

Lista 02

Disciplina: Inteligência Artificial

Prof<sup>a</sup>.: Cristiane Neri Nobre

Data: 22/08/2025

Aluna: Alessandra Faria Rodrigues

Matrícula: 828333

### Questão 01

1) Como uma árvore de decisão é gerada? Qual o significado do atributo que está na raiz da árvore?

Ela é gerada de forma recursiva, dividindo o conjunto de dados em partes menores com base nos atributos que melhor separam as classes. O algoritmo avalia todos os atributos disponíveis e escolhe aquele que melhor divide os dados, para cada valor do atributo escolhido, cria-se um ramo que representa uma condição, e assim o conjunto de dados é dividido de acordo com os valores desse atributo, e para cada subconjunto o algoritmo escolhe o melhor atributo seguinte e repete todo esse processo até que não haja mais atributos ou quando todos os nós são folhas.

O atributo que está na raiz da árvore de decisão é o mais importante na primeira divisão dos dados, pois é ele que mais reduz a incerteza ou que melhor separa os exemplos de acordo com o alvo.

2) O que você pode fazer com uma árvore de decisão gerada a partir de uma base de dados?

- a) Fazer previsões (classificação e regressão): Dado um novo registro é possível percorrer a árvore seguindo as condições até chegar em uma folha, que dará a classe prevista (ex.: “vai a aula” ou “não vai a aula”). No caso de regressão, cada folha pode conter um valor médio (ex.: prever nota do aluno com base no tempo de estudo e frequência).
- b) Interpretar o modelo e extrair conhecimento para tomada de decisão: Cada caminho da raiz até uma folha pode ser traduzido em uma regra lógica "SE-ENTÃO". Essas regras são facilmente compreensíveis por humanos e podem ser usadas para entender o que levam a um determinado resultado.
- c) Entendimento dos dados: A árvore mostra quais atributos são mais importantes na classificação (a raiz e os primeiros níveis geralmente tem maior impacto), o que ajuda a descobrir padrões nos dados de forma que seja possível interpretá-los.

Em resumo, com uma árvore de decisão você pode **prever, classificar, gerar regras interpretáveis, identificar padrões** e apoiar a **tomada de decisão**. Além disso, serve também como explicação simples e visual de como os dados se comportam.

### 3) Quais as vantagens e desvantagens de um algoritmo de árvore de decisão?

#### Vantagens:

- **Flexibilidade:** As árvores de decisão são métodos não paramétricos, o que significa que não assumem nenhuma distribuição específica para os dados. Elas fornecem uma cobertura completa do espaço de instâncias.
- **Seleção de Atributos:** O próprio processo de construção da árvore seleciona os atributos a serem utilizados, o que torna os modelos robustos contra atributos irrelevantes e redundantes.
- **Interpretabilidade:** Decisões complexas e globais podem ser aproximadas por uma série de decisões locais mais simples, e todas as decisões se baseiam nos valores dos atributos do problema.
- **Eficiência:** O algoritmo de aprendizado é do tipo guloso e construído de cima para baixo (top-down). Sua complexidade de tempo é linear em relação ao número de exemplos.

#### Desvantagens:

- **Valores Ausentes:** É necessário que os algoritmos utilizem mecanismos específicos para lidar com a falta de valores nos dados.
- **Atributos Contínuos:** A ordenação de atributos contínuos representa um problema. Estima-se que essa operação possa consumir 70% do tempo necessário para gerar uma árvore de decisão em grandes bases de dados com muitos atributos contínuos.
- **Instabilidade:** Pequenas variações nos dados de treinamento podem causar grandes variações na árvore final. Além disso, as inferências feitas próximo às folhas tendem a ser menos confiáveis do que as feitas perto da raiz.

### 4) Como avaliamos a qualidade de uma árvore?

Para avaliar a qualidade da árvore é preciso considerar o tipo de aprendizado. Em problemas de classificação, essa avaliação é feita pela matriz de confusão. Em problemas de agrupamento, avalia utilizando métricas específicas como silhouette index. Em problemas de regressão, utilizam-se as métricas: Mean squared error (MSE), Mean absolute error (MAE), Normalized MSE and MAE. Em problemas de associação, utilizam-se as métricas: suporte, confiança, lift, dentre outras.

