



Prova da Residência - IMD/LIGA - Caderno 2

Professor(a): Comitê de Processo Seletivo Residência - IMD/LIGA

Os cartões de resposta devem ser lidos
usando o aplicativo android Multiprova Corretor.



2 provas

III

Matrícula

Multiprova - UFRN

CARTÃO TIPO A1

Código da prova

2	0													
Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15

Data: / /

Assinatura:

SIGA OS MODELOS AO LADO.

SE RASURAR O CARTÃO, PEÇA OUTRO.

0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J V

O LEITOR DE CARTÕES **NÃO ENTENDE**
OUTROS ESTILOS DE LETRAS E NÚMEROS.

O LEITOR DE CARTÕES NAO ENTENDE
OUTROS ESTILOS DE LETRAS E NÚMEROS.

0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J V

SIGA OS MODELOS AO LADO.
SE RASURAR O CARTÃO, PEÇA OUTRO.

[illegible]

Nome:

Matrícula:

Assinatura: _____

Questão 1

Valor da questão: 1,00

Um Data Warehouse pode ser subdividido em pequenas porções para atender um departamento da empresa, por exemplo. Essa subdivisão do Data Warehouse é conhecida como:

- a) Data Mart
- b) Database
- c) Data Center
- d) Data Store
- e) Data Driven

Questão 2

Valor da questão: 1,00

Qual das etapas a seguir é executada pelo cientista de dados após a aquisição dos dados?

- a) Replicação de dados
- b) Modelagem de dados
- c) Nenhuma delas
- d) Limpeza de dados
- e) Integração de dados

Questão 3

Valor da questão: 1,00

Qual dos seguintes termos se concentra na descoberta de propriedades (anteriormente) desconhecidas nos dados?

- a) Aprendizado de Máquina
- b) Disputa de dados
- c) Análise de Clustering
- d) Big Data
- e) Mineração de dados

Questão 4

Valor da questão: 1,00

Todos são tipos de Aprendizado de Máquina, exceto:

- a) Semissupervisionado
- b) Estocástico
- c) Não supervisionado
- d) Supervisionado
- e) Por Reforço

Questão 5

Valor da questão: 1,00

Qual dos métodos de agrupamento abaixo se baseia na variância e no desvio padrão dos dados?

- a) Hierarchical agglomerative
- b) Decision Tree
- c) DBScan
- d) Gaussian mixture model
- e) k-means

Questão 6

Valor da questão: 1,00

Dentre as seguintes opções, identifique aquela que está relacionada a redução da dimensionalidade.

- a) entropia
- b) acurácia
- c) colinearidade
- d) desempenho
- e) estocasticidade

Questão 7

Valor da questão: 1,00

Um conjunto de dados possui os seguintes valores numéricos para um determinado atributo: 1; 3; 7; 5; 2; 6; 10. Após processo de normalização pelo valor máximo, os valores desse atributo são:

- a) 0,1; 0,3; 0,7; 0,5; 0,2; 0,6; 1
- b) 10; 10; 10; 10; 10; 10; 10
- c) -9; -7; -5; -8; -4; 0
- d) 1; 1; 1; 1; 1; 1; 1
- e) 1; 3; 7; 5; 2; 6; 10

Questão 8

Valor da questão: 1,00

Um centro médico de pesquisas estuda uma doença rara em uma população. Em uma etapa do estudo, percebe-se que o conjunto de dados rotulados está desbalanceado. Sobre este conjunto de dados, é CORRETO afirmar que:

- a) o número de objetos (exemplos) para um dado rótulo é superior ao número de objetos para um outro rótulo
- b) alguns objetos (exemplos) não estão rotulados
- c) poderá ser utilizado apenas em algoritmos supervisionados
- d) contém objetos (exemplos) com atributos inválidos
- e) poderá ser utilizado apenas em algoritmos não-supervisionados

Questão 9

Valor da questão: 1,00

Em problemas reais de fluxo contínuo de dados, métodos de aprendizado de máquina devem ser capazes de reagir a mudanças de conceitos (*concept drift*). A mudança de conceito está ligada:

- a) a uma mudança no método de normalização dos dados.
- b) ao surgimento de novas classes e atributos.
- c) ao surgimento de novos atributos.
- d) a uma mudança na taxa de amostragem dos dados.
- e) a uma mudança na distribuição dos dados ao longo do tempo.

