

Teste - Versão 1  
08/05/2025

Aprendizagem Automática  
90 minutos

Nome: \_\_\_\_\_

ID: \_\_\_\_\_

Problema	Valores	Classificação
1	5	
2	2.5	
3	2	
4	2	
5	2	
6	2	
7	2	
8	3	
Total	20.5	

## Problema 1 (Escolha Múltipla - 5 Valores)

1.) Para cada uma das seguintes questões circula a opção correta.

1.1) Por que é importante aplicar *scaling* aos dados antes de usar algoritmos como k-NN ou SVM?

- a) Para reduzir o tempo de treino
- b) Porque estes algoritmos são muito sensíveis à escala das variáveis**
- c) Para melhorar a interpretabilidade dos dados
- d) Porque melhora a codificação de variáveis categóricas

1.2) Qual das alternativas abaixo é mais apropriada para lidar com valores omissos em conjuntos de dados numéricos?

- a) Excluir todos os dados do conjunto de treino
- b) Preencher com zero, sempre
- c) Usar imputação, como média ou mediana, dependendo da distribuição**
- d) Substituir por valores aleatórios

1.3) Qual das alternativas a seguir descreve corretamente um método do tipo filtro (*filter*) para seleção de atributos (*feature selection*)?

- a) Treinar um modelo e observar os coeficientes das variáveis
- b) Utilizar medidas estatísticas como correlação ou teste qui-quadrado**
- c) Aplicar um algoritmo de busca recursiva por subconjuntos
- d) Usar regularização L1 para eliminar atributos irrelevantes

1.4) Qual das opções abaixo é um sintoma típico de *overfitting*?

- a) Performance baixa no treino e alta no teste
- b) Performance alta tanto no treino quanto no teste
- c) Performance muito alta no treino e baixa no teste**
- d) Performance constante em todos os conjuntos

1.5) Num pipeline supervisionado típico de *Machine Learning*, qual das seguintes sequências de etapas é mais adequada para garantir um bom e justo desempenho na generalização do modelo?

a) Recolha dos dados → Divisão treino/teste → Treino do modelo → Pré-processamento completo (inclusive imputação e *scaling*) → Avaliação com dados de teste

b) Recolha dos dados → Pré-processamento completo (inclusive imputação e *scaling*) → Divisão treino/teste → Treino do modelo → Avaliação com dados de teste

c) Recolha dos dados → Divisão em treino/teste → Pré-processamento completo (inclusive imputação e *scaling*) → Treino do modelo → Avaliação com dados de teste

d) Recolha dos dados → Divisão em treino/teste → Pré-processamento completo (inclusive imputação e *scaling*) → Avaliação com dados de teste → Treino do modelo

## Problema 2 (Verdadeiro ou Falso – 2.5 Valores)

2.) Para cada uma das afirmações seguintes diz se esta é verdadeira (V) ou falsa (F).

2.1) Um modelo com alta variância tende a apresentar *overfitting*. \_\_\_\_V\_\_\_\_

2.2) Para um problema de deteção de fraudes em cartões de crédito, onde a classe de fraude é muito rara, a *accuracy* é a melhor métrica a ser utilizada. \_\_\_\_F\_\_\_\_

2.3) Ao aplicar PCA, os componentes principais são sempre ortogonais entre si. \_\_\_\_V\_\_\_\_

2.4) No método de *clustering* hierárquico aglomerativo, cada ponto de dados começa como um cluster e, a cada iteração, os clusters são combinados. \_\_\_\_V\_\_\_\_

2.5) Num problema de classificação com SVM, os vetores de suporte são os pontos mais distantes do hiperplano de separação. \_\_\_\_F\_\_\_\_

## Problema 3 (Clustering - 2 Valores)

3.1) Usa a matriz de distâncias apresentada na Tabela 1 para efetuar o agrupamento hierárquico pelo método de ligação simples (*single linkage*). O resultado deve ser apresentado através de um dendrograma, o qual deverá indicar claramente a ordem de agrupamento dos pontos.

Tabela 1 – Matriz de distâncias.

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

For the single link or MIN version of hierarchical clustering, the proximity of two clusters is defined as the minimum of the distance (maximum of the similarity) between any two points in the two different clusters.

Steps:

- Using graph terminology, start with all points as singleton clusters.
- Add links between points one at a time (shortest links first).
- These single links combine the points into clusters.

