

1- a)

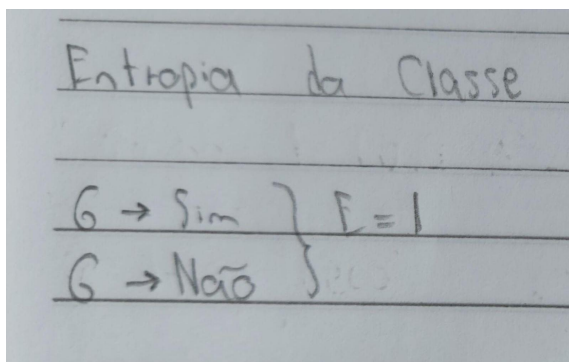
- **Classificação:** prever ou descrever uma classe nominal
 - Ex.: determinar se alguém vai jogar tênis com base nos atributos de como está o clima no dia (SIM ou NÃO)
- **Regressão:** prever ou descrever uma classe numérica
 - Ex.: prever o valor da comida em um restaurante, com base nos atributos (valor numérico)
- **Agrupamento (clusterização):** agrupar as instâncias, de acordo com os atributos de entrada.
 - Ex.: identificar perfis de usuários em redes sociais
- **Regras de associação:** buscar semelhanças entre os elementos
 - Ex.: (e-commerce) quem comprou A também comprou B

b)

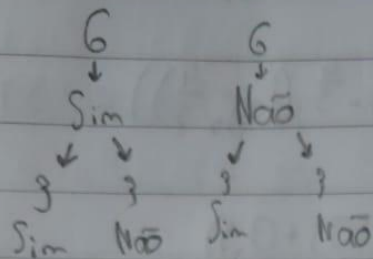
- **Supervisionado:** indutor recebe conjunto de exemplos (entrada, rótulo_desejado)
- **Não supervisionado:** indutor receber apenas atributos de entrada
- **Semi-supervisionado:** utiliza dados rotulados e não rotulados para aprender o modelo
- **Reforço:** maximização das recompensas do resultado
- **Deep Learning:** rede neural com muitas camadas

c) KDD é o processo de descoberta de padrões interessantes e desconhecidos em grandes bancos de dados. Em suma, o KDD é capaz de detectar dados e informações armazenadas nas bases de dados, transformando-as em conhecimento.

2- a) Através dos cálculos, o melhor atributo para a raiz da árvore é o atributo "Cliente".

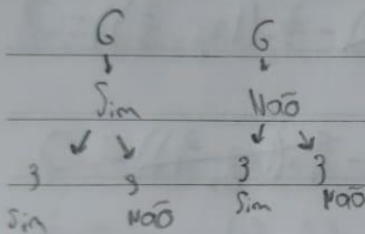


Ganho - "Alternativo":



Confusão máxima
Ganho = 0

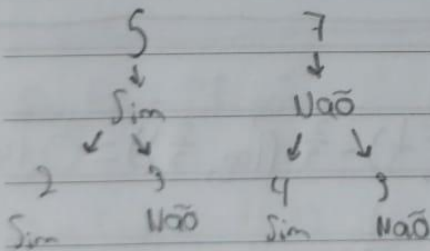
Ganho - "Bar":



Confusão máxima
Ganho = 0

Ganho - "Sex/Sab":

$$\frac{5}{12} \left(-\frac{2}{5} \cdot \left(\log_2 \frac{2}{5} \right) - \frac{3}{5} \cdot \left(\log_2 \frac{3}{5} \right) \right) = 0,404$$

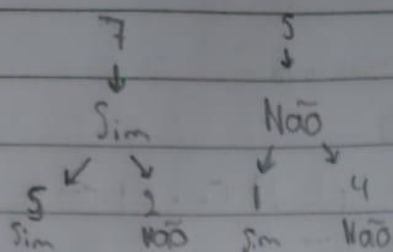


$$\frac{7}{12} \left(-\frac{4}{7} \cdot \left(\log_2 \frac{4}{7} \right) - \frac{3}{7} \cdot \left(\log_2 \frac{3}{7} \right) \right) = 0,574$$

$$\text{Ganho} \approx 1 - (0,404 + 0,574)$$

$$\approx 1 - 0,978 = 0,022$$

Ganho - "Fome"



$$\frac{7}{12} \left(-\frac{5}{7} \cdot (\log_2 \frac{5}{7}) - \frac{2}{7} \cdot (\log_2 \frac{2}{7}) \right) \approx 0,503$$

$$\frac{5}{12} \left(-\frac{1}{5} \cdot (\log_2 \frac{1}{5}) - \frac{4}{5} \cdot (\log_2 \frac{4}{5}) \right) \approx 0,300$$

$$\text{Ganho} \approx 1 - (0,503 + 0,3)$$

$$\approx 1 - 0,803 = 0,197$$

Ganho - "Cliente"

