

PROVA D

Nome: _____

Respostas:

1	2	3	4	5	6	7	8	9	10	11
B	B	E	A	A	A	C	B	E	D	A

Questão 1

Considere as seguintes afirmativas sobre o modelo kmeans:

- i. É um modelo de aprendizado Supervisionado
- ii. Sua acuracidade é medida comparando-se as respostas do conjunto de testes
- iii. Os centróides são escolhidos inicialmente de modo a minimizar a distância de elementos do mesmo grupo

Estão corretas:

- a. Somente iii
- b. Nenhuma das alternativas**
- c. Somente i, ii
- d. Somente ii, iii
- e. Somente ii

Questão 2

Considere um conjunto de dados com 1.000.000 de elementos e a aplicação de um kmeans com k=3. O número de distâncias calculadas a cada passo do algoritmo é:

- a. Depende dos valores iniciais selecionados aleatoriamente
- b. 3.000.000**
- c. 9.000.000
- d. 1.000.000
- e. 333.333

Questão 3

Considere as seguintes afirmativas sobre o número ótimo de clusters no algoritmo kmeans:

- i. É obtido a partir do valor mínimo de inércia dos grupos
- ii. Deve também considerar o critério de tamanho dos grupos
- iii. É determinado automaticamente na aplicação do kmeans

Estão corretas:

- a. Somente ii, iii
- b. Somente iii
- c. Todas as alternativas
- d. Somente i, ii
- e. Somente ii**

Questão 4

Exercício. Considere unicamente a construção de Clusters pelo método Hierárquico. Qual dos itens abaixo não está associado à formação de Clusters diferentes pelo Método Hierárquico?

- a. Determinação dos Centróides dos Grupos**
- b. Forma de Construção (aglomerativa, divisiva)
- c. Função de similaridade (função distância)
- d. Função de Ligação (complete, single etc.)
- e. Normalização dos dados

Questão 5

Considere uma série de dados de empresas como os valores de ações, patrimônio e receita ao longo do tempo. Você deseja fazer uma predição de valor das empresas para um futuro próximo. Quais métodos são mais adequados para isso (assinale a melhor alternativa):

a. Métodos de Aprendizado Supervisionado e de Séries Temporais

- b. Redes neurais profundas
- c. Métodos de Aprendizado não Supervisionado
- d. Métodos de Regressão não Linear
- e. Métodos de Aprendizado Supervisionado

Questão 6

Considere das seguintes afirmativas sobre alguns dos modelos Estatísticos de Séries Temporais:

- i. O Modelo de Média Móvel Integrada Autoregressiva adiciona ao modelo ARMA uma outra variável externa, a variável 'integrada'.
- ii. Uma Série Temporal é melhor definida como uma variável $S_t = f(S_t, S_{t-1}, \dots, S_{t-2})$ que $S_t = f(t)$
- iii. No modelo ARIMA a parte integrada corresponde a etapa de diferenciação do modelo em que a Série S_t é transformada em uma Série de diferenças do tipo $V_t = S_t - S_{t-1}$ que pode ser aplicada mais que uma vez.

São corretas as afirmativas:

- a. Somente ii., iii.**
- b. Nenhuma
- c. Somente iii.
- d. Somente i., iii.
- e. Somente i., ii.

Questão 7

Considere das seguintes afirmativas sobre Séries Temporais:

- i. Séries não Estacionárias ou apresentam Tendência ou Sazonalidade, mas não os dois comportamentos ao mesmo tempo.
- ii. Em uma Série não Estacionária a variância dos dados sempre muda ao longo do tempo.
- iii. Séries não Estacionárias são sempre decompostas de forma Multiplicativa.

São corretas as afirmativas:

- a. Somente ii., iii.
- b. Somente i., iii.
- c. Nenhuma**
- d. Somente i., ii.
- e. Somente ii.

Questão 8

Considere o gráfico ACF (de Autocorrelação) e os seguintes modelos:

- i. Regressão Linear Múltipla.
 - ii. AR, Autoregressivo.
 - iii. ARIMA, Autoregressivo Integrado de Médias Móveis.
- O gráfico ACF é útil para avaliar que modelos?

- a. Somente i., iii.
- b. Somente iii.**
- c. Somente i., ii.
- d. Somente ii., iii.
- e. Nenhuma

Questão 9

Você trabalha em uma Grande empresa de Exploração de Petróleo que quer empregar modelos de Ciências de Dados para prever possíveis falhas em equipamentos e antecipar suas manutenções. Ela deseja avaliar modelos de Árvore de Decisão, knn e de Deep Learning. Você pode esperar que (assinale a melhor alternativa):

a. Os modelos de Deep Learning, terão um número de camadas superior a quantidade de níveis da Árvores de Decisão para obter a mesma acuracidade

- b. Os modelos de Deep Learning terão resultados melhores que o Knn quando comparados com o uso dos mesmos atributos numéricos
- c. Os modelos de Deep Learning não terão resultados melhores que as Árvores de Decisão quando empregados atributos categóricos
- d. Os modelos de Deep Learning terão certamente um resultado melhor que outros modelos
- e. Os modelos de Deep Learning poderão fazer uso de recursos avançados de processamento como uso de GPU e paralelismo**

Questão 10

Considere as seguintes afirmativas sobre Modelos Neurais/Deep Learning:

- i. Apresentam sempre um custo computacional menor para o treinamento que outros modelos
- ii. O aumento do número de camadas em geral aumenta a capacidade de representação dos dados (capacidade de solução de problemas mais complexos)
- iii. Uma Rede Neural em camadas pode ser empregada para tarefas supervisionadas tanto em problemas de Classificação como de Regressão

Estão corretas:

- a. Somente i, ii
- b. Somente iii
- c. Somente ii
- d. Somente ii, iii**
- e. Todas as alternativas

Questão 11

Você está aplicando um modelo de Árvore de Decisão para a seleção de novos fornecedores de peças para Tratores com base nos dados históricos dos fornecimentos passados. Os fornecedores serão classificados entre Rejeitados, Preferencial e Contingente (quando o Preferencial não pode atender a demanda). São 32 as características disponíveis dos fornecedores em sua base de dados. Na escolha do modelo: (assinale a melhor alternativa)

- a. Você usa diferentes atributos de entrada e parâmetros da Árvore de Decisão, como profundidade máxima e função de entropia ou gini, optando pelo modelo de maior acuracidade**
- b. Você emprega um número de níveis máximo igual ao número de atributos de entrada
- c. Você usa o critério de maior Entropia para escolha do modelo
- d. Você usa diferentes atributos de entrada e parâmetros da Árvore de Decisão, como profundidade máxima e função de entropia ou gini, optando pela Árvore maior (maior número de níveis) que irá proporcionar a maior acuracidade
- e. Você usa diferentes parâmetros da Árvore de Decisão, como profundidade máxima e função de entropia ou gini, optando pelo modelo de maior acuracidade, mas sem alterar os atributos de entrada que são dados