

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Instituto de Ciências Exatas e Informática

Curso de Ciência da Computação - Coração Eucarístico

Profa.: Camila Laranjeira - mila.laranjeira@gmail.com

Disciplina: Inteligência Artificial / 1o Semestre de 2022

Aluno(a):	Rafael Amauri Diniz Augusto
-----------	-----------------------------

Lista 05 - Árvores de decisão

1. Considere a base de dados seguinte em que a classe a ser predita é estar ou não com COVID-19 e responda as questões a seguir (apresente os cálculos).

Instância	Faixa etária	Falta de ar?	Tosse?	Covid-19?
1	<=18	Não	Sim	Sim
2	18 a 40	Não	Sim	Não
3	18 a 40	Não	Sim	Não
4	acima de 60	Sim	Não	Sim
5	acima de 60	Sim	Não	Não
6	acima de 60	Não	Não	Sim
7	18 a 40	Não	Sim	Sim
8	18 a 40	Sim	Sim	Não
9	40 a 60	Sim	Não	Sim
10	40 a 60	Não	Sim	Sim
11	acima de 60	Sim	Não	Não
12	acima de 60	Sim	Sim	Sim

a) Qual a quantidade de informação do conjunto original?

Faixa etária, Tosse, Falta de ar.

b) Apresente o ganho de informação ao definir cada um dos atributos como nó raiz de uma árvore de decisão.

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

Faixa Etária

≤ 18

1

18 a 40

7		
2	3	8

40 a 60

9	10
---	----

Acima de 60

4	6	12
5	11	

Ganho de informação

$$\begin{aligned} & 1 - \left[\frac{1}{12} * E(1, 0) + \frac{4}{12} * E\left(\frac{1}{4}, \frac{3}{4}\right) + \frac{2}{12} * E(1, 0) + \frac{5}{12} * E\left(\frac{3}{5}, \frac{2}{5}\right) \right] = \\ & 1 - \left[\frac{1}{12} * 0 + \frac{4}{12} * 0.81 + \frac{2}{12} * 0 + \frac{5}{12} * 0.97 \right] = \\ & 1 - \left[\frac{4}{12} * 0.81 + \frac{5}{12} * 0.97 \right] = \\ & 1 - [0.27 + 0.34] = \\ & 1 - 0.61 = 0.39 \end{aligned}$$

Falta de ar

Sim

4	9	12
5	8	11

Não

1	6	7	10
2	3		

Ganho de informação

$$1 - \left[\frac{6}{12} * E\left(\frac{3}{6}, \frac{3}{6}\right) + \frac{6}{12} * E\left(\frac{4}{6}, \frac{2}{6}\right) \right] =$$

