

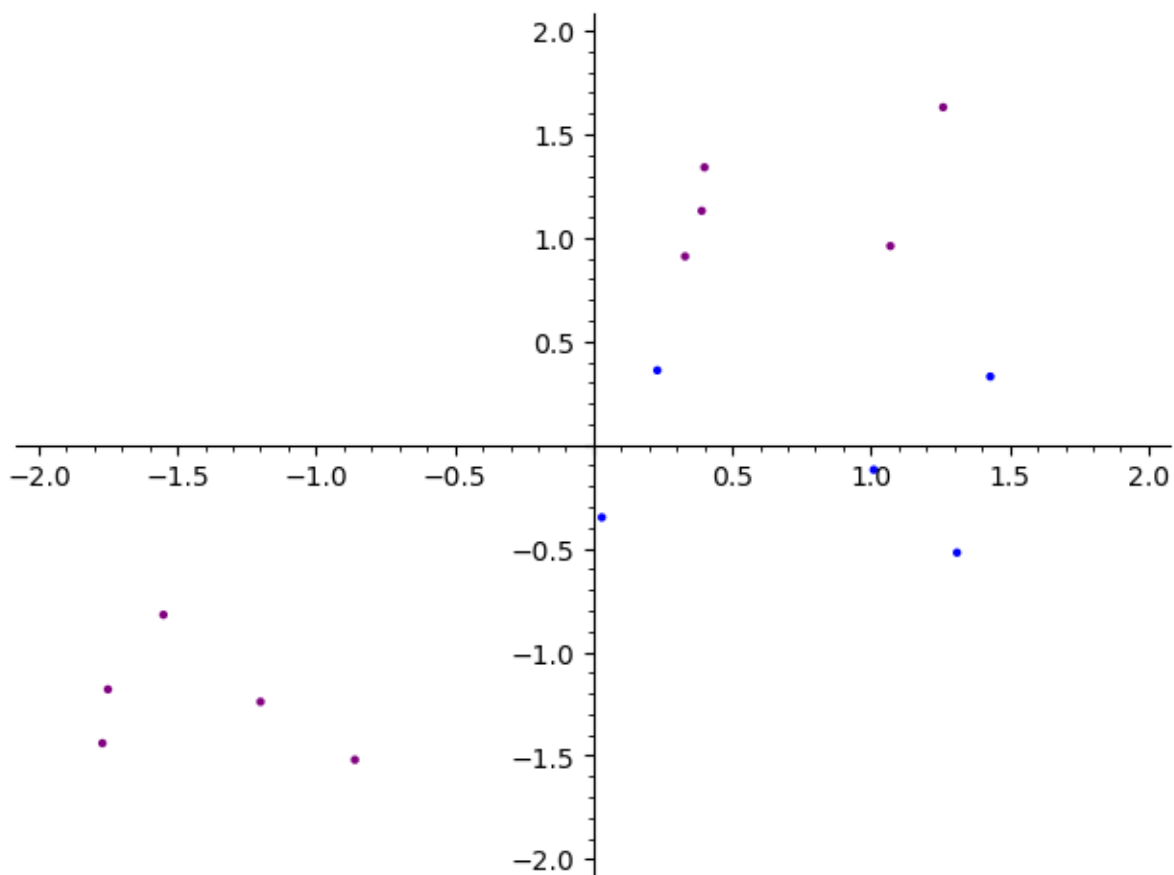
Exercício 1

$$H(\mathcal{X}) = \sum_{i=1}^n p(x_i) \log_2 \left(\frac{1}{p(x_i)} \right).$$

1.a

Represente os dados em um plano cartesiano com símbolos ou cores diferentes para cada categoria:

Resolução.



1.b

Defina uma regra que classifique corretamente os exemplos da tabela.

Resolução.

Uma possível regra para classificar corretamente os exemplos da tabela seria:

$$h(x_1, x_2) = \begin{pmatrix} \text{se } x_1 > 0 & \text{e } x_2 < 0.5 & \text{então } 1 \\ & \text{se não } & 0 \end{pmatrix} \quad (1)$$

Exercício 2

2.a

Calcule a entropia do conjunto de treinamento

Resolução.

