
Aprendizagem Automática

FICHA N. 4

ENUNCIADO

Nome: Diogo Sancho Pires Lobo

Número: A48168

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 `A48168_Q001_data.p`, encontra-se um conjunto de dados bidimensionais divididos em duas classes $\Omega = \{\varpi_0, \varpi_1\}$ (negativos e positivos). Há duas variáveis num dicionário: \mathbf{X} é uma matriz de dados, e \mathbf{y} é um array com as classes dos dados. Considere o seguinte modelo linear de classificação:

$\hat{y} = \mathbf{w}^\top \mathbf{x} = w_0 + w_1 x_1 + w_2 x_2$, com $\mathbf{x} \in \varpi_1$ para $\hat{y} \geq 0$, e para $\mathbf{w} = [w_0, w_1, w_2]^\top$ e $\mathbf{x} = [1, x_1, x_2]^\top$.

Considere ainda que o vetor \mathbf{w}_{MSE} é o vetor de pesos que minimiza o erro quadrático médio deste conjunto: $\mathcal{E} = \frac{1}{N} \sum_{n=1}^N (y[n] - \hat{y}[n])^2$, onde N é o número total de pontos, e $n = 1, \dots, N$. As saídas desejadas são: $y[n] = -1$ para $\mathbf{x}[n] \in \varpi_0$ e $y[n] = +1$ para $\mathbf{x}[n] \in \varpi_1$.

- (a) Considere o classificador com o seguinte vetor de pesos $\mathbf{w} = [0.00, 0.99, 0.15]$.

- O número de erros na classe ϖ_0 é de 21.
- O número de acertos na classe ϖ_1 é de 1652.
- Todas as respostas anteriores.
- Nenhuma das respostas anteriores.

- (b) Considere o classificador com vetor de pesos, \mathbf{w}_{MSE} , que minimiza o erro quadrático médio do conjunto.

- O vetor que minimiza o erro quadrático médio é $\mathbf{w}_{\text{MSE}} = [1.254, 1.208, -0.923]$.
- O erro quadrático médio é igual a 1.084.
- Todas as respostas anteriores.
- Nenhuma das respostas anteriores.

2. Considere o conjunto “diabetes” disponível em `sklearn.datasets` (usar a função `load_diabetes()`). Pretende-se estimar e avaliar modelos de regressão polinomial com os dados deste conjunto: use as primeiras 316 amostras para treino e as restantes para teste.

- (a) Considere o modelo de regressão polinomial de 2ª ordem que minimiza o erro quadrático médio no conjunto de treino.

- No conjunto de treino, o erro absoluto médio é igual a 37.97.
- No conjunto de treino, o coeficiente de determinação, R^2 , é igual a 0.79.

- iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
- (b) Considere que utiliza função `Lasso` (sub-módulo `linear_model` do `sklearn`) para uma regressão polinomial de 4ª ordem dos dados de treino. Instancie o regressor somente com os seguintes parâmetros:
- ```
Lasso(random_state=42, alpha=0.01)
```
- i. No conjunto de teste, o coeficiente de determinação,  $R^2$ , é igual a 0.526.
  - ii. Excluindo  $w_0$ , o número de coeficientes do polinómio diferentes de zero é 32.
  - iii. Todas as respostas anteriores.
  - iv. Nenhuma das respostas anteriores.
3. No ficheiro `A48168_Q003_data.p` encontra-se disponível uma variável independente  $x$  e uma variável dependente  $y$ . Pretende-se estimar a variável  $y$  através de uma regressão polinomial da variável  $x$ , minimizando o erro quadrático médio.
- (a) Considere o conjunto de treino composto pelo “fold” 1, e o conjunto de teste composto pelo “fold” 0. Considere ainda que, através da minimização do erro quadrático médio do conjunto de treino, se estimou um modelo regressão polinomial de 3ª ordem:  $\hat{y} = w_0 + w_1x + w_2x^2 + w_3x^3$ .
- i. No conjunto de teste, o erro quadrático médio é igual a 0.91.
  - ii. No conjunto de treino, o coeficiente de determinação,  $R^2$ , é igual a 0.78.
  - iii. Todas as respostas anteriores.
  - iv. Nenhuma das respostas anteriores.
- (b) Considere o conjunto de treino composto pelo “fold” 1, e o conjunto de teste composto pelo “fold” 0. Considere ainda que, através da minimização do erro quadrático médio do conjunto de treino, se estimou um modelo regressão polinomial de 4ª ordem:  $\hat{y} = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4$ .
- i. No conjunto de teste, o erro absoluto médio é igual a 0.66.
  - ii. Arredondando a zero casas decimais, o valor de  $w_0$  é 2.
  - iii. Todas as respostas anteriores.
  - iv. Nenhuma das respostas anteriores.