



**Iniciado em** Sunday, 24 Mar 2024, 18:53

**Estado** Finalizada

**Concluída em** Sunday, 24 Mar 2024, 18:53

**Tempo  
empregado** 29 segundos

**Avaliar** 10,00 de um máximo de 10,00(100%)

Questão 1

Correto

Atingiu 2,00 de 2,00

Comparando os resultados dos modelos KNN treinados com os dados normalizados e não normalizados, utilizando valores de  $k = \{1, 3, 5, 7\}$ , podemos concluir que (marque todas as verdadeiras):

- ☐ a. A normalização possui pouco impacto na acurácia do modelo KNN
- ☐ b. O desempenho do modelo sempre aumenta ao normalizarmos os atributos
- ☐ c. O desempenho do modelo sempre diminui ao normalizarmos os atributos
- ☒ d. O uso de normalização causa alterações na fronteira de decisão de modelo, podendo impactar na sua acurácia
- ☒ e. Aplicando a normalização de dados, a acurácia aumentou para  $k=5$  e  $k=7$
- ☐ f. Aplicando a normalização de dados, a acurácia diminuiu ou manteve-se igual para  $k=5$  e  $k=7$




## Questão 2

Correto

Atingiu 2,00 de 2,00

Considerando o modelo treinado com  $k=5$  utilizando dados não normalizados, verifique quem são os 5 vizinhos mais próximos da instância de teste **N1**. Informe os IDs dos 5 vizinhos mais próximos, seguindo o formato "T1;T2;T3;T4;T5": separados por ponto e vírgula, sem espaços em branco, e em ordem crescente de ID.


Resposta: 

## Questão 3

Correto

Atingiu 2,00 de 2,00

Considerando o modelo treinado com  $k=5$  utilizando **dados não normalizados**, verifique quem são os 5 vizinhos mais próximos da instância de teste **N1** e como eles estão dispostos no espaço de entrada. Isto é, a posição destes vizinhos em relação à instância N1 e às coordenadas x (Total Sulfur Dioxide) e y (Citric Acid). Podemos afirmar que:

- ☐ a. Os k-vizinhos mais próximos tendem a estar dispostos ao redor de N1, em uma área que se aproxima do formato de uma elipse.
- ☐ b. Os k-vizinhos mais próximos tendem a estar distribuídos ao longo do eixo que possui intervalo de valores em escalas maiores
- ☒ c. A distribuição dos k-vizinhos mais próximos tende a estar alinhada com o eixo que possui intervalo de valores em escalas menores 


## Questão 4

Correto

Atingiu 2,00 de 2,00

Considere M2 o modelo treinado com o conjunto de dados contendo 2 atributos e M11 o modelo treinado com o conjunto de dados contendo 11 atributos. Em ambos os casos, foram utilizados dados normalizados,  $k=5$ , e a **distância Euclidiana**.

Analisando o efeito da perturbação do atributo "citric acid" para a instância de teste **N4** na saída destes modelos que utilizam como base a distância Euclidiana, é incorreto afirmar que:


- ☐ a. Ao utilizar a distância Euclidiana, o aumento da dimensionalidade dos dados (em M11) torna o algoritmo menos invariante a alterações pontuais em atributos, mesmo quando ocorrem alterações significativas (por exemplo, alterar o valor do atributo citric acid de 1.0 para 0.3).
- ☐ b. As distâncias entre instâncias (de treino e teste) aumenta naturalmente conforme aumentamos o número de dimensões (isto é, atributos) do problema, e tornam-se menos distintas entre si.
- ☒ c. O KNN é um algoritmo insensível a ruídos ou perturbações nos dados. 
- ☐ d. A perturbação no valor de *citric acid* para a instância de teste N4 não causou alterações nos vizinhos mais próximos encontrados em M11.
- ☐ e. A perturbação no valor de *citric acid* para a instância de teste N4 causou alterações substanciais nos vizinhos mais próximos encontrados em M2.
- ☐ f. Dados com alta dimensionalidade quebram a suposição básica do KNN de que todos os vizinhos semelhantes estão mais próximos do que quaisquer outros pontos de dados no conjunto de treinamento.

## Questão 5

Correto

Atingiu 2,00 de 2,00

Sobre o método Holdout estudado em aula (denominado versão de duas vias/2-way, conforme discutiremos mais adiante na disciplina), é **incorreto** afirmar que:

- ☐ a. Ele divide os dados aleatoriamente em dois conjuntos disjuntos, treinamento e teste.
- ☐ b. A proporção de dados utilizada para treinamento e teste deve ser definida manualmente pelo programador, e depende de fatores como a quantidade de instâncias disponíveis, número de instâncias por classe, etc.
- ☐ c. O conjunto de teste é usado para estimar o desempenho do modelo treinado. Se satisfatório, é recomendado que o modelo final seja treinado com todo o conjunto de dados.
- ☒ d. Ao ser utilizado, o método holdout garante que toda instância será usada uma vez para teste do modelo treinado. 
- ☐ e. O uso de métodos de avaliação de modelos preditivos, como o holdout, ajuda a evitar a produção de modelos com overfitting.