Questão 10

Valor da questão: 1,00

Uma pesquisadora possui um banco de dados rotulado no qual cada exemplo é representado por 15 atributos relacionados à ocorrência de um tumor. Cada exemplo é rotulado como sendo relativo à fase A, B ou C do tumor. A pesquisadora deseja construir um sistema classificador a partir desse banco de dados. Para isso, pretende utilizar uma rede neural artificial do tipo perceptron de múltiplas camadas junto ao algoritmo de aprendizado por retropropagação de erro (*back-propagation*). É CORRETO afirmar que:

- a) a ordem de apresentação dos exemplos de treinamento pode influenciar o desempenho de aprendizado do sistema classificador
- b) a ordem de apresentação dos exemplos de treinamento não influencia o desempenho de aprendizado do sistema classificador
- c) o desempenho de aprendizado do sistema classificador independe dos pesos iniciais atribuídos às conexões da rede neural artificial
- d) sistemas classificadores com base em redes neurais artificiais só podem ser utilizados em problemas de classificação binários (2 classes ou rótulos), portanto a pesquisadora deve escolher outro método de aprendizado supervisionado
- e) para a construção desse sistema classificador, os dados não precisam ser rotulados

Questão 11

Valor da questão: 1,00

Um centro de pesquisas possui um banco de dados de pacientes formado por diversos atributos (peso, altura, ingestão diária de calorias, medidas corporais, etc.) e rotulado em três classes: paciente com diabetes tipo 1 (D1), paciente com diabetes tipo 2 (D2) e paciente saudável (S). Um pesquisador recebeu a tarefa de ajudar o corpo clínico de um hospital a construir um fluxograma para ajudar a identificar a doença a partir dos valores dos atributos. Para isso, recomenda-se ao pesquisador utilizar:

- a) máquinas de vetores de suporte
- b) k-médias
- c) redes neurais artificiais
- d) agrupamento hierárquico
- e) árvores de decisão

Questão 12

Valor da questão: 1,00

O problema conhecido como Maldição da Dimensionalidade é observado quando, para um determinado número de amostras (exemplos) em um banco de dados, o desempenho de um modelo classificador degrada:

- a) pelo aumento do número de atributos de cada exemplo
- b) pela diminuição do número de atributos de cada exemplo
- c) pelo aumento do número de rótulos (classes) do problema
- d) pelo aumento do valor máximo de cada atributo do problema
- e) pela diminuição do número de rótulos (classes) do problema

Questão 13

Valor da questão: 1,00

Uma rede com centenas de clínicas construiu um conjunto de dados composto por diversos atributos de cada unidade (número de consultórios, número de médicos, área do estacionamento, etc.). Cada clínica foi rotulada como "boa", "regular" e "ruim", após uma pesquisa de opinião com seus clientes. A direção da rede gostaria de visualizar esse conjunto multidimensional de dados em apenas 3 dimensões. Qual técnica abaixo poderia ser utilizada para essa tarefa?

- a) máquina de vetor de suporte (SVM)
- b) curvas ROC (Curva Característica de Operação do Receptor)
- c) *Naive Bayes*
- d) análise de componentes principais (PCA)
- e) normalização pelo valor máximo do atributo

Questão 14

Valor da questão: 1,00

Uma rede com centenas de clínicas construiu um conjunto de dados composto pela coordenada geográfica bidimensional de cada clínica. A direção da rede deseja construir 3 centros de distribuição de insumos para essas clínicas, considerando-se apenas as coordenadas geográficas, de forma que cada clínica fique próxima a pelo menos um centro de distribuição. Em uma primeira abordagem a esse problema, pode-se utilizar o método de:

- a) agrupamento *k*-médias com *k* igual ao número de clínicas
- b) *k*-vizinhos mais próximos com *k* igual ao número de clínicas
- c) *k*-vizinhos mais próximos com *k* igual a 3
- d) agrupamento *k*-médias com *k* igual a 3
- e) agrupamento *k*-médias com *k* igual a 2

Questão 15

Valor da questão: 1,00

Uma virologista recebeu um conjunto de dados sobre vírus coletados em 25 cidades do país. Esse conjunto de dados inclui uma série de atributos para cada vírus, porém nenhum deles foi previamente identificado. Em uma análise preliminar, a virologista deseja visualizar e estimar quais vírus são mais parecidos, de acordo com um determinado critério de similaridade. Qual das técnicas abaixo poderia ser utilizada?

