



**Inteligência Artificial ao alcance de todos.**

[www.escolalivre-ia.com.br](http://www.escolalivre-ia.com.br)

# Apresentação

- Nome: Diego Santos
- Msc em Engenharia da Computação pela Universidade de Pernambuco
- Cientista de Dados na Reclame Aqui
- Áreas de Pesquisa: Visão Computacional, NLP, PDI

# Agenda

O que é uma árvore de decisão

Formalização do Problema

Exemplo

Problemas

Algoritmos

# Árvore de Decisão

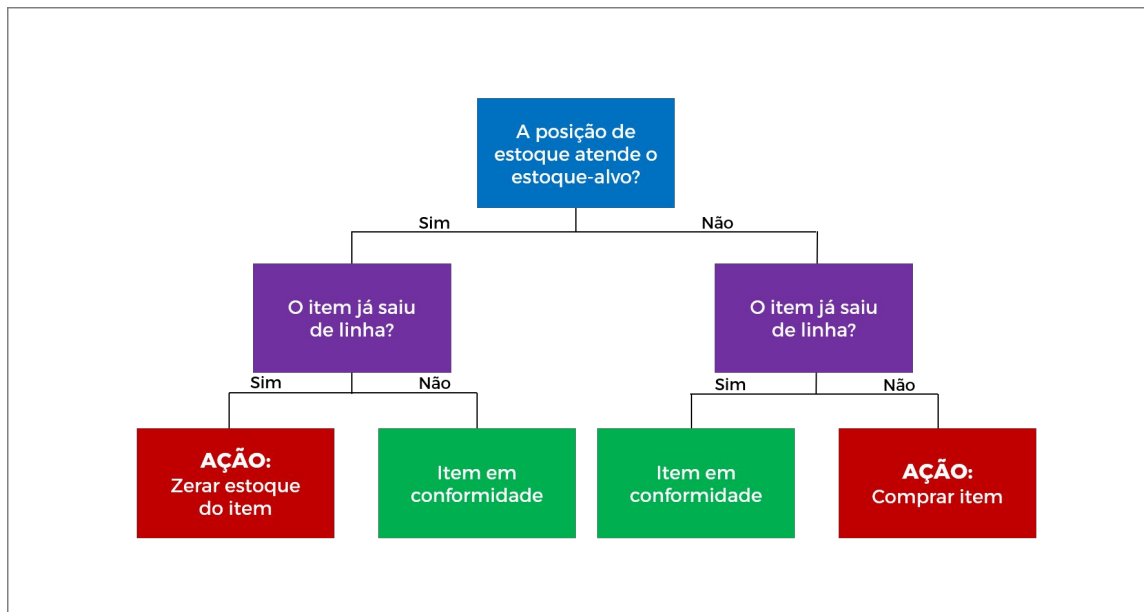


- O que é?
  - Uma árvore de decisão, é uma árvore onde cada nó representa uma escolha entre um número possível de alternativas e cada folha uma decisão;
  - Um tipo de algoritmo de aprendizado supervisionado.

- É um dos métodos mais usados para inferência.



# Exemplo



# Problemas apropriados

- Instâncias são representadas por pares atributo-valor
- A função alvo possui valores de saída discretos
- Os dados do treinamento podem conter erros
- Os dados de treinamento podem ter valores ausentes em algum atributo.

# Formalização do Problema



# Formalização do Problema

Dado um conjunto de treino  $D^{\text{train}}$ , encontre uma árvore

$$g: X \rightarrow Y$$

Tal que para um conjunto de teste  $D^{\text{test}}$  desconhecido durante o treino, o erro de classificação no teste

$$\text{err}(g; D^{\text{test}}) := \frac{1}{|D^{\text{test}}|} \sum_{(x,y) \in D^{\text{test}}} \delta(g(x), y)$$

# Formalização do Problema

- Como  $D^{test}$  é desconhecido, procuramos a árvore que minimize o erro de classificação em  $D^{train}$ .
- Para isso, assume-se que a distribuição de instâncias nas classes do treino é proporcional a distribuição de instâncias nas classes de teste.
- Uma abordagem força bruta é inviável pois o número de árvores no espaço de buscas cresce exponencialmente com o número de atributos.

