



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
PRÓ-REITORIA DE GESTÃO DE PESSOAS
COORDENADORIA DE CONCURSOS
Campus Universitário – Praça Cívica - Natal/RN - 59078-970
Fone (84) 3342 2296 Fax (84) 3215 3270
www.progesp.ufrn.br | concursos@reitoria.ufrn.br



SELEÇÃO SIMPLIFICADA PARA CONTRATAÇÃO TEMPORÁRIA – PROFESSOR SUBSTITUTO

EDITAL Nº 076/2023

ÁREA: Aprendizado de Máquina e Ciência de Dados

PROVA ESCRITA

INSTRUÇÕES

1	Na parte inferior desta capa, escreva o seu nome completo de forma legível.
2	A pessoa candidata deverá guardar, em embalagem porta-objetos fornecida pela equipe de aplicação ou comissão de seleção, telefone celular desligado ou quaisquer equipamentos eletrônicos, também desligados, ainda que os sinais de alarme estejam nos modos de vibração e silencioso, garantindo que nenhum som seja emitido, inclusive do despertador, caso esteja ativado, sob pena de ser eliminado do processo seletivo.
3	A embalagem porta-objetos devidamente lacrada e identificada pela pessoa candidata, assim como bolsas e sacolas, deverão ser mantidos embaixo da carteira até o término de suas provas. A embalagem somente poderá ser deslacrada fora do local de realização das provas.
4	Este Caderno contém 20 questões de múltipla escolha . Verifique se ele está completo. Se estiver incompleto ou contiver imperfeição gráfica que impeça a leitura, solicite imediatamente ao Fiscal que o substitua.
5	A Prova Objetiva (questões de múltipla escolha) vale 10,0 pontos e cada uma de suas questões tem o mesmo valor.
6	Cada questão de múltipla escolha apresenta quatro opções de resposta, das quais apenas uma é correta.
7	Somente é permitido o uso de caneta esferográfica de tinta preta ou azul. Não será considerada a questão marcada com lápis grafite.
8	Utilize o verso das páginas deste Caderno para rascunhos.
9	Você dispõe de, no máximo, quatro horas para responder às questões de múltipla escolha.
10	Será considerada como resposta definitiva a cada questão o preenchimento do Gabarito Final , na última folha do caderno de provas.
11	Antes de se retirar definitivamente da sala, devolva ao Fiscal este Caderno de Provas.

