

Questão 1: Um banco de dados tem um cadastro de 500 registros de pessoas que usam do serviço de uma empresa, sendo 380 homens e 120 mulheres. Para retirar uma amostra de 10% desses cadastros obedecendo a amostragem estratificada proporcional, determine o número de cadastros de homens e mulheres que devem estar presentes nessa amostra.

Questão 2: Explique a diferença entre seleção de atributos por ordenação e por subconjunto.

Questão 3: Converta para uma das escalas numéricas:

- a) carro: família, luxo, esporte
- b) sexo: masculino, feminino
- c) nota musical: dó, ré, mi, fa, só, lá, si

Questão 4: Discretize por largura e por frequência para $k = 3$: 25, 11, 10, 8, 22, 3, 30, 18, 12, 21, 18, 40, 14, 23, 7.

Questão 5: Calcule as distâncias Manhattan, Euclidiana e Suprema para:

- a) $x_1 = (-3, 0, 5, 2)$ e $x_2 = (-5, -10, 3, 11)$;
- b) $x_3 = (0, 0, 1)$ $x_4 = (1, 0, 1)$

Questão 6: Calcule a similaridade do cosseno para os vetores apresentados no exercício anterior (questão 5)

Questão 7: Explique a diferença entre classificação binária, unária, multiclasse e multi-rótulo

Questão 8: Monte o dendograma com a escala de similaridade da matriz a seguir

$$P_0^{SM} = \begin{bmatrix} 1 & 0.75 & 0.26 & 0.21 & 0.18 \\ 0.75 & 1 & 0.44 & 0.35 & 0.20 \\ 0.26 & 0.44 & 1 & 0.96 & 0.90 \\ 0.21 & 0.35 & 0.96 & 1 & 0.98 \\ 0.18 & 0.20 & 0.90 & 0.98 & 1 \end{bmatrix}$$

Questão 9: Calcule a silhueta para x_3 e x_7 . Qual está melhor situado em seu cluster?

cluster 1

$x_1 = (0.0, 0.5, 0.9, 0.5)$

$x_2 = (1.5, 1.1, 0.9, 0.8)$

$x_3 = (1.1, 0.9, 1.3, 1.2)$

cluster 2

$x_4 = (2.0, 2.5, 2.9, 2.5)$

$x_5 = (3.5, 3.1, 2.9, 2.8)$

$x_6 = (3.1, 2.9, 3.3, 3.2)$

$x_7 = (5.0, 4.2, 3.9, 4.1)$

Questão 11: O que é “aprendizado de conceitos”?

Questão 12: Explique o conceito de ordenação de hipóteses mais gerais para mais específicas.

Questão 13: O que é o “espaço de versões”?

Questão 14: Descreva o algoritmo de aprendizado denominado “eliminação de candidatos”.

Questão 15: Quais o papel das instâncias positivas, das instâncias negativas de classificação binária? Relacione com o papel das instâncias utilizadas em classificação unária.

Questão 16: Dado o exemplo da tabela abaixo, use o algoritmo de eliminação de candidato geral para específica para aprender o conceito bola (atributos tamanho e cor).

Tamanho	Cor	Forma	Classe
Pequena	Vermelha	Bola	Sim
Pequena	Branca	Bola	Sim
Grande	Azul	Bola	Sim

Tabela 1: Conjunto de trinamento

Questão 17: Dado o exemplo da tabela abaixo, use estratégia de pesquisa geral para específica para aprender o conceito bola.

Tamanho	Cor	Forma	Classe
Pequena	Vermelha	Tijolo	Não
Grande	Branca	Bola	Sim
Grande	Azul	Cubo	Não
Pequena	Azul	Bola	Sim

Tabela 2: Conjunto de trinamento

Questão 18: Dado o exemplo da tabela abaixo, use o algoritmo de eliminação candidata para aprender o conceito bola vermelha.

Tamanho	Cor	Forma	Saída
Pequena	Vermelha	Bola	Sim
Pequena	Azul	Bola	Não
Grande	Vermelha	Bola	Sim
Grande	Vermelha	Cubo	Não

Tabela 3: Conjunto de treinamento

Questão 19: Para cada uma das funções booleanas seguintes apresente uma árvore de decisão que as representem:

a) $A \wedge \neg B$

b) $(A \wedge B) \vee (C \wedge D)$

Questão 20: Considere o seguinte conjunto de treinamento

Instância	Classificação	a_1	a_2
1	+	V	V
2	+	V	V
3	-	V	F
4	+	F	F

5	-	F	V
6	-	F	V

- a)** Qual a entropia desta coleção de exemplos de treinamento com que diz respeito a classificação da função alvo?
- b)** Qual é o ganho de informação de a_1 em relação a estes exemplos de treinamento?
- c)** Qual é o ganho de informação de a_2 em relação a estes exemplos de treinamento?
- d)** Induza a árvore usando a medida Gini Média Ponderada (GiniP)