# Busca Gulosa

Sendo assim, uma busca gulosa é usada de forma que:

- Árvores são construídas a partir da raiz em uma sequência de passos até que a árvore final seja encontrada.
- Em cada passo a escolha deve ser
  1. ótima localmente.
  1. irrevogável.
- Hipótese: uma sequência de seleções ótimas localmente levarão a uma solução ótima global no final.

# Indução de Árvores de Decisão

Sendo assim, uma busca gulosa é usada de forma que:

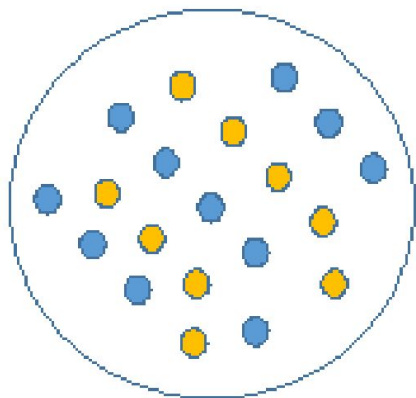
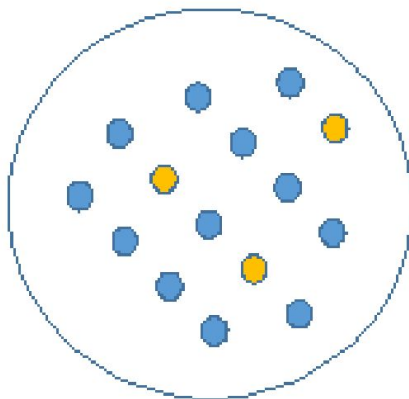
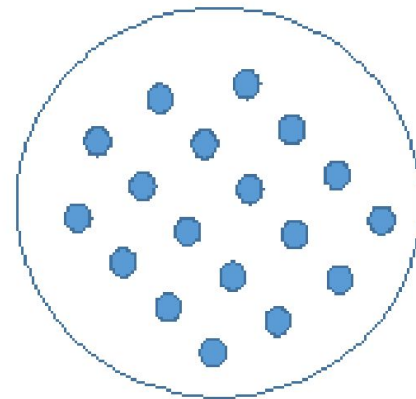
- Árvores são construídas a partir da raiz em uma sequência de passos até que a árvore final seja encontrada.
- Em cada passo a escolha deve ser
  1. ótima localmente.
  2. irrevogável.
- Hipótese: uma sequência de seleções ótimas localmente levarão a uma solução ótima global no final.

# Indução de Árvores de Decisão

- ▶ Ideia: testar os atributos mais importantes primeiro.
- ▶ Atributos importantes tem maior poder de classificação.
- ▶ Condição de parada:
  1. expandir um nó até que (quase) todas as instâncias possuam a mesma classe, ou
  2. nenhum dos atributos apresentem “ganho de informação”.
  3. não existam mais atributos para discriminar as instâncias.
  4. a árvore atingiu uma altura predefinida.

# Seleção de Atributos

- ▶ Ideia: testar os atributos mais importantes primeiro.
- ▶ Qual nó pode ser descrito facilmente?

**A****B****C**

# Então...

- ▶ Nós com menos impureza, precisam de menos informações para serem descritos.
- ▶ Mais impureza, mais informação requer.

A teoria que mede e define o nível de desorganização de um sistema é conhecida como **Entropia**

# Entropia - medir a homogeneidade de um conjunto de aprendizado

- ▶ Entropia é uma medida da incerteza sobre uma fonte de mensagens.
- ▶ Dada uma coleção D, contendo exemplos positivos e negativos de algum alvo, a entropia relativa a D é:

$$E = \sum_{i=1}^D -p_i * \log_2(p_i)$$

Onde,  $p_i$  é a proporção de D pertencente a classe i.