A entropia do conjunto de treinamento pode ser calculado usando a fórmula:

$$H(V) = -\left(\frac{10}{15} \log_2 \frac{10}{15} + \frac{5}{15} \log_2 \frac{5}{15}\right)$$

Aplicando a fórmula:

$$H(V) = -\left(\frac{10}{15} \log_2 \frac{10}{15} + \frac{5}{15} \log_2 \frac{5}{15}\right)$$

$$H(V) = -\left(\frac{10}{15} - 0.58496250072 + \frac{5}{15} - 1.5849625007\right)$$

$$H(V) = -\left(\frac{2}{3} - 0.58496250072 + \frac{1}{3} - 1.5849625007\right)$$

$$H(V) = -(-0.389975 - 0.528320833)$$

$$H(V) = -(-0.389975 - 0.528320833)$$

$$H(V) = 0.918295833$$

2.b - Errado

Considere a pergunta $x_1 > 0$ e calcule a entropia média após obter a resposta a essa pergunta.

Resolução.

PS: Calculei quando $x_1 > 1$, corrigir depois

A entropia do conjunto de treinamento pode ser calculado usando a fórmula:

$$H(V) = -\left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5}\right)$$

Aplicando a fórmula:

$$\begin{aligned}H(V) &= -(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5}) \\H(V) &= -(\frac{2}{5} - 1.3219280949 + \frac{3}{5} - 0.73696559417) \\H(V) &= -(-0.528771237 - 0.442179356) \\H(V) &= -(-0.528771237 - 0.442179356) \\H(V) &= 0.970950593\end{aligned}$$

2.b

Considere a pergunta $x_1 > 0$ e calcule a entropia média após obter a resposta a essa pergunta.

Resolução.

PS: Calculei quando $x_1 > 1$, corrigir depois

A entropia do conjunto de treinamento pode ser calculado usando a fórmula:

$$H(V) = -(\frac{5}{10} \log_2 \frac{5}{10} + \frac{5}{10} \log_2 \frac{5}{10})$$

Aplicando a fórmula:

$$\begin{aligned}H(V) &= -(\frac{5}{10} \log_2 \frac{5}{10} + \frac{5}{10} \log_2 \frac{5}{10}) \\H(V) &= -(\frac{2}{5} - 1 + \frac{3}{5} - 1) \\H(V) &= -(-\frac{2}{5} - \frac{3}{5}) \\H(V) &= 1\end{aligned}$$

2.c

Considere a pergunta $x_2 > 0$ e calcule a entropia média após obter a resposta a essa pergunta.

Resolução.

A entropia do conjunto de treinamento pode ser calculado usando a fórmula:

$$H(V) = -(\frac{5}{8} \log_2 \frac{5}{8} + \frac{3}{8} \log_2 \frac{3}{8})$$

$$\begin{aligned}H(V) &= -(\frac{5}{8} \log_2 \frac{5}{8} + \frac{3}{8} \log_2 \frac{3}{8}) \\H(V) &= -(\frac{5}{8} - 0.67807190511 + \frac{3}{8} - 1.4150374993) \\H(V) &= -(-0.42379494 - 0.543890622) \\H(V) &= 0.967685562\end{aligned}$$

2.d

Calcule os ganhos de informação obtidos para as duas respostas anteriores

Resolução.

Para calcular os ganhos de informação para as duas respostas anteriores, é possível usar a fórmula abaixo:

$$H(V) = -\left(\frac{5}{8} \log_2 \frac{5}{8} + \frac{3}{8} \log_2 \frac{3}{8}\right)$$

2.e

Escolha a pergunta com maior ganho de informação para ser a raiz da árvore e complete a árvore para classificar corretamente todos os exemplos da tabela

Resolução.

Answers

- 2.a - 0,9183
- 2.b - 0,667
- 2.c - 0,9118
- 2.d - $x_1 > 0 \rightarrow 0,2516$
- 2.d - $x_2 > 0 \rightarrow 0,0065$
- 3.c - $x_1 = 0,1076$
- 3.c - $x_2 = 0,6076$
- 4.e - 0,5027
- 6.b - $\frac{10}{12}$
- 6.c - $\frac{7}{10}$