o atributo que melhor fornece uma classificação exata dos exemplos
- **Atributo perfeito:** divide os exemplos em conjuntos que são todos positivos ou todos negativos
 - *Cientes* não é perfeito, mas é “bastante bom”.
- **Atributo inútil:** deixa os conjuntos de exemplos com a mesma proporção do conjunto original
 - *Tipo* é um atributo “realmente inútil”.

Árvores de Decisão

- ❑ Algoritmo (Ad): É bom se ele produz hipóteses que classificam (predizem) bem exemplos ainda não vistos.
- ❑ A predição é boa se ela se torna verdadeira
- ❑ Como avaliar a qualidade de uma hipótese?
- ❑ É possível checar sua previsão com uma classificação correta já conhecida - **conjuntos de teste**.



Árvores de Decisão

Avaliação do desempenho

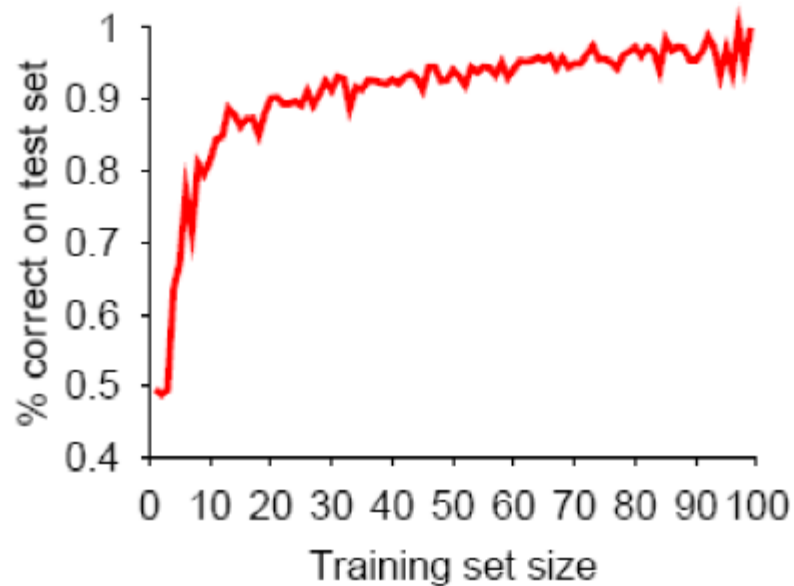
1. Coletar um grande conjunto de exemplos
2. Dividi-lo em dois conjuntos disjuntos: Conjunto de treinamento e conjunto de teste
3. Aplicar o algoritmo ao conjunto de treinamento, gerando uma hipótese h
4. Medir a quantidade de exemplos do conjunto de teste classificados corretamente por h
5. Repetir as etapas 2 a 4 para
 - Diferentes tamanhos de conjuntos de treinamento
 - Diferentes conjuntos de treinamento de cada tamanho



Árvores de Decisão

Exemplo de Análise: Curva de Aprendizagem

Curva de aprendizagem para o algoritmo de árvore de decisão sobre 100 exemplos gerados aleatoriamente no domínio do restaurante. O gráfico resume 20 testes.



Cuidado: Overfitting

Observação: É um bom sinal de que existe realmente algum padrão nos dados e de que o algoritmo de aprendizagem o está incorporando.