$$1 - \left[\frac{6}{12} * 1 + \frac{6}{12} * 0.92 \right] =$$

$$1 - [0.5 + 0.46] =$$

$$1 - 0.96 = 0.04$$

Tosse

Sim

1	7	10	12
2	3	8	

Não

4	6	9
5	11	

Ganho de informação

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

$$1 - \left[\frac{7}{12} * 0.98 + \frac{5}{12} * 0.97 \right] =$$

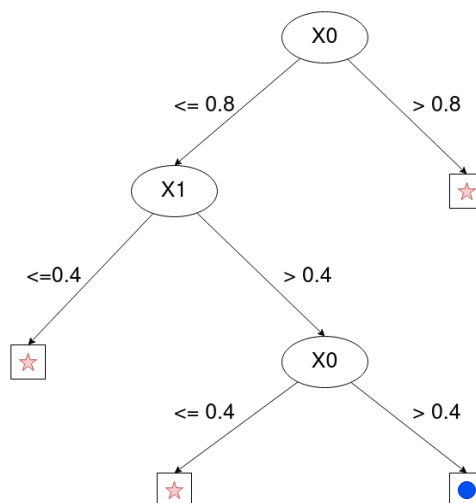
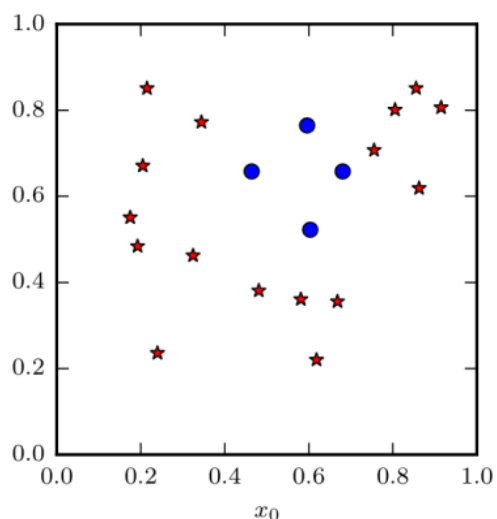
$$1 - [0.57 + 0.40] =$$

$$1 - 0.97 = 0.03$$

c) Qual o melhor atributo para a raiz da árvore considerando uma abordagem gulosa?

Faixa Etária, pois apresenta o maior ganho de informação.

2. Dado o conjunto de dados a seguir, use o bom-senso para propor uma árvore de decisão que separe as estrelas vermelhas das bolinhas azuis. Apresente a representação geométrica (pode riscar o próprio desenho) e o modelo da árvore que deve ter pelo menos 3 níveis.



Calcule a entropia das folhas de sua árvore:

P(Estrela): $16/20 = 4/5$

P(Bola): $4/20 = 1/5$

$$\begin{aligned}
 E &= - \left(\frac{4}{5} * \log_2 * \frac{4}{5} + \frac{1}{5} * \log_2 * \frac{1}{5} \right) \\
 &= - (0.8 * (-0.32) + 0.2 * (-2.32)) \\
 &= - (-0.256 - 0.464) \\
 &= - (-0.72) \\
 &= 0.72
 \end{aligned}$$

Classifique os pontos a seguir usando a sua árvore (estrela/círculo):

x0: 0.4, x1: 1.0	Estrela
x0: 0.8, x1: 0.6	Círculo
x0: 0.6, x1: 0.4	Estrela

3. Dada a matriz de confusão a seguir, preencha o reporte de classificação da tabela.

		Foi classificado como		
		A	B	C
Era da classe	A	30	10	20
	B	10	60	5
	C	20	10	60

	A	B	C	Média
Acurácia	0,5	0,8	0,66	0.66
Precisão	0,5	0,75	0,7	0.66
Sensibilidade	0,5	0,8	0,66	0.66

Desenvolvimento

Foi classificado como					
Era da classe		A	B	C	Total
	A	30	10	20	60
	B	10	60	5	75
	C	20	10	60	90
	Total	60	80	85	225

Acurácia de A = $30 / 60 = 0.5$

Precisão de A = $30 / 60 = 0.5$

Sensibilidade de A = $30 / 60 = 0.5$

Acurácia de B = $60 / 75 = 0.8$

Precisão de B = $60 / 80 = 0.75$

Sensibilidade de B = $60 / 75 = 0.8$

Acurácia de C = $60 / 90 = 0.66$

Precisão de C = $60 / 85 = 0.7$

Sensibilidade de C = $60 / 90 = 0.66$

Acurácia da Média = $(30 + 60 + 60) / 60 = 150 / 225 = 0.66$

Precisão(Média)

$$(60 * 0.5) / 225 + (75 * 0.75) / 225 + (90 * 0.7) / 225 =$$

$$30 / 225 + 56.25 / 225 + 63 / 225 =$$

$$0.13 + 0.25 + 0.28 = 0.66$$

Sensibilidade(Média)

$$(60 * 0.5) / 225 + (75 * 0.8) / 225 + (90 * 0.66) / 225 =$$

$$30 / 225 + 60 / 225 + 59.4 / 225 =$$

$$0.13 + 0.27 + 0.26 = 0.66$$