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

Combine P1 and P2:

$$\begin{aligned} \text{dist}(\{P1, P2\}, \{P3\}) &= \min(\text{dist}(P1, P3), \text{dist}(P2, P3)) \\ &= \min(0.41, 0.64) \\ &= 0.41 \end{aligned}$$

$$\begin{aligned} \text{dist}(\{P1, P2\}, \{P4\}) &= \min(\text{dist}(P1, P4), \text{dist}(P2, P4)) \\ &= \min(0.55, 0.47) \\ &= 0.47 \end{aligned}$$

$$\begin{aligned} \text{dist}(\{P1, P2\}, \{P5\}) &= \min(\text{dist}(P1, P5), \text{dist}(P2, P5)) \\ &= \min(0.35, 0.98) \\ &= 0.35 \end{aligned}$$

	P12	P3	P4	P5
P12	0.00	0.41	0.55	0.35
P3	0.41	0.00	0.44	0.85
P4	0.55	0.44	0.00	0.76
P5	0.35	0.85	0.76	0.00

Combine P12 and P5:

$$\begin{aligned} dist(\{P12, P5\}, \{P3\}) &= \min(dist(P12, P3), dist(P5, P3)) \\ &= \min(0.41, 0.85) \\ &= 0.41 \end{aligned}$$

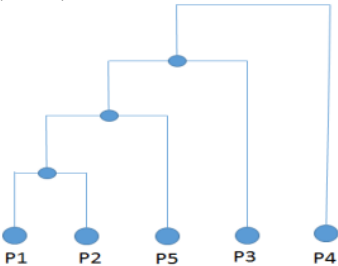
$$\begin{aligned} dist(\{P12, P5\}, \{P4\}) &= \min(dist(P12, P4), dist(P5, P4)) \\ &= \min(0.55, 0.76) \\ &= 0.55 \end{aligned}$$

	P125	P3	P4
P125	0.00	0.41	0.55
P3	0.41	0.00	0.44
P4	0.55	0.44	0.00

Combine P125 and P3:

$$\begin{aligned} dist(\{P125, P3\}, \{P4\}) &= \min(dist(P125, P4), dist(P3, P4)) \\ &= \min(0.55, 0.44) \\ &= 0.44 \end{aligned}$$

	P1235	P4
P1235	0.00	0.44
P4	0.44	0.00



Single Link Dendrogram

## Problema 4 (Naïve Bayes - 2 Valores)

4.) Com base nos dados de treino da tabela abaixo (Tabela 2), prevê a classe para o novo exemplo seguinte, utilizando o classificador Naïve Bayes:

idade<=30, rendimento=médio, estudante=sim, avaliação\_de\_crédito=razoável

#	idade	rendimento	estudante	avaliação_de_credito	Classe: comprar_computador
1	<=30	alto	não	razoável	não
2	<=30	alto	não	excelente	não
3	>30; <= 40	alto	não	razoável	sim
4	>40	médio	não	razoável	sim
5	>40	baixo	sim	razoável	sim
6	>40	baixo	sim	excelente	não
7	>40	baixo	sim	excelente	sim
8	<=30	médio	não	razoável	não
9	<=30	baixo	sim	razoável	sim
10	>40	médio	sim	razoável	sim
11	<=30	médio	sim	excelente	sim
12	>30; <= 40	médio	não	excelente	sim
13	>30; <= 40	alto	sim	razoável	sim
14	>40	médio	não	excelente	não

### *Solution:*

E= age<=30, income=medium, student=yes, credit-rating=fair  
 E<sub>1</sub> is age<=30, E<sub>2</sub> is income=medium, student=yes, E<sub>4</sub> is credit-rating=fair  
 We need to compute P(yes|E) and P(no|E) and compare them.

$$P(\text{yes} | E) = \frac{P(E_1 | \text{yes}) P(E_2 | \text{yes}) P(E_3 | \text{yes}) P(E_4 | \text{yes}) P(\text{yes})}{P(E)}$$