2 → nenhum (0, 2)

4 → alguns (4, 0)

6 → cheio (2, 4)

$$\frac{2}{12} \left(-\frac{0}{2} \cdot (\log_2 \frac{0}{2}) - \frac{2}{2} \cdot (\log_2 \frac{2}{2}) \right) = 0$$

$$\frac{4}{12} \left(-\frac{4}{4} \cdot (\log_2 \frac{4}{4}) - \frac{0}{4} \cdot (\log_2 \frac{0}{4}) \right) = 0$$

$$\frac{6}{12} \left(-\frac{2}{6} \cdot (\log_2 \frac{2}{6}) - \frac{4}{6} \cdot (\log_2 \frac{4}{6}) \right) \approx 0,459$$

$$\text{Ganho} = 1 - (0,459) \approx 0,541$$

Ganho - "Preço"

3 → RRR (1, 2)

2 → RR (2, 0)

7 → R (3, 4)

$$\frac{3}{12} \left(-\frac{1}{3} \cdot (\log_2 \frac{1}{3}) - \frac{2}{3} \cdot (\log_2 \frac{2}{3}) \right) \approx 0,144$$

$$\frac{2}{12} \left(-\frac{2}{2} \cdot (\log_2 \frac{2}{2}) \right) = 0$$

$$\frac{7}{12} \left(-\frac{3}{7} \cdot (\log_2 \frac{3}{7}) - \frac{4}{7} \cdot (\log_2 \frac{4}{7}) \right) \approx 0,574$$

$$\text{Ganho} = 1 - (0,144 + 0,574)$$

$$\text{tilibra} = 1 - 0,718 \approx 0,282$$

Ganho - "Chuva"

$$\frac{7}{12} \left(-\frac{2}{7} \cdot (\log_2 \frac{3}{7}) - \frac{4}{7} \cdot (\log_2 \frac{4}{7}) \right) \approx 0,472$$

7 → Não (3,4)

$$5 \rightarrow \text{Sim} (3,2) \quad \frac{5}{12} \left(-\frac{3}{5} \cdot (\log_2 \frac{3}{5}) - \frac{2}{5} \cdot (\log_2 \frac{2}{5}) \right) \approx 0,404$$

$$\begin{aligned} \text{Ganho} &= 1 - (0,472 + 0,404) \\ &\approx 0,124 \end{aligned}$$

Ganho - "Res"

$$\frac{7}{12} \left(-\frac{2}{7} \cdot (\log_2 \frac{3}{7}) - \frac{4}{7} \cdot (\log_2 \frac{4}{7}) \right) \approx 0,472$$

7 → Não (3,4)

$$5 \rightarrow \text{Sim} (3,2) \quad \frac{5}{12} \left(-\frac{3}{5} \cdot (\log_2 \frac{3}{5}) - \frac{2}{5} \cdot (\log_2 \frac{2}{5}) \right) \approx 0,404$$

$$\begin{aligned} \text{Ganho} &= 1 - (0,472 + 0,404) \\ &\approx 0,124 \end{aligned}$$

Ganho - "Tipo"

$$\frac{2}{12} \left(-\frac{1}{2} \cdot (\log_2 \frac{1}{2}) - \frac{1}{2} \cdot (\log_2 \frac{1}{2}) \right) \approx 0,166$$

2 → Francês (1,1)

$$2 \rightarrow \text{Italiano} (1,1) \quad \frac{4}{12} \left(-\frac{2}{4} \cdot (\log_2 \frac{2}{4}) - \frac{2}{4} \cdot (\log_2 \frac{2}{4}) \right) \approx 0,333$$

4 → Tailandês (2,2)

$$\begin{aligned} 4 \rightarrow \text{Hamburget} (2,2) \quad \text{Ganho} &= 1 - (0,166 + 0,166 + 0,333 + 0,333) \\ &\approx 0,002 \end{aligned}$$

Ganho - "Tempo"

$$\frac{6}{12} \left(-\frac{4}{6} \cdot (\log_2 \frac{4}{6}) - \frac{2}{6} \cdot (\log_2 \frac{2}{6}) \right) \approx 0,459$$

6 → 0-10 (4,2)

$$2 \rightarrow 10-30 (1,1) \quad \frac{2}{12} \left(-\frac{1}{2} \cdot (\log_2 \frac{1}{2}) - \frac{1}{2} \cdot (\log_2 \frac{1}{2}) \right) \approx 0,166$$

2 → 30-60 (1,1)

$$2 \rightarrow >60 (0,2) \quad \frac{2}{12} \left(-\frac{0}{2} \cdot (\log_2 \frac{0}{2}) - \frac{2}{2} \cdot (\log_2 \frac{2}{2}) \right) \approx 0$$

$$\text{Ganho} = 1 - (0,459 + 0,166 + 0,166) \\ \approx 0,209$$

2-b) Através dos cálculos, o melhor atributo para estar no segundo nível da árvore é o atributo "Fome".