5) Como podemos obter as regras a partir de uma árvore de decisão?

Cada caminho completo, da raiz da árvore até um nó folha , representa uma regra lógica no formato "**SE... ENTÃO...**".

Para obter essas regras basta ir ao nó folha (ENTÃO) e traçar o caminho até a raiz, coletando as condições (SE). A partir disso, para construir a regra junte todas as condições coletadas com o operador lógico "E". A conclusão da regra será o resultado ou a classe presente no nó folha que você escolheu no início. Repita esse mesmo processo para todos os nós folhas para obter o conjunto completo de regras que descreve o modelo.

### **Questão 02**

Considerando-se a base de dados sobre "Esperar ou não pelo restaurante" (verificar base de dados 'Restaurante.csv' disponibilizada no CANVAS), pede-se:

1) Calcular o ganho de informação de cada atributo. Que atributo é a raiz da árvore?

Lista 02

Questão 02.1)

$$\text{ganho}(\text{atributo}) = \text{entropia}(\text{classe}) - \text{entropia}(\text{atributo})$$

$$\text{ganho}(\text{cliente}) = \underbrace{\text{entropia}(\text{classe})}_{\text{entropia da classe}} - \text{entropia}(\text{cliente})$$

6 Sim

6 Não

$$\text{Entropia}(s) = \sum -p_i \log_2 p_i$$

$$E(\text{classe}) = E\left(\frac{1}{2}, \frac{1}{2}\right)$$

$$= -\frac{1}{2} \times \log_2 \frac{1}{2} - \frac{1}{2} \times \log_2 \frac{1}{2} = 1$$

entropia da classe

$$1 - \left[ \frac{2}{12} E(0,1) + \frac{4}{12} E(1,0) + \frac{6}{12} E\left(\frac{2}{6}, \frac{4}{6}\right) \right]$$

S N  
Nenhum = 2 0 1  
Alguns = 4 1 0  
Cheio = 6 2 4  
          6 6 6

$$= 1 - \left[ \frac{6}{12} E\left(\frac{2}{6}, \frac{4}{6}\right) \right]$$

$$= -\frac{2}{6} \times \log_2 \frac{2}{6} - \frac{4}{6} \times \log_2 \frac{4}{6} \approx 0,91829$$

$$= 1 - \left[ \frac{6}{12} \times 0,91829 \right]$$

ganho cliente

$$= 1 - [0,45915] = 0,54085$$

$$\text{ganho}(\tau: \text{po}) = \text{entropia}(\text{classe}) - \text{entropia}(\tau: \text{po})$$

$$1 - \left[ \frac{2}{12} \epsilon\left(\frac{1}{3}, \frac{2}{3}\right) + \frac{4}{12} \epsilon\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} \epsilon\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} \epsilon\left(\frac{1}{3}, \frac{2}{3}\right) \right]$$

	S	N
francês	$\frac{1}{2}$	$\frac{1}{2}$
Tailorades	$\frac{2}{3}$	$\frac{1}{3}$
Hamburger	$\frac{1}{3}$	$\frac{2}{3}$
Italiano	$\frac{1}{3}$	$\frac{2}{3}$

$$1 - \left[ \frac{2}{12} + \frac{4}{12} + \frac{4}{12} + \frac{2}{12} \right]$$

$$= 1 - \left[ \frac{12}{12} \right] = 1 - [1] = \boxed{0}$$

↳ ganho do  $\tau: \text{po}$

$$\text{ganho}(\text{tempo}) = \text{entropia}(\text{classe}) - \text{entropia}(\text{tempo})$$

$$1 - \left[ \frac{6}{12} \epsilon\left(\frac{4}{6}, \frac{2}{6}\right) + \frac{2}{12} \epsilon\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} \epsilon\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} \epsilon(0, 1) \right]$$