$$P(\text{yes}) = 9/14 = 0.643 \quad P(\text{no}) = 5/14 = 0.357$$

$$P(E_1 | \text{yes}) = 2/9 = 0.222 \quad P(E_1 | \text{no}) = 3/5 = 0.6$$

$$P(E_2 | \text{yes}) = 4/9 = 0.444 \quad P(E_2 | \text{no}) = 2/5 = 0.4$$

$$P(E_3 | \text{yes}) = 6/9 = 0.667 \quad P(E_3 | \text{no}) = 1/5 = 0.2$$

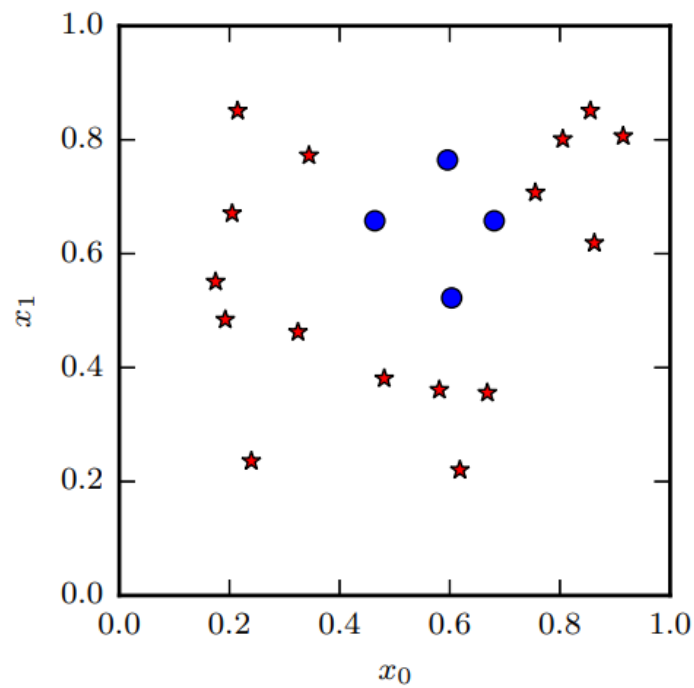
$$P(E_4 | \text{yes}) = 6/9 = 0.667 \quad P(E_4 | \text{no}) = 2/5 = 0.4$$

$$P(\text{yes} | E) = \frac{0.222 \cdot 0.444 \cdot 0.667 \cdot 0.668 \cdot 0.443}{P(E)} = \frac{0.028}{P(E)} \quad P(\text{no} | E) = \frac{0.6 \cdot 0.4 \cdot 0.2 \cdot 0.4 \cdot 0.357}{P(E)} = \frac{0.007}{P(E)}$$

Hence, the Naïve Bayes classifier predicts buys\_computer=yes for the new example.

## Problema 5 (Árvores de Decisão - 2 Valores)

5.1) Calcula a entropia do dataset seguinte. (0.5 Valores)

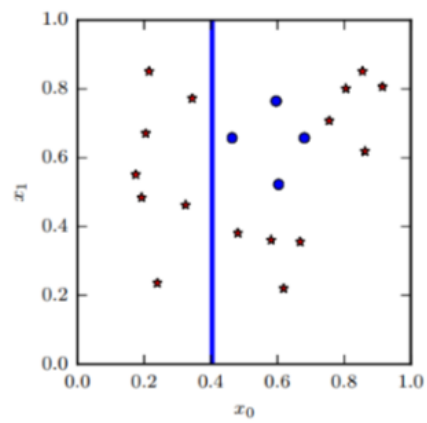


$$E = - \sum_{i=1}^n p_i \log_2(p_i)$$

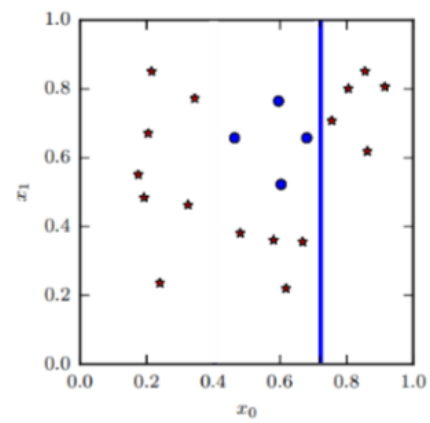
$$E = - (16/20 \log_2(16/20) + 4/20 \log_2(4/20))$$

5.2) Qual dos seguintes *splits* iniciais escolherias? Apresenta todos os cálculos efetuados. (1.5 valores)