NOME COMPLETO DA PESSOA CANDIDATA

--

QUESTÕES DE MÚLTIPLA ESCOLHA

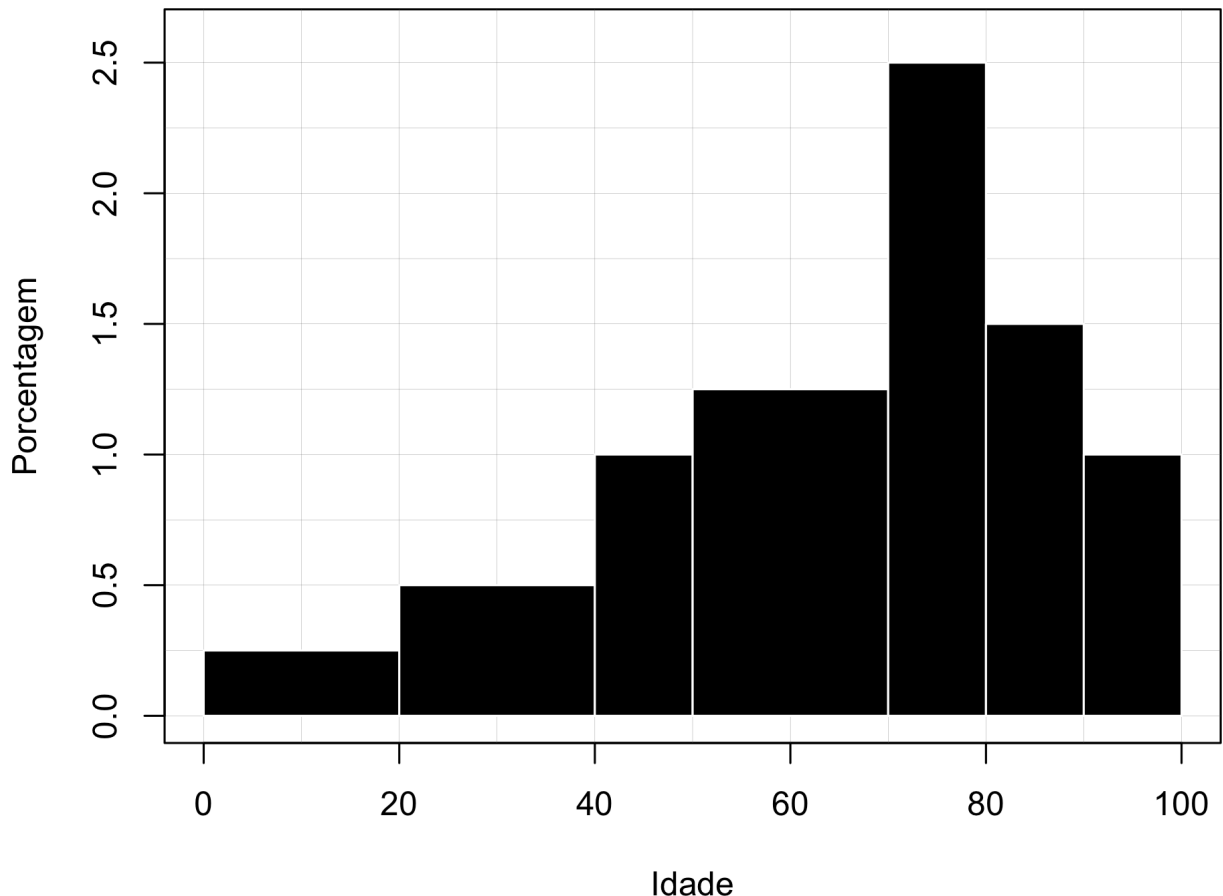
QUESTÃO 1) Suponha que estamos interessados em realizar uma tarefa de aprendizado de máquina não-supervisionada. É recomendável um método de agrupamento de dados baseado na Distância de Manhattan em vez da Distância Euclidiana exceto quando

- a) os dados apresentam muitos outliers.
- b) o domínio do problema possui alta dimensionalidade.
- c) os dados se comportam de maneira aproximadamente normal.
- d) o problema apresenta dados inteiros.

QUESTÃO 2) Em uma tarefa de classificação binária, suponha que não é possível determinar uma curva da forma $y = \beta_0 + \beta_1 x$ que separe de maneira adequada as duas classes. Nesse caso, dois algoritmos que podem ser testados a fim de se obterem boas métricas de avaliação são

- a) SVM com kernel radial e Random Forest.
- b) Regressão Logística e Random Forest.
- c) Regressão Logística e Naïve Bayes.
- d) Random Forest e Naïve Bayes.

QUESTÃO 3) O histograma da Figura abaixo exhibe as idades e as frequências de morte estimadas para uma determinada população. Este resultado foi obtido através de uma amostra de 2376 pessoas. Para facilitar seu raciocínio, assuma que a distribuição das idades dentro de cada intervalo é uniforme.



Considere as seguintes afirmações a respeito deste histograma:

- I- A mediana da amostra é 50 anos.

II- É possível determinar com precisão a idade, em anos, da pessoa mais velha a falecer nesta amostra.

III- Se cada pessoa da amostra tivesse vivido um ano a mais, a média da amostra seria alterada, mas não é possível determinar a magnitude desta alteração.

Baseando-se nessas afirmações, escolha a alternativa correta:

- a) Apenas a afirmação I está correta.
- b) Apenas as afirmações I e II estão corretas.
- c) Apenas as afirmações II e III estão corretas.
- d) Nenhuma afirmação está correta.

QUESTÃO 4) Considerando que as siglas RMSE e MAE se referem a Raiz do Erro Quadrático Médio e Erro Médio Absoluto, respectivamente, é correto afirmar que

- a) o RMSE é mais sensível a outliers do que o MAE.
- b) ambas são métricas indicadas para verificar o desempenho de um algoritmo de classificação.
- c) não é possível determinar um valor teórico mínimo para ambas as métricas.
- d) se busca maximizar os valores de RMSE e MAE em tarefas de regressão.