	S	N
0 - 10	4	2
10 - 30	1	1
30 - 60	1	1
>60	0	2

$$= 1 - \left[ \frac{6}{12} \times \epsilon\left(\frac{4}{6}, \frac{2}{6}\right) + \frac{2}{12} + \frac{2}{12} + 0 \right]$$

$$= 1 - \left[ \frac{6}{12} \times \left( -\frac{4}{6} \times \log_2 \frac{4}{6} - \frac{2}{6} \times \log_2 \frac{2}{6} \right) + \frac{2}{12} + \frac{2}{12} + 0 \right] \approx 0,91829$$

$$= 1 - \left[ \frac{6}{12} \cdot 0,91829 + \frac{2}{12} + \frac{2}{12} \right]$$

$$= 1 - [0,79248] = \boxed{0,20752}$$

↳ ganho do tempo

$$\text{ganho (fome)} = \text{entropia (classe)} - \text{entropia (fome)}$$

S N

Sim → 7 → 5 2

Não → 5 → 1 4

$$1 - \left[ \frac{7}{12} \in \left( \frac{5}{7}, \frac{2}{7} \right) + \frac{5}{12} \in \left( \frac{1}{5}, \frac{4}{5} \right) \right]$$

$$- \frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7} \approx 0,86312$$

$$- \frac{1}{5} \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5} \approx 0,72192$$

$$= 1 - \left[ \frac{7}{12} \times 0,86312 + \frac{5}{12} \times 0,72192 \right]$$

$$= 1 - \left[ \frac{6,0418}{12} + \frac{3,6096}{12} \right]$$

$$= 1 - [0,8042] = \boxed{0,1958} \rightarrow \text{ganho de fome}$$

$$\text{ganho (preço)} = \text{entropia (classe)} - \text{entropia (preço)}$$

S N

R → 7 → 3 4

RR → 2 → 2 0

RRR → 3 → 1 2

$$1 - \left[ \frac{7}{12} \in \left( \frac{3}{7}, \frac{4}{7} \right) + \frac{2}{12} \in (1, 0) + \frac{3}{12} \in \left( \frac{1}{3}, \frac{2}{3} \right) \right]$$

$$\approx 0,98522$$

$$\approx 0,91829$$

$$1 - \left[ \frac{7}{12} \times 0,98522 + 0 + \frac{3}{12} \times 0,91829 \right]$$

$$1 - [0,8042] = \boxed{0,1958} \rightarrow \text{ganho de preço}$$

$$\text{ganho (chuva)} = \text{entropia (classe)} - \text{entropia (chuva)}$$

S N

Sim → 5 → 2 3

Não → 7 → 3 4

$$1 - \left[ \frac{5}{12} \in \left( \frac{2}{5}, \frac{3}{5} \right) + \frac{7}{12} \in \left( \frac{3}{7}, \frac{4}{7} \right) \right]$$

$$\approx 0,97095$$

$$\approx 0,98522$$

$$1 - [0,9792] = \boxed{0,0208} \rightarrow \text{ganho de chuva}$$

$$\text{ganho}(\text{res}) = \text{entropia}(\text{classe}) - \text{entropia}(\text{res})$$

	S	N
Sim	5	3
Não	7	4

$$1 - \left[ \frac{5}{12} \in \left( \frac{3}{5}, \frac{2}{5} \right) + \frac{7}{12} \in \left( \frac{3}{7}, \frac{4}{7} \right) \right]$$

$\approx 0,97095$        $\approx 0,9852$

$$1 - [0,9792] = \boxed{0,0208} \sim \text{ganho de res}$$

$$\text{ganho}(\text{sexSub}) = \text{entropia}(\text{classe}) - \text{entropia}(\text{sexSub})$$

	S	N
Sim	5	2
Não	7	4

$$1 - \left[ \frac{5}{12} \in \left( \frac{2}{5}, \frac{3}{5} \right) + \frac{7}{12} \in \left( \frac{4}{7}, \frac{3}{7} \right) \right]$$