- a) agrupamento hierárquico e dendrograma
- b) *naive Bayes*
- c) *k*-médias com *k*=25
- d) *k*-vizinhos mais próximos
- e) máquinas de vetor de suporte

Nome:

Matrícula:

Assinatura: _____

Questão 1

Valor da questão: 1,00

Um Data Warehouse pode ser subdividido em pequenas porções para atender um departamento da empresa, por exemplo. Essa subdivisão do Data Warehouse é conhecida como:

- a) Database
- b) Data Center
- c) Data Store
- d) Data Mart
- e) Data Driven

Questão 2

Valor da questão: 1,00

Qual das etapas a seguir é executada pelo cientista de dados após a aquisição dos dados?

- a) Nenhuma delas
- b) Replicação de dados
- c) Limpeza de dados
- d) Integração de dados
- e) Modelagem de dados

Questão 3

Valor da questão: 1,00

Qual dos seguintes termos se concentra na descoberta de propriedades (anteriormente) desconhecidas nos dados?

- a) Mineração de dados
- b) Disputa de dados
- c) Análise de Clustering
- d) Big Data
- e) Aprendizado de Máquina

Questão 4

Valor da questão: 1,00

Todos são tipos de Aprendizado de Máquina, exceto:

- a) Estocástico
- b) Semissupervisionado
- c) Supervisionado
- d) Não supervisionado
- e) Por Reforço

Questão 5

Valor da questão: 1,00

Qual dos métodos de agrupamento abaixo se baseia na variância e no desvio padrão dos dados?

- a) k-means
- b) Hierarchical agglomerative
- c) Decision Tree
- d) Gaussian mixture model
- e) DBScan

Questão 6

Valor da questão: 1,00

Dentre as seguintes opções, identifique aquela que está relacionada a redução da dimensionalidade.

- a) estocasticidade
- b) acurácia
- c) entropia
- d) colinearidade
- e) desempenho

Questão 7

Valor da questão: 1,00

Um conjunto de dados possui os seguintes valores numéricos para um determinado atributo: 1; 3; 7; 5; 2; 6; 10. Após processo de normalização pelo valor máximo, os valores desse atributo são:

- a) -9; -7; -5; -8; -4; 0
- b) 0,1; 0,3; 0,7; 0,5; 0,2; 0,6; 1
- c) 1; 1; 1; 1; 1; 1; 1
- d) 1; 3; 7; 5; 2; 6; 10
- e) 10; 10; 10; 10; 10; 10; 10

Questão 8

Valor da questão: 1,00

Um centro médico de pesquisas estuda uma doença rara em uma população. Em uma etapa do estudo, percebe-se que o conjunto de dados rotulados está desbalanceado. Sobre este conjunto de dados, é CORRETO afirmar que:

- a) poderá ser utilizado apenas em algoritmos supervisionados
- b) o número de objetos (exemplos) para um dado rótulo é superior ao número de objetos para um outro rótulo
- c) poderá ser utilizado apenas em algoritmos não-supervisionados
- d) alguns objetos (exemplos) não estão rotulados
- e) contém objetos (exemplos) com atributos inválidos

Questão 9

Valor da questão: 1,00

Em problemas reais de fluxo contínuo de dados, métodos de aprendizado de máquina devem ser capazes de reagir a mudanças de conceitos (*concept drift*). A mudança de conceito está ligada:

- a) a uma mudança no método de normalização dos dados.
- b) ao surgimento de novos atributos.
- c) ao surgimento de novas classes e atributos.
- d) a uma mudança na taxa de amostragem dos dados.
- e) a uma mudança na distribuição dos dados ao longo do tempo.

