

---

## Aprendizagem Automática

FICHA N. 2

ENUNCIADO

---

Nome: Duarte Alexandre Pedro Gonçalves

Número: A46484

---

ATENÇÃO: Fixa de respostas múltiplas. Só uma única resposta em cada alínea está correta. Cada alínea vale 2 valores. Respostas erradas descontam 0.5 valores.

---

1. No ficheiro `A46484_Q001_data.p`, encontram-se dois conjuntos de dados bidimensionais, um para treino e outro para teste, divididos em 5 classes (índices de 0 a 4). Há quatro variáveis num dicionário: as chaves `trueClassTrain` e `trueClassTest` contêm os índices das classes dos dados de treino e de teste, enquanto as chaves `Xtrain` e `Xtest` contêm os dados. Classifique o conjunto de teste com o classificador  $k$  vizinhos mais próximos ( $k$ NN), com  $k = 13$ , e utilize a métrica de distância,  $\mathcal{D}_\diamond(\mathbf{x}, \mathbf{y})$ , entre dois vetores bidimensionais  $\mathbf{x}$  e  $\mathbf{y}$  (nota: caso haja empate, escolha a classe com menor índice).

$$\mathcal{D}_\diamond(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^\top \begin{bmatrix} +1 & -2 \\ +1 & +1 \end{bmatrix} (\mathbf{x} - \mathbf{y})$$

- (a) Para esta alínea, use valores inteiros.

- i. O número de pontos da classe  $\varpi_2$  classificados na classe  $\varpi_2$  é 146.
- ii. O número de pontos da classe  $\varpi_4$  classificados na classe  $\varpi_3$  é 17.
- iii. Todas as respostas anteriores.
- iv. Nenhuma das respostas anteriores.

- (b) Para esta alínea, use valores inteiros.

- i. O número total de erros é de 77.
- ii. O número de acertos na classe  $\varpi_3$  é de 179.
- iii. Todas as respostas anteriores.
- iv. Nenhuma das respostas anteriores.

2. Considere o conjunto de 9 vetores bi-dimensionais, divididos em duas classes  $\Omega = \{\varpi_0, \varpi_1\}$ , representados na matriz  $\mathbf{X} = \begin{bmatrix} -2 & -1 & -2 & -1 & 5 & 5 & 5 & 2 & 2 \\ 5 & 7 & 5 & 4 & 5 & 4 & 3 & 4 & 3 \end{bmatrix}$  (os 4 primeiros vetores do conjunto pertencem à classe  $\varpi_0$ ).

- (a) Considere que o conjunto  $\mathbf{X}$  foi classificado usando o classificador de distância ao centroide com a métrica de distância euclidiana. Considere ainda que  $\varpi_0$  corresponde à classe dos negativos e  $\varpi_1$  à dos positivos. Para esta alínea, use valores inteiros.
- i. O número de Verdadeiros Positivos é: 5.
  - ii. O número de Verdadeiros Negativos é: 3.
  - iii. Todas as respostas anteriores.

- iv. Nenhuma das respostas anteriores.
  - (b) Considere que se pretende classificar novos vetores com o classificador de distância ao centroide usando a métrica de distância de Manhattan.
    - i. O vetor  $\mathbf{x} = [-4, -1]^T$  é classificado na classe  $\varpi_0$ .
    - ii. O vetor  $\mathbf{x} = [-3, 4]^T$  é classificado na classe  $\varpi_1$ .
    - iii. Todas as respostas anteriores.
    - iv. Nenhuma das respostas anteriores.
  - (c) Considere que se pretende classificar novos vetores com o classificador de distância ao centroide usando a métrica de distância de Mahalanobis.
    - i. O vetor  $\mathbf{x} = [3, -1]^T$  é classificado na classe  $\varpi_1$ .
    - ii. O vetor  $\mathbf{x} = [-1, -1]^T$  é classificado na classe  $\varpi_1$ .
    - iii. Todas as respostas anteriores.
    - iv. Nenhuma das respostas anteriores.
  - (d) Considere que o conjunto  $\mathbf{X}$  foi classificado usando o classificador de distância ao centroide com a métrica de distância de cosseno. Considere ainda que  $\varpi_0$  corresponde à classe dos negativos e  $\varpi_1$  à dos positivos. Para esta alínea, use valores inteiros.
    - i. O número de acertos na classe dos positivos é de 3.
    - ii. O número de acertos na classe dos negativos é de 4.
    - iii. Todas as respostas anteriores.
    - iv. Nenhuma das respostas anteriores.
  - (e) Para esta alínea, arredonde os valores pedidos a 2 casas decimais.
    - i. A distância de Manhattan entre os vetores de média das duas classes é: 1.75.
    - ii. A distância de cosseno entre os vetores de média das duas classes é: 1.77.
    - iii. Todas as respostas anteriores.
    - iv. Nenhuma das respostas anteriores.
3. No ficheiro `A46484_Q003_data.p` estão os resultados de um classificador, num problema de duas classes (negativos e positivos). Há duas variáveis num dicionário: a chave `trueClass` contém os índices da verdadeira classe, enquanto a chave `estClass` contém o resultado da classificação (ambas as variáveis são `np.array`s com valores inteiros 0-negativos, 1-positivos).
- (a) Para esta alínea, arredonde os valores pedidos a 1 casa decimal.
    - i. A percentagem de acertos da classe dos negativos é de 89.1%.
    - ii. A percentagem de exemplos da classe dos negativos é de 80.0%.
    - iii. Todas as respostas anteriores.
    - iv. Nenhuma das respostas anteriores.
  - (b) Para esta alínea, arredonde os valores pedidos a 3 casas decimais.
    - i. O F-score é: 1.000.
    - ii. A taxa de re-chamada (recall) é: 0.906.
    - iii. Todas as respostas anteriores.

- iv. Nenhuma das respostas anteriores.
- (c) Para esta alínea, use valores inteiros.
  - i. O número de Falsos Positivos é: 223.
  - ii. O número de Verdadeiros Positivos é: 290.
  - iii. Todas as respostas anteriores.
  - iv. Nenhuma das respostas anteriores.