

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

Aluna(o):	Gustavo Lopes Rodrigues
-----------	-------------------------

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?

A quantidade de informação é 48

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

Faixa etária

<=18	18 a 40	40 a 60	Acima de 60
Sim: 1	Sim: 1	Sim: 2	Sim: 3
Não: 0	Não: 3	Não: 0	Não: 2
Total: 1	Total: 4	Total: 2	Total: 5

Ganho(Faixa etária) : $1 - [1/12 * E(1,0) + 4/12 * E(1/4, 3/4) + 2/12 * E(1,0) + 5/12 * E(2/3, 1/3)] =$
 Ganho(Faixa etária) : $1 - [0 + 4/12 * (-1/4 * \log_2(1/4) - 3/4 * \log_2(3/4)) + 0 + 5/12 * (-2/3 * \log_2(2/3) - 1/3 * \log_2(1/3))] =$

Ganho(Faixa etária) : $1 - [0 + 0.27 + 0 + 0.40] = 0.33$

Falta de Ar

Não	Sim
Sim: 4	Sim: 3
Não: 2	Não: 3
Total: 6	Total: 6

Ganho(Falta de Ar) : $1 - [6/12 * E(2/3, 1/3) + 6/12 * E(1/2, 1/2)] =$

Ganho(Falta de Ar) : $1 - [0.45 + 0.5] = 0.05$

Tosse:

Não	Sim
Sim: 3	Sim: 4
Não: 2	Não: 3
Total: 5	Total: 7

Ganho(Tosse) : $1 - [5/12 * E(3/5, 2/5) + 7/12 * E(4/7, 3/7)] =$

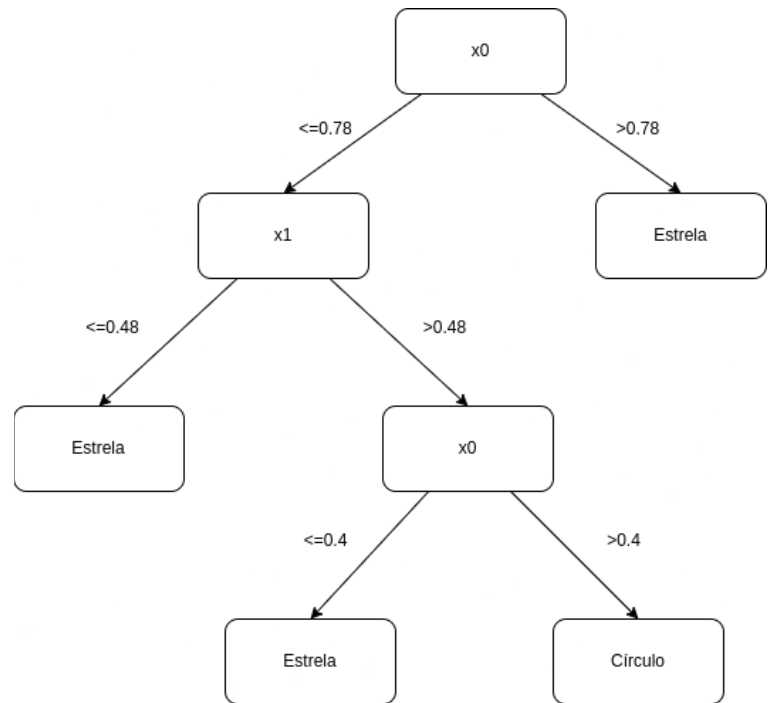
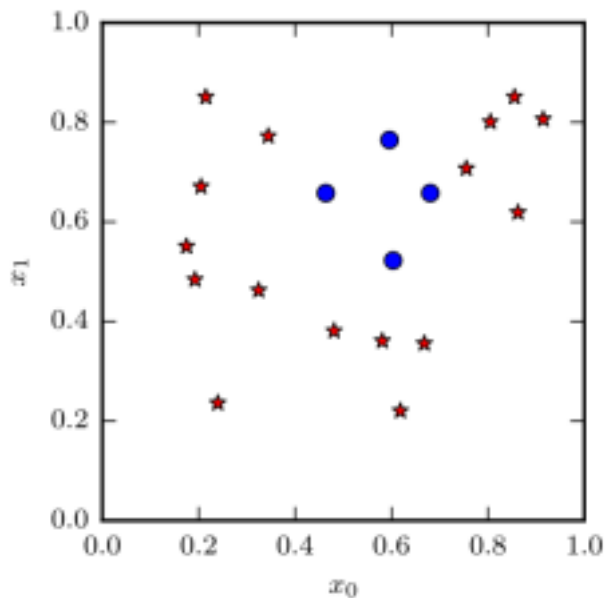
Ganho(Tosse) : $1 - [0.40 + 0.57] = 0.03$

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

O melhor atributo será a faixa etária por possuir 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.

// Cole aqui o desenho de sua árvore



Calcule a entropia das folhas de sua árvore:

$$\text{Prob}(\text{Estrela}) = 16/20 = \frac{4}{5}$$

$$\text{Prob}(\text{Círculo}) = 4/20 = \frac{1}{5}$$

$$E(\frac{4}{5}, \frac{1}{5}) = -(\frac{4}{5} * \log(\text{base2})(\frac{4}{5})) + \frac{1}{5} * \log(\text{base2})(\frac{1}{5}))$$

$$-(0.78 * (-0.32) + 2 * (-2.32))$$

$$0.72192$$

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

$x_0: 0.4, x_1: 1.0$	Estrela
$x_0: 0.8, x_1: 0.6$	Círculo

$x_0: 0.6, x_1: 0.4$	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	$30/60 = 0.5$	$60/75 = 0.8$	$60/90 = 0.67$	$150/225 = 0.66$
Precisão	$30/60 = 0.5$	$60/80 = 0.75$	$60/85 = 0.71$	$60/225 \cdot 0.5 + 75/225 \cdot 0.75 + 90/225 \cdot 0.71 = 0.667$
Sensibilidade	$30/60 = 0.5$	$60/75 = 0.8$	$60/90 = 0.67$	$60/225 \cdot 0.5 + 75/225 \cdot 0.8 + 90/225 \cdot 0.66 = 0.664$