Entropia da classe

$$2 \rightarrow \text{sim} \quad E\left(\frac{2}{6}, \frac{4}{6}\right) = -\frac{2}{6} \cdot (\log_2 \frac{2}{6}) - \frac{4}{6} \cdot (\log_2 \frac{4}{6}) \approx 0,918$$

4 → não

Segundo Nível:

Ganho - "Altetnat"

$$\frac{5}{6} \left(-\frac{2}{5} \cdot (\log_2 \frac{2}{5}) - \frac{3}{5} \cdot (\log_2 \frac{3}{5}) \right) \approx 0,809$$

5 → sim (2,3)

$$1 \rightarrow \text{Não} (0,1) \quad \frac{1}{6} \left(-\frac{1}{1} \cdot (\log_2 \frac{1}{1}) \right) \approx 0$$

$$\text{Ganho} = 0,918 - 0,809 = 0,109$$

Ganho - "Bar"

$$\frac{2}{6} \left(-\frac{1}{3} \cdot (\log_2 \frac{1}{3}) - \frac{2}{3} \cdot (\log_2 \frac{2}{3}) \right) = 0,459$$

3 → sim (1,2)

3 → não (1,2)

$$\text{Ganho} = 0,918 - (0,459 \cdot 2)$$

$$= 0$$

Ganho - "Sex/Sab"

Reaproveitar os cálculos do atributo "Alternativo"

5 → sim (2,3)

1 → Não (0,1)

$$\text{Ganho} = 0,918 - 0,809 = 0,109$$

Ganho - "fome"

$$\frac{4}{6} \left(-\frac{2}{4} \cdot (\log_2 \frac{2}{4}) - \frac{2}{4} \cdot (\log_2 \frac{2}{4}) \right) = 0,666$$

4 → sim (2,2)

2 → Não (0,2)

$$\frac{2}{6} \left(-\frac{2}{2} \cdot (\log_2 \frac{2}{2}) \right) = 0$$

$$\text{Ganho} = 0,918 - 0,666 = 0,252$$

Ganho - "Preço"

2 → RRR (0,2)

4 → R (2,2)

Reaproveitar os cálculos do atributo "fome"

$$\text{Ganho} = 0,252$$

Ganho - "Chuva"

$$\frac{2}{6} \left(-\frac{1}{2} \cdot (\log_2 \frac{1}{2}) - \frac{1}{2} \cdot (\log_2 \frac{1}{2}) \right) = 0,333$$

2 → Sim (1,1)

4 → Não (1,3)

$$\frac{4}{6} \left(-\frac{1}{4} \cdot (\log_2 \frac{1}{4}) - \frac{3}{4} \cdot (\log_2 \frac{3}{4}) \right) = 0,540$$

$$\text{Ganho} = 0,918 - 0,873 = 0,045$$

Ganho - "Res"

$$\frac{2}{6} \left(-\frac{2}{2} \cdot (\log_2 \frac{2}{2}) - \frac{2}{2} \cdot (\log_2 \frac{2}{2}) \right) = 0$$

2 → Sim (0,2)

4 → Não (2,2)

$$\frac{4}{6} \left(-\frac{2}{4} \cdot (\log_2 \frac{2}{4}) - \frac{2}{4} \cdot (\log_2 \frac{2}{4}) \right) = 0,666$$

$$\text{Ganho} = 0,252$$

Ganho - "tipo"

$$\frac{2}{6} \left(-\frac{1}{2} \cdot (\log_2 \frac{1}{2}) - \frac{1}{2} \cdot (\log_2 \frac{1}{2}) \right) \approx 0,333$$

2 → tai. (1,1)

1 → fran. (0,1) $\frac{1}{6} \left(-\frac{1}{1} \cdot (\log_2 \frac{1}{1}) \right) = 0$

2 → hamb. (1,1)

1 → ital. (0,1)

$$\text{Ganho} = 0,918 - 0,666 = 0,252$$

Ganho - "tempo"

$$\frac{2}{6} \left(-\frac{1}{2} \cdot (\log_2 \frac{1}{2}) - \frac{1}{2} \cdot (\log_2 \frac{1}{2}) \right) \approx 0,333$$

2 → 30-60 (1,1)

2 → >60 (0,2)

2 → at/ 30 (1,1)

$$\frac{2}{6} \left(-\frac{2}{2} \cdot (\log_2 \frac{2}{2}) \right) = 0$$

$$\text{Ganho} = 0,252$$