SPLIT 1



SPLIT 2



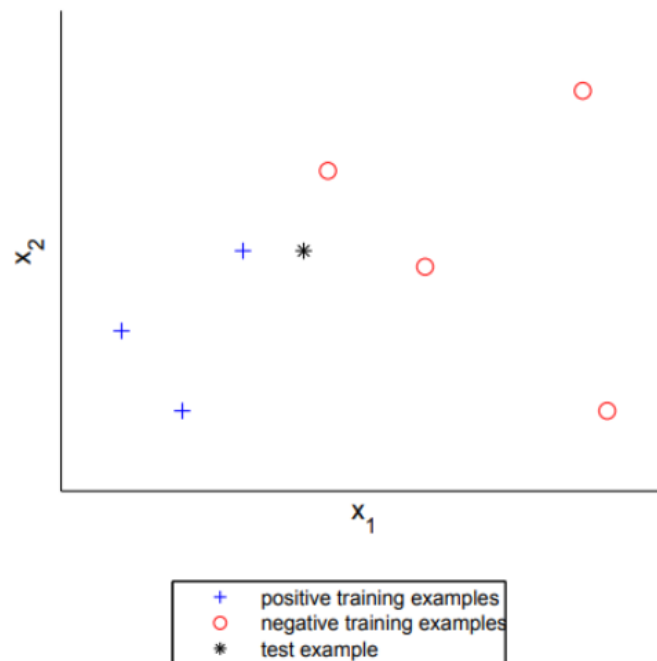
SPLIT 1

CALCULAR INFORMATION GAIN PARA OS DOIS CASOS



## Problema 6 (KNN - 2 Valores)

6.1) Na figura seguinte estão representados dados de treino e um único ponto de teste para a tarefa de classificação dados dois atributos contínuos  $X_1$  e  $X_2$ . Para cada valor de  $k$ , circule a previsão do classificador *k-nearest neighbors* para o ponto de teste representado (\*). (1 Valor)



6.1.1) Previsão para  $k=1$ : (0.5 Valores)

a. positive (+)

b. negative (o)

6.1.2) Previsão para  $k=3$ : (0.5 Valores)

a. positive (+)

b. negative (o)

6.1.3) Previsão para  $k=7$ : (0.5 Valores)

a. positive (+)

b. negative (o)

6.2) Supõe que a figura anterior se trata de um problema de regressão. Usando um KNN com  $k=3$  os exemplos mais próximos do exemplo de teste (\*) estão representados na tabela seguinte (Tabela 3). Qual será a previsão do modelo? (1 Valor)

Tabela 3 – Vizinhos mais próximos.

	x1	x2	y
Vizinho 1	5	3	10
Vizinho 2	7	5	20
Vizinho 3	6	6	15

$$y = (10 + 20 + 15) / 3 = 15$$

### Problema 7 (Regressão Linear - 2 Valores)

7.) Considera a tabela abaixo (Tabela 4). Supõe que queremos prever a variável  $y$  sabendo as features  $x_1$ ,  $x_2$  usando um modelo de regressão linear  $\hat{y} = \beta_1 x_1 + \beta_2 x_2$ .

Tabela 4

x1	x2	y
1	1	1
-2	1	-5
3	0	6

7.1) Quais os valores dos coeficientes  $\beta = (\beta_1, \beta_2)$  que minimizam a soma dos erros ao quadrado ( $\sum (y - \hat{y})^2$ )? (1 Valor)

a. (2, -1)

b. (3, -1)

c. (-1, 2)

d. (-1, 3)

e. nenhuma das anteriores

7.2) Usando os coeficientes da alínea anterior qual o valor da variável  $y$  tendo em conta que  $x_1=3$  e  $x_2=3$ . Apresenta todos os cálculos efetuados. (1 Valor)

$$y = 2 * 3 + -1 * 3 = 3$$

## Problema 8 (Perguntas de Desenvolvimento - 3 Valores)

8.1) Supõe que estás a utilizar um SVM para classificar dados que claramente não são linearmente separáveis. De forma muito breve responde às seguintes questões.

8.1.1) O que aconteceria se usasses um *kernel* linear? (0.5 Valores)

Modelo iria ter uma má performance porque os dados não são linearmente separáveis.

8.1.2) Qual solução poderias aplicar para melhorar a performance do modelo? (0.5 Valores)

Usar um kernel não linear. RFB, polinomial, etc

8.2) Qual a principal diferença entre *bagging* e *boosting*? (1 Valor)

Bagging treina modelos em paralelo com dados amostrados aleatoriamente; boosting treina modelos sequencialmente, corrigindo os erros do anterior.

8.3) Por que se deve usar um conjunto de validação separado do teste final durante a otimização de hiperparâmetros?

Para evitar que o desempenho no teste final esteja inflacionado por overfitting causado pela escolha de hiperparâmetros.