# Entropia

- ▶ Entropia é 0 se todos os membros de  $D$  pertencerem a mesma classe.
- ▶ Entropia é 1 quando o número de exemplos de cada classe é igual.
- ▶ E por conseguinte, entre 0 e 1 se o número de exemplos em cada classe for diferente.

# Ganho de informação

- ▶ Mede a redução da entropia
- ▶ Decide que atributo deve ser usado em cada nó de decisão

Sendo  $D$  nossa coleção:

$$G(D, A) = E(D) - \sum_{v \in \text{values}(A)} \frac{|D_v|}{|D|} \cdot E(D_v)$$

onde:

- $D$  é cada valor  $v$  de todos os possíveis valores de um atributo  $A$
- $D_v$  é o subconjunto de  $D$  para o qual o atributo  $A$  possui o valor  $v$
- $|D_v|$  = número de elementos em  $D_v$

# Entropia

► 14 instancias, 9 yes e 5 não

► Vamos calcular a entropia:

$$E = \sum_{i=1}^D -p_i * \log_2(p_i)$$

$$E = -\left(\frac{9}{14}\right) * \log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) * \log_2\left(\frac{5}{14}\right)$$

$$E = 0,41 + 0,53 = 0.94$$

	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no

# Exemplo

# Qual nó?

Outlook?

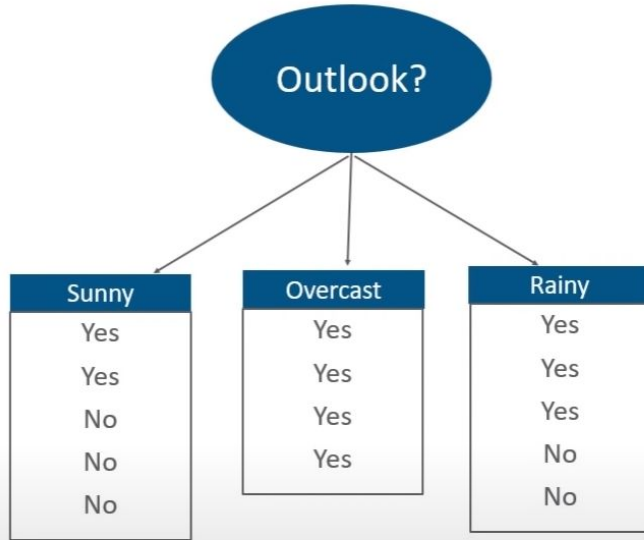
Temperatura?

Humidity?

Windy?

	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no

# Entropia



	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no

# Entropia

$$E(outlook|sunny) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.971$$

$$E(outlook|overcast) = -1 \log_2 1 - 0 \log_2 0 = 0$$

$$E(outlook|rainy) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.971$$

*Ganho de informação*

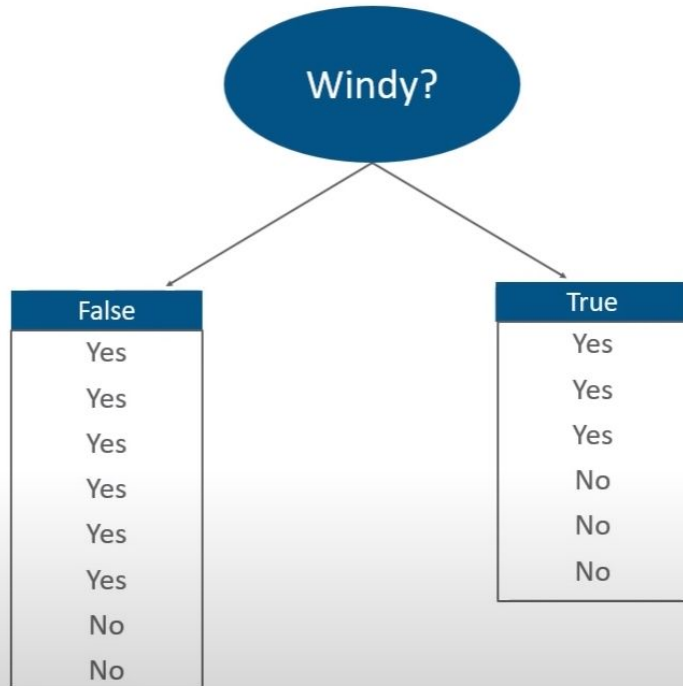
$$I(Outlook) = \frac{5}{14} \times 0.971 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.971 = 0.693$$