QUESTÃO 5) Em relação à Análise de Componentes Principais (ACP), avalie se as afirmativas a seguir são falsas (F) ou verdadeiras (V). Considere que a ACP será aplicada em uma matriz X de dimensão $n \times p$ criada a partir de p variáveis com n observações cada uma.

- I. Um dos objetivos da ACP é obter, a partir de combinações lineares das variáveis originais, novas variáveis não-correlacionadas entre si.
- II. A ACP é uma técnica de redução de dimensionalidade que gera seus melhores resultados quando as variáveis analisadas apresentam relações não-lineares entre si.
- III. Caso estejam em escalas diferentes, é fundamental transformar as variáveis da matriz antes de calcular a ACP, de modo que todas tenham média 0 e desvio padrão 1.

As afirmativas acima são, respectivamente,

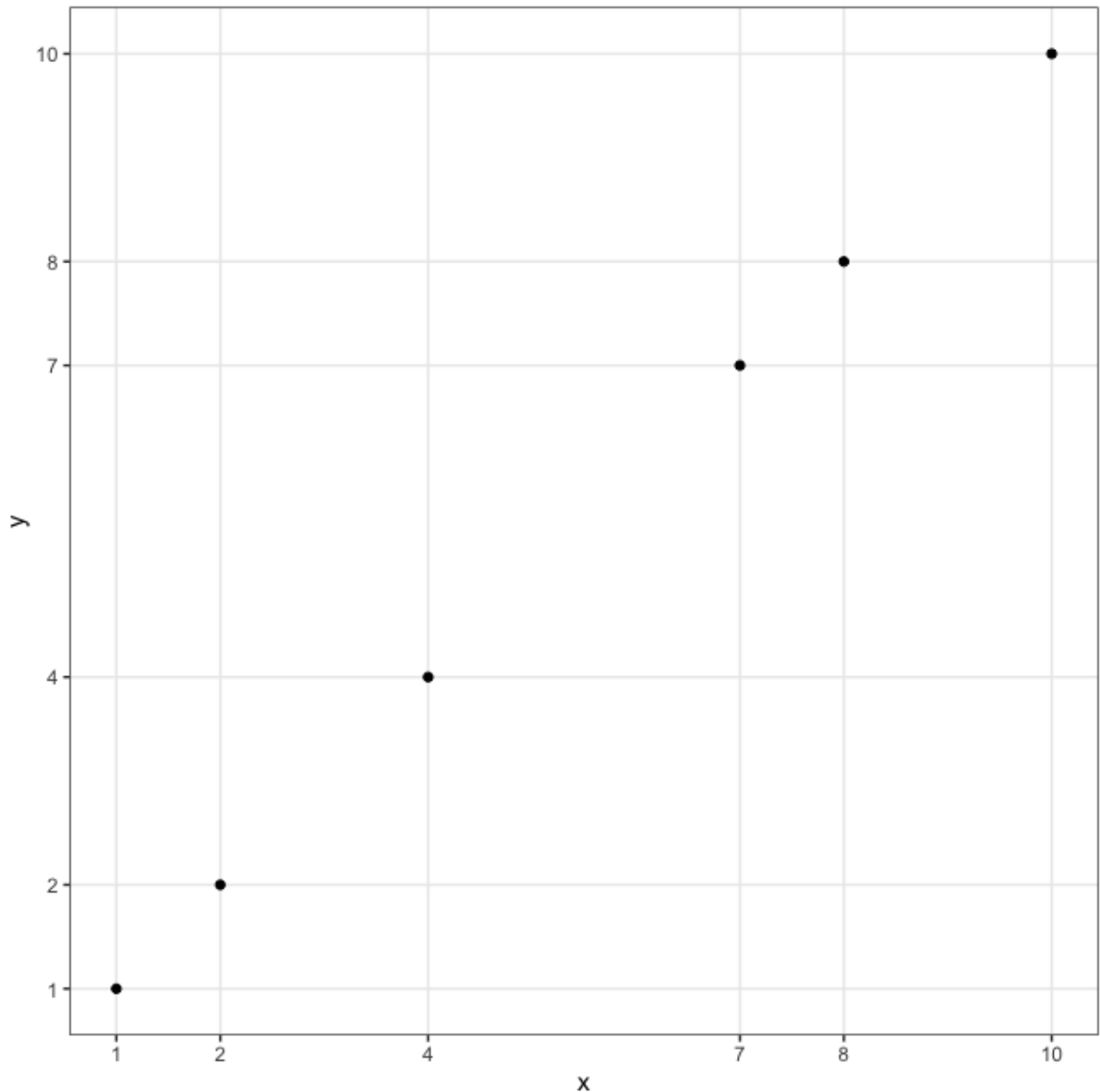
- a) V, V, V
- b) V, F, V
- c) V, F, F
- d) F, F, V

QUESTÃO 6) Quais das estatísticas a seguir podem ser deduzidas a partir de um boxplot?

