

Questionário 1

Questão 1

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Quando padronizamos os dados?

- ☒ a. quando os dados possuem escalas diferentes.
- ☐ b. quando os dados são correlacionados.
- ☐ c. quando os dados são independentes.
- ☐ d. nenhuma das outras opções.
- ☐ e. quando os dados são ordinais.

Questão 2

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Usando a distância interquantil e o critério de que uma observação é um outlier se ao menos uma das variáveis está fora dos limites máximos do boxplot, quantos outliers há nos dados da base Vehicles? (Veja o exemplo da aula 2).

- ☐ a. 15
- ☐ b. 0
- ☐ c. 10
- ☐ d. 20
- ☐ e. 4

33

Questão 3

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Com relação ao modelo de regressão, assinale a alternativa correta.

- ☒ a. O resíduo é definido pela diferença entre o valor real (do conjunto de treinamento) e o valor predito.
- ☐ b. Não é possível usar o teste de hipóteses no modelo de regressão. ✗
- ☐ c. A acurácia é uma medida que pode ser usada no modelo de regressão. ✗
- ☐ d. A medida R^2 é uma medida de associação não-linear. ✗
- ☐ e. Valores de R^2 próximo de 1 indicam que uma grande proporção da variabilidade dos dados são explicadas pelo modelo de regressão. ✗

Questão 4

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Assinale a alternativa INCORRETA:

- ☐ a. A regularização não deve ser feita ajustando o coeficiente R^2 .
- ☐ b. Altos valores dos coeficientes de regressão são indicativos de overfitting.
- ☐ c. Ridge regression pode não é usada para selecionar atributos.
- ☐ d. Lasso mantém os valores dos parâmetros pequenos e diferentes de zero.
- ☐ e. Um modelo de regressão linear é linear nos parâmetros e não nas variáveis, que podem ser não lineares (ex. X^2 , $\sin(X)$, $\ln(X)$).

Questão 5

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Considere o conjunto de dados: $X = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]$, $Y = [0, 3, 4, 5, 10, 8, 12, 15, 15, 19, 22, 21, 26, 28, 27, 29]$. Calcule os coeficiente b_0 e b_1 na regressão linear, como vimos na aula. Ou seja, no modelo: $Y \approx b_0 + b_1 X$ ajustado a esses dados. (Veja o exemplo no notebook. Escolha o valor mais próximo das alternativas).

```
import matplotlib.pyplot as plt

import numpy as np

from statistics import variance

import math

x = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15])
y = np.array([0, 3, 4, 5, 10, 8, 12, 15, 15, 19, 22, 21, 26, 28, 27, 29])

plt.plot(x, y, 'bo')

plt.xlabel("x", fontsize = 15)

plt.ylabel("y", fontsize = 15)

plt.show(True)
```

- ☒ a. $b_0: 0.16, b_1: 2.01$
- ☐ b. $b_0: 2.5, b_1: 2.5$
- ☐ c. $b_0: 0.50, b_1: 10.0$
- ☐ d. $b_0: 2.0, b_1: 2.0$
- ☐ e. $b_0: 1.0, b_1: 0.1$

Questão 6

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Ainda considerando os dados anteriores, qual o valor de R^2 ? (considere o valor mais próximo).

$X = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]$, $Y = [0, 3, 4, 5, 10, 8, 12, 15, 15, 19, 22, 21, 26, 28, 27, 29]$.

- ☐ a. 0.45
- ☐ b. 0.71
- ☒ c. 0.98
- ☐ d. 0.89
- ☐ e. 0.65

Questão 7

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Com relação ao algoritmo k-vizinhos, assinale a alternativa correta:

- ☐ a. No método k-vizinhos nunca ocorre overfitting.
- ☐ b. O número de vizinhos k é um parâmetro do modelo.
- ☐ c. É um classificador linear.
- ☐ d. O algoritmo não "aprende" um modelo, apenas memoriza objetos de treinamento.
- ☐ e. O número de vizinhos k não influencia na classificação. **Y**

Questão 8

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Com relação ao algoritmo Naive Bayes:

- ☐ a. Todas as probabilidades exigidas não podem ser calculadas dos dados de treinamento em uma passagem.
- ☐ b. Só pode ser usado para dados numéricos.
- ☐ c. Assume que as variáveis que descreve os atributos são independentes.
- ☐ d. Assume que os atributos são correlacionados.
- ☐ e. Não é robusto a ruídos e atributos irrelevantes.

Questão 9

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Considere a base de dados 'vertebralcolumn-2C'. Calcule a acurácia na classificação usando Naive Bayes. Considere 20% dos dados no conjunto de teste e o método hold-out. Arredonde para uma casa decimal.

- ☐ a. 0.3
- ☐ b. 1.0
- ☒ c. 0.8
- ☐ d. 0.6
- ☐ e. 0.5

Questão 10

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Considere os dados gerados com o código a seguir. Usando $h=1$ no algoritmo Bayesiano não-paramétrico, com 20% dos dados no conjunto de teste, qual a acurácia? (Selecione o valor mais próximo).

(Use `x_train, x_test, y_train, y_test = train_test_split(X, y, train_size = p, random_state = 42)`).

```
import sklearn.datasets as skdata
from matplotlib import pyplot
import numpy as np
import matplotlib.pyplot as plt
np.random.seed(42)
# Gera os dados em duas dimensões
n_samples = 100 # número de observações
# centro dos grupos
centers = [(0, 0), (2, 2)]
X, y = skdata.make_blobs(n_samples=200, n_features=2, cluster_std=1.0, centers=centers, shuffle=False, random_state=42)
plt.figure(figsize=(6,4))
plt.scatter(X[:,0], X[:,1], c=y, cmap='viridis', s=50, alpha=0.7)
plt.show(True)
```

- ☐ a. 0.64
- ☒ b. 0.95
- ☐ c. 0.75
- ☐ d. 0.56
- ☐ e. 0.83

Questão 11

Ainda não respondida

Vale 1,00 ponto(s).

🚩 Marcar questão

⚙ Editar questão

Considere a base winequality-red. Realize a classificação usando o classificador Bayesiano paramétrico (que assume que os dados seguem uma normal multivariada, como feito na aula). Qual o valor da acurácia (escolha o valor mais próximo).

- ☐ a. 0.8
- ☐ b. 0.6
- ☒ c. 0.4
- ☐ d. 0.7
- ☐ e. 0.9