$$G(Outlook) = E(D) - I(Outlook)$$

$$0.94 - 0.693 = 0.247$$

	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no

# Entropia



	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no



# Entropia

$$E(\text{windy}|\text{True}) = 1$$

$$E(\text{windy}|\text{False}) = 0.811$$

*Ganho de informação*

$$I(\text{windy}) = \frac{8}{14} \times 0.811 + \frac{6}{14} \times 1 = 0.892$$

$$G(\text{Windy}) = E(D) - I(\text{Windy})$$

$$0.94 - 0.892 = 0.048$$

	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no

# Entropia

$$E(windy|True) = 1$$

$$E(windy|False) = 0.811$$

*Ganho de informação*

$$I(windy) = \frac{8}{14} \times 0.811 + \frac{6}{14} \times 1 = 0.892$$

$$G(Windy) = E(D) - I(Windy)$$

$$0.94 - 0.892 = 0.048$$

	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no

# Entropia

Outlook:  
Ganho: 0.247

Temperature:  
Ganho: 0.029

Humidity:  
Ganho: 0.152

Windy:  
Ganho: 0.048

**Ganho Máximo: 0.247**  
**Nossa Raiz é o Outlook**

	outlook	temp.	humidity	windy	play
D1	sunny	hot	high	false	no
D2	sunny	hot	high	true	no
D3	overcast	hot	high	false	yes
D4	rainy	mild	high	false	yes
D5	rainy	cool	normal	false	yes
D6	rainy	cool	normal	true	no
D7	overcast	cool	normal	true	yes
D8	sunny	mild	high	false	no
D9	sunny	cool	normal	false	yes
D10	rainy	mild	normal	false	yes
D11	sunny	mild	normal	true	yes
D12	overcast	mild	high	true	yes
D13	overcast	hot	normal	false	yes
D14	rainy	mild	high	true	no

# Poda

# Poda

- ▶ Considere cada um dos nós de decisão na árvore como candidatos à remoção.
- ▶ A remoção de um nó de decisão consiste em remover a sub-árvore enraizada nesse nó, transformá-lo em uma folha.
- ▶ Os nós serão removidos apenas se a árvore resultante não apresentar desempenho pior que a original no conjunto de validação
- ▶ A remoção de nós continua até que a remoção adicional seja prejudicial.

# Problemas

# Problemas

- ▶ Quão profunda deve ser a árvore de decisão?
- ▶ Atributos contínuos
- ▶ Manipulando dados de treinamento com valores de atributo ausentes

# Algoritmos



# Algoritmos

Existem vários algoritmos de classificação que utilizam a árvore de decisão. Não se pode determinar qual é a melhor. Dependendo do problema, um algoritmo pode ser mais eficiente que outro.

- ▶ ID3 (Quinlan 1979)
- ▶ CART – Classification and Regression Trees (Breiman et al. 1984)
- ▶ Assistant (Cestnik et al. 1987)
- ▶ C4.5 (Quinlan 1993)
- ▶ C5 (See5, Quinlan) disponível no WEKA, Clementine o CHAID (Chi Square Automatic Interaction Detection)

# Muito obrigado!



Inteligência Artificial ao alcance de todos.

[www.escolalivre-ia.com.br](http://www.escolalivre-ia.com.br)