- a) Média, Variância e Desvio Padrão
- b) Média, Mediana e Máximo
- c) Mediana, Mínimo e Variância
- d) Mediana, Desvio Interquartilico e Máximo

QUESTÃO 7) Suponha que desejamos ajustar um modelo de regressão linear simples nos dados da figura abaixo. Assuma que representa os dados de entrada e os de saída. Considere três casos de ajuste do modelo:

- I. Todos os pontos serão utilizados no conjunto de treinamento.
- II. Apenas os quatro pontos mais à direita (4, 7, 8, 10 para os dados de entrada) serão utilizados no conjunto de treinamento.
- III. O método de validação cruzada *leave-one-out* será utilizado no conjunto de treinamento.



Os valores da métrica RMSE (Raiz do Erro Quadrático Médio, em inglês) para cada um dos casos considerados são, respectivamente

- a) 0, 0, 0
- b) 0, 0, >0
- b) >0, 0, >0
- d) >0, >0, 0

QUESTÃO 8) Dado um conjunto de treinamento $(x_i, y_i)_{i=1}^n$, onde $x_i \in X$ e $y_i \in Y$, o objetivo do aprendizado supervisionado é aprender $f: X \rightarrow Y$. Com base nessa definição, qual das seguintes afirmações é verdadeira?

- a) f é a função que desejamos aprender e Y são os exemplos de treinamento.
- b) X representa o espaço de entrada, que são as variáveis independentes ou features no conjunto de treinamento.
- c) y_i representa um par de entrada-saída no conjunto de treinamento.
- d) x_i e y_i representam, respectivamente, o espaço de entrada e o espaço de saída.

QUESTÃO 9) No aprendizado de máquina fracamente supervisionado, como é tratada a questão da precisão e confiabilidade dos rótulos?

- a) Os rótulos são considerados imprecisos, mas confiáveis.
- b) Os rótulos são considerados precisos, mas não confiáveis.
- c) Os rótulos são considerados imprecisos e não confiáveis.
- d) A precisão e confiabilidade dos rótulos não são consideradas.

QUESTÃO 10) *Active Learning* é uma estratégia de aprendizado de máquina onde um algoritmo seleciona os exemplos mais informativos a serem rotulados para treinamento. Qual das seguintes afirmações melhor descreve uma das principais vantagens do *Active Learning*?

- a) *Active Learning* elimina a necessidade de dados rotulados.
- b) *Active Learning* sempre melhora o desempenho do modelo, independentemente do domínio do problema.
- c) *Active Learning* pode ajudar a reduzir a quantidade de dados de treinamento necessários para atingir um determinado nível de desempenho.
- d) *Active Learning* permite que o modelo aprenda a rotular os dados automaticamente.

QUESTÃO 11) A função *softmax* é comumente usada em aprendizado de máquina, especialmente em problemas de classificação multiclasse. Qual das seguintes afirmações melhor descreve uma característica importante da função *softmax*?

- a) A função *softmax* produz um vetor que é frequentemente interpretado como a distribuição de probabilidade de um evento multiclasse.
- b) A função *softmax* pode aceitar valores negativos e positivos, mas sempre retorna valores negativos.
- c) A função *softmax* retorna sempre o maior valor da entrada.
- d) A função *softmax* é uma função linear.

QUESTÃO 12) As funções de ativação desempenham um papel crítico nas redes neurais, permitindo que elas aprendam e modelam relações complexas. Qual das seguintes afirmações é verdadeira sobre as funções de ativação?