$\approx 0,97095$        $\approx 0,9852$

$$1 - [0,9792] = \boxed{0,0208} \sim \text{ganho de sexSub}$$

$$\text{ganho}(\text{bar}) = \text{entropia}(\text{classe}) - \text{entropia}(\text{bar})$$

	S	N
Sim	6	3
Não	6	3

$$1 - \left[ \frac{6}{12} \in \left( \frac{1}{2}, \frac{1}{2} \right) + \frac{6}{12} \in \left( \frac{1}{2}, \frac{1}{2} \right) \right]$$

$= 1 - \left[ \frac{6}{12} + \frac{6}{12} \right]$

$$\frac{3}{6} = \frac{1}{2}$$

$$= 1 - [1] = \boxed{0} \sim \text{ganho de bar}$$

$$\text{ganho}(\text{alternativo}) = \text{entropia}(\text{classe}) - \text{entropia}(\text{alternativo})$$

$$1 - \left[ \frac{6}{12} \in \left( \frac{1}{2}, \frac{1}{2} \right) + \frac{6}{12} \in \left( \frac{1}{2}, \frac{1}{2} \right) \right] = \boxed{0}$$

$\sim \text{ganho alternativo}$

	S	N
Sim	6	3
Não	6	3

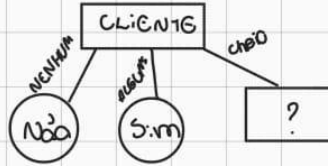
Resposta: A raiz da árvore será "cliente", pois dentre todos os atributos foi o que apresentou maior ganho



2) Que atributo estará no segundo nível da árvore? Faça os cálculos e apresente a árvore gerada até o segundo nível da árvore.

Lista 02

Questão 02 - 2)



entropia (classe)  $\rightarrow$   
 $H = 0,91829$

Sim  $\rightarrow 2$   
 Não  $\rightarrow 4$

ganho (tempo) =

$$0,9 - \left[ \frac{2}{6} \in \left( \frac{1}{2}, \frac{1}{2} \right) + \frac{2}{6} \in \left( \frac{1}{2}, \frac{1}{2} \right) + \frac{2}{6} \in (0,1) \right]$$

$$0,9 - \left[ \frac{2}{6} + \frac{2}{6} \right] = 0,9 - [0,6666] \approx \boxed{0,3} \sim \text{ganho de tempo}$$

S N  
 10-30  $\rightarrow 2 \rightarrow 1 \ 1$   
 30-60  $\rightarrow 2 \rightarrow 1 \ 1$   
 760  $\rightarrow 2 \rightarrow 0 \ 2$

ganho (tipo) =

$$0,9 - \left[ \frac{2}{6} \in \left( \frac{1}{2}, \frac{1}{2} \right) + \frac{1}{6} \in (0,1) + \frac{2}{6} \in \left( \frac{1}{2}, \frac{1}{2} \right) + \frac{1}{6} \in (0,1) \right]$$

$$0,9 - \left[ \frac{2}{6} + \frac{2}{6} \right] \approx \boxed{0,3} \sim \text{ganho de tipo}$$

S N  
 To:briles  $\rightarrow 2 \rightarrow 1 \ 1$   
 Frances  $\rightarrow 1 \rightarrow 0 \ 1$   
 Hamb  $\rightarrow 2 \rightarrow 1 \ 1$   
 Italiano  $\rightarrow 1 \rightarrow 0 \ 1$

ganho (res) =

$$0,9 - \left[ \frac{2}{6} \in (0,1) + \frac{4}{6} \in \left( \frac{1}{2}, \frac{1}{2} \right) \right]$$

$$0,9 - \left[ \frac{4}{6} \right] \approx \boxed{0,23} \sim \text{ganho de res}$$

S N  
 Sim  $\rightarrow 2 \rightarrow 0 \ 2$   
 Não  $\rightarrow 4 \rightarrow 2 \ 2$



ganho(chuva) =

$$0,9 - \left[ \frac{2}{6} \epsilon \left( \frac{3}{2}, \frac{1}{2} \right) + \frac{4}{6} \epsilon \left( \frac{1}{4}, \frac{3}{4} \right) \right]$$

