

PROVA B

Nome: \_\_\_\_\_

Respostas:

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

Questão 1

Considere as seguintes afirmativas sobre o modelo kmeans:

- É um modelo de aprendizado não Supervisionado
- Sendo um modelo de médias pode apenas empregar a distância Euclidiana
- O valor de k é inicialmente definido aleatoriamente

Estão corretas:

- Somente ii
- Somente i, ii
- Somente ii, iii
- Somente iii

**e. Somente i**

Questão 2

Considere as seguintes afirmativas sobre o algoritmo kmeans:

- O valor de k é determinado ao final do algoritmo
- O algoritmo termina quando o cálculo da soma da distância dos centróides é obtida e seu valor é máximo
- A cada iteração o novo valor de cada centróide é obtido a partir da média dos valores dos elementos mais próximos de cada respectivo centróide

Estão corretas:

- Nenhuma das alternativas
- Somente i, ii
- c. Somente iii**
- Somente ii
- Somente i, iii

Questão 3

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

- É obtido a regra do cotovelo a curva do gráfico de soma de distâncias
- A regra do cotovelo fornece um ponto de equilíbrio do número de grupos para minimizar as distâncias intra grupo e maximizar a distância entre grupos distintos
- O número de cluster é ótimo quando o tamanho dos grupos (número de elementos de cada clusters) é igual

Estão corretas:

- Somente ii
- b. Somente i, ii**
- Somente ii, iii
- Todas as alternativas
- Somente iii

Questão 4

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**
- Função de similaridade (função distância)
- Normalização dos dados
- Função de Ligação (complete, single etc.)

e. Forma de Construção (aglomerativa, divisiva)

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):

- Redes neurais profundas
- Métodos de Aprendizado não Supervisionado
- Métodos de Regressão não Linear
- Métodos de Aprendizado Supervisionado
- e. Métodos de Aprendizado Supervisionado e de Séries Temporais**

Questão 6

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

- O Modelo de Média Móvel Integrada Autoregressiva adiciona ao modelo ARMA uma outra variável externa, a variável 'integrada'.
- Uma Série Temporal é melhor definida como uma variável  $S_t = f(S_t, S_{t-1}, \dots, S_{t-2})$  que  $S_t = f(t)$
- 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:

- Nenhuma
- Somente iii.
- Somente i., ii.
- d. Somente ii., iii.**
- Somente i., iii.

Questão 7

Considere das seguintes afirmativas sobre Séries Temporais:

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

São corretas as afirmativas:

- Somente ii., iii.
- b. Nenhuma**
- Somente ii.
- Somente i., iii.
- Somente i., ii.

Questão 8

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

- MA, Médias Móveis.
- AR, Autoregressivo.
- ARIMA, Autoregressivo Integrado de Médias Móveis.

O gráfico PACF é útil para avaliar que modelos?

- Somente i., ii.
- Somente iii.
- c. Somente ii., iii.**
- Somente i., iii.
- Todos

Questão 9

Considere as seguintes afirmativas sobre Modelos Neurais/Deep Learning:

- i. As funções de ativação são modificadas a cada rodada do treinamento para minimizar o erro produzido
- ii. Para efetuar o treinamento os pesos são ajustados a cada rodada para produzir novas saídas que tentam ser mais próxima da saída desejada
- iii. Ao longo do treinamento o erro é constantemente decrescente

Estão corretas:

a. Todas as alternativas

**b. Somente ii**

c. Somente ii, iii

d. Somente i, ii

e. Somente iii

#### Questão 10

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 certamente um resultado melhor que outros modelos

b. Os modelos de Deep Learning não terão resultados melhores que as Árvores de Decisão quando empregados atributos categóricos

c. Os modelos de Deep Learning terão resultados melhores que o Knn quando comparados com o uso dos mesmos atributos numéricos

**d. Os modelos de Deep Learning poderão fazer uso de recursos avançados de processamento como uso de GPU e paralelismo**

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

#### Questão 11

Você está aplicando um modelo knn para prever que peças da linha de produção serão aprovadas e quais não. Com isso você espera poder aplicar o controle de qualidade, que é dispendioso, a um número reduzido de peças e reduzir assim o custo de produção. Na escolha do modelo: (assinale a melhor alternativa)

a. Você usa todos os atributos de entrada e varia os valores e k, optando pelo modelo com todos os atributos e o valor de k que retorna maior acuracidade

b. Você apenas emprega o menor valor de k

c. Você usa diferentes atributos de entrada e valores e k, optando pelo maior número de atributos e maior k

d. Você usa diferentes atributos de saída e valores e k, optando pelo modelo de maior acuracidade

**e. Você usa diferentes atributos de entrada e valores e k, optando pelo modelo de maior acuracidade**