- a) A função identidade permite que uma rede neural aprenda relações não-lineares.
- b) A função de ativação ReLU retorna valores no intervalo de -1 a 1.
- c) A função de ativação tangente hiperbólica sempre retorna 0.
- d) A função de ativação sigmóide pode sofrer do problema do desaparecimento de gradientes.

QUESTÃO 13) Na validação cruzada para seleção de modelos de aprendizado de máquina para problemas de classificação e regressão, durante a avaliação de cada modelo:

- a) Os conjuntos de treinamento e validação são fixos.
- b) Os conjuntos de validação devem se sobrepor entre *folds*.
- c) O número de *folds* deve ser igual ao tamanho do dataset.
- d) Os *folds* devem ser escolhidos aleatoriamente sem reposição.

QUESTÃO 14) - Qual das seguintes afirmações sobre a metodologia CRISP-DM é correta?

- a) A metodologia CRISP-DM é linear, cada fase deve ser concluída antes de passar para a próxima.
- b) A fase de Implementação na metodologia CRISP-DM envolve apenas a colocação do modelo em produção.
- c) A metodologia CRISP-DM é iterativa, permitindo que feedback e aprendizados de uma fase influenciem atividades anteriores.
- d) A avaliação dos modelos na metodologia CRISP-DM ocorre imediatamente após a fase de Entendimento dos Dados.

QUESTÃO 15) - Selecione a alternativa que contém, respectivamente, exemplos de metaheurísticas (i) populacionais, (ii) construtivas, (iii) bioinspiradas e (iv) de inteligência de enxame.

- a) GRASP, algoritmos genéticos, otimização por nuvem de partículas, colônias de abelhas artificiais.
- b) Algoritmos de estimação de distribuição, GRASP, algoritmos meméticos, otimização por nuvem de partículas.
- c) Algoritmos meméticos, otimização por colônias de formigas, algoritmos culturais e busca tabu.
- d) Busca em vizinhança variável, guloso iterativo, sistemas imunológicos artificiais e otimização por colônia de formigas.

QUESTÃO 16) - Selecione a alternativa que contém, respectivamente, exemplos de problemas de otimização que envolvem (i) roteamento, (ii) *assignment* e (iii) *scheduling*.

- a) Caixeiro viajante, alocação quadrática e mochila.
- b) Roteamento de veículos, caminho mais curto, *nurse rostering*.
- c) Roteamento de veículos, caminho mais curto, mochila.
- d) Caixeiro viajante, alocação quadrática e *flowshop*.

QUESTÃO 17) - Selecione a alternativa que contém, em ordem crescente de tamanho de espaço de busca, exemplos de operadores de vizinhança utilizados por técnicas de busca local em problemas cuja solução seja representada por uma permutação.

- a) 2-swap, 2-exchange, *insert*.
- b) 2-swap, *insert* e 2-exchange.
- c) 2-exchange, 2-swap, *insert*.
- d) 2-exchange, *insert*, 2-swap

QUESTÃO 18) - Selecione a alternativa que contém, respectivamente, exemplos de arquiteturas de redes neurais que oferecem como uma de suas principais vantagens serem fortemente paralelizáveis.

- a) CNN, Transformer e RNN.
- b) TCN, BERT e LSTM.
- c) CNN, Transformer e BERT.
- d) TCN, RNN e LSTM.

QUESTÃO 19) - Selecione a alternativa que contém, respectivamente, exemplos de algoritmos tradicionalmente associados ao aprendizado supervisionado, não-supervisionado e por reforço.

- a) CNN, kNN e RNN.
- b) CNN, mistura de modelos Gaussianos e AlphaGo.
- c) kNN, CNN e BERT.
- d) mistura de modelos Gaussianos, kNN e LSTM.

QUESTÃO 20) - Selecione a alternativa que contém, exclusivamente, técnicas apropriadas para a modelagem de séries temporais em um problema de aprendizado de máquina supervisionado.

- a) Suavização, dessazonalização e validação cruzada progressiva (*walk-forward*).
- b) Normalização, suavização e validação cruzada.
- c) Imputação, embaralhamento e *holdout*.
- d) Dessazonalização, embaralhamento e validação cruzada estratificada.

GABARITO FINAL			
01		11	
02		12	
03		13	
04		14	
05		15	
06		16	
07		17	
08		18	
09		19	
10		20	
Assinatura da Pessoa Candidata			