$\approx 0,8113$

$$0,9 - \left[ \frac{2}{6} + \frac{4}{6} \times 0,8113 \right]$$

$$= 0,9 - [0,8742] = \boxed{0,0258} \sim \text{ganho de chuva}$$

S N  
Sim → 2 - 1 1  
Não → 4 - 1 3

ganho(preço) =

$$0,9 - \left[ \frac{4}{6} \epsilon \left( \frac{1}{2}, \frac{1}{2} \right) + \frac{2}{6} \epsilon (0,1) \right]$$

$$0,9 - \left[ \frac{4}{6} \right] = \boxed{0,23} \sim \text{ganho de preço}$$

S N  
R → 4 2 2  
RRR → 2 → 0 2

ganho(fome) =

$$0,9 - \left[ \frac{4}{6} \epsilon \left( \frac{1}{2}, \frac{1}{2} \right) + \frac{2}{6} \epsilon (0,1) \right]$$

$$0,9 - \left[ \frac{4}{6} \right] \approx \boxed{0,23} \sim \text{ganho de fome}$$

S N  
Sim → 4 - 2 2  
Não → 2 - 0 2

ganho(sex Sab)

$$0,9 - \left[ \frac{5}{6} \epsilon \left( \frac{2}{5}, \frac{3}{5} \right) + \frac{1}{6} \epsilon (0,1) \right]$$

$\approx 0,97095$

$$= 0,9 - \left[ \frac{5}{6} \times 0,97095 \right] \approx \boxed{0,0908} \sim \text{ganho sex Sab}$$

S N  
Sim → 5 - 2 3  
Não → 1 - 0 1

ganho(bor) =

$$0,9 - \left[ \frac{3}{6} \in \left( \frac{1}{3}, \frac{2}{3} \right) + \frac{3}{6} \in \left( \frac{1}{3}, \frac{2}{3} \right) \right]$$

$$= 0,9 - \left[ \frac{3}{6} \times 0,9182 + \frac{3}{6} \times 0,9182 \right]$$

$$= 0,9 - [0,9182] \approx \boxed{0} \sim \text{ganho de bor}$$

	S	N
Sim → 3 →	1	2
Não → 3 →	1	2

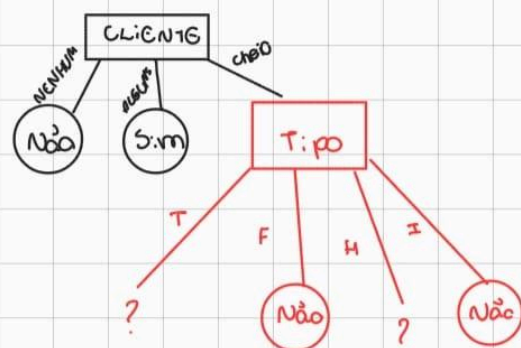
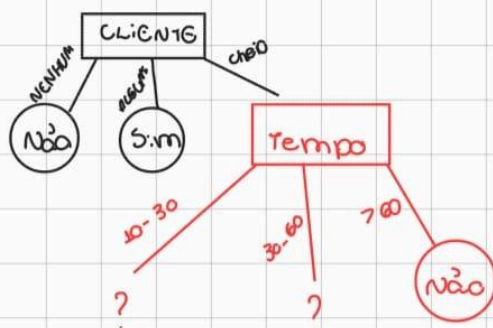
ganho(alternat:uo)

$$0,9 - \left[ \frac{5}{6} \in \left( \frac{2}{5}, \frac{3}{5} \right) + \frac{1}{6} \in (0,1) \right]$$

$\approx 0,97095$

$$0,9 - [0,8] \approx \boxed{0,1} \sim \text{ganho alternat:uo}$$

	S	N
Sim → 5 →	2	3
Não → 1 →	0	1



Resposta: Para o segundo nível da árvore podemos considerar duas opções, "TEMPO" e "TIPO", pois foram esses atributos que apresentaram maior ganho.