
Aprendizagem Automática

FICHA N. 2

ENUNCIADO

Nome: Luis Carlos Semedo Da Fonseca

Número: A45125

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 `A45125_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 $k = 23$, 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 acertos na classe ϖ_3 é de 137.
- ii. O número de acertos na classe ϖ_4 é de 206.
- iii. Todas as respostas anteriores.
- iv. Nenhuma das respostas anteriores.

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

- i. O número de pontos da classe ϖ_0 classificados na classe ϖ_1 é 7.
- ii. O número de pontos da classe ϖ_3 classificados na classe ϖ_3 é 137.
- iii. Todas as respostas anteriores.
- iv. Nenhuma das respostas anteriores.

2. Considere o conjunto de 10 vetores bi-dimensionais, divididos em duas classes $\Omega = \{\varpi_0, \varpi_1\}$, representados na matriz $\mathbf{X} = \begin{bmatrix} 1 & -1 & 1 & -2 & -1 & 1 & -2 & 2 & -1 & 1 \\ 0 & -1 & -3 & 0 & -2 & -1 & 1 & 2 & 0 & -1 \end{bmatrix}$ (os 5 primeiros vetores do conjunto pertencem à classe ϖ_0).

- (a) 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 ϖ_0 corresponde à classe dos negativos e ϖ_1 à dos positivos. Para esta alínea, use valores inteiros.

- i. O número de acertos na classe dos positivos é de 1.
- ii. O número total de erros é 2.
- 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 $\mathbf{x} = [-4, 2]^T$ é classificado na classe ϖ_1 .
 - ii. O vetor $\mathbf{x} = [1, 2]^T$ é classificado na classe ϖ_1 .
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
 - (c) 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: 0.00.
 - ii. A distância de cosseno entre os vetores de média das duas classes é: 1.89.
 - 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 euclidiana. Considere ainda que ϖ_0 corresponde à classe dos negativos e ϖ_1 à dos positivos. Para esta alínea, use valores inteiros.
 - i. O número de Verdadeiros Positivos é: 3.
 - ii. O número de Verdadeiros Negativos é: 5.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
 - (e) Considere que se pretende classificar novos vetores com o classificador de distância ao centroide usando a métrica de distância de Mahalanobis.
 - i. O vetor $\mathbf{x} = [-4, -2]^T$ é classificado na classe ϖ_1 .
 - ii. O vetor $\mathbf{x} = [-3, 2]^T$ é classificado na classe ϖ_1 .
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
3. No ficheiro `A45125_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 80.3%.
 - ii. A percentagem de exemplos da classe dos negativos é de 89.0%.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
 - (b) Para esta alínea, arredonde os valores pedidos a 3 casas decimais.
 - i. A taxa de falso positivos (falsos alarmes) é: 0.197.
 - ii. O F-score é: 0.225.
 - iii. Todas as respostas anteriores.

- iv. Nenhuma das respostas anteriores.
- (c) Para esta alínea, use valores inteiros.
 - i. O número de Falsos Positivos é: 189.
 - ii. O número de Verdadeiros Positivos é: 186.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.