Questão 10

Valor da questão: 1,00

Uma pesquisadora possui um banco de dados rotulado no qual cada exemplo é representado por 15 atributos relacionados à ocorrência de um tumor. Cada exemplo é rotulado como sendo relativo à fase A, B ou C do tumor. A pesquisadora deseja construir um sistema classificador a partir desse banco de dados. Para isso, pretende utilizar uma rede neural artificial do tipo perceptron de múltiplas camadas junto ao algoritmo de aprendizado por retropropagação de erro (*back-propagation*). É CORRETO afirmar que:

- a) o desempenho de aprendizado do sistema classificador independe dos pesos iniciais atribuídos às conexões da rede neural artificial
- b) a ordem de apresentação dos exemplos de treinamento não influencia o desempenho de aprendizado do sistema classificador
- c) sistemas classificadores com base em redes neurais artificiais só podem ser utilizados em problemas de classificação binários (2 classes ou rótulos), portanto a pesquisadora deve escolher outro método de aprendizado supervisionado
- d) para a construção desse sistema classificador, os dados não precisam ser rotulados
- e) a ordem de apresentação dos exemplos de treinamento pode influenciar o desempenho de aprendizado do sistema classificador

Questão 11

Valor da questão: 1,00

Um centro de pesquisas possui um banco de dados de pacientes formado por diversos atributos (peso, altura, ingestão diária de calorias, medidas corporais, etc.) e rotulado em três classes: paciente com diabetes tipo 1 (D1), paciente com diabetes tipo 2 (D2) e paciente saudável (S). Um pesquisador recebeu a tarefa de ajudar o corpo clínico de um hospital a construir um fluxograma para ajudar a identificar a doença a partir dos valores dos atributos. Para isso, recomenda-se ao pesquisador utilizar:

- a) agrupamento hierárquico
- b) árvores de decisão
- c) redes neurais artificiais
- d) k-médias
- e) máquinas de vetores de suporte

Questão 12

Valor da questão: 1,00

O problema conhecido como Maldição da Dimensionalidade é observado quando, para um determinado número de amostras (exemplos) em um banco de dados, o desempenho de um modelo classificador degrada:

- a) pelo aumento do valor máximo de cada atributo do problema
- b) pelo aumento do número de atributos de cada exemplo
- c) pela diminuição do número de atributos de cada exemplo
- d) pela diminuição do número de rótulos (classes) do problema
- e) pelo aumento do número de rótulos (classes) do problema

Questão 13

Valor da questão: 1,00

Uma rede com centenas de clínicas construiu um conjunto de dados composto por diversos atributos de cada unidade (número de consultórios, número de médicos, área do estacionamento, etc.). Cada clínica foi rotulada como "boa", "regular" e "ruim", após uma pesquisa de opinião com seus clientes. A direção da rede gostaria de visualizar esse conjunto multidimensional de dados em apenas 3 dimensões. Qual técnica abaixo poderia ser utilizada para essa tarefa?

- a) curvas ROC (Curva Característica de Operação do Receptor)
- b) normalização pelo valor máximo do atributo
- c) máquina de vetor de suporte (SVM)
- d) *Naive Bayes*
- e) análise de componentes principais (PCA)

Questão 14

Valor da questão: 1,00

Uma rede com centenas de clínicas construiu um conjunto de dados composto pela coordenada geográfica bidimensional de cada clínica. A direção da rede deseja construir 3 centros de distribuição de insumos para essas clínicas, considerando-se apenas as coordenadas geográficas, de forma que cada clínica fique próxima a pelo menos um centro de distribuição. Em uma primeira abordagem a esse problema, pode-se utilizar o método de:

- a) *k*-vizinhos mais próximos com *k* igual ao número de clínicas
- b) *k*-vizinhos mais próximos com *k* igual a 3
- c) agrupamento *k*-médias com *k* igual ao número de clínicas
- d) agrupamento *k*-médias com *k* igual a 2
- e) agrupamento *k*-médias com *k* igual a 3

Questão 15

Valor da questão: 1,00

Uma virologista recebeu um conjunto de dados sobre vírus coletados em 25 cidades do país. Esse conjunto de dados inclui uma série de atributos para cada vírus, porém nenhum deles foi previamente identificado. Em uma análise preliminar, a virologista deseja visualizar e estimar quais vírus são mais parecidos, de acordo com um determinado critério de similaridade. Qual das técnicas abaixo poderia ser utilizada?

- a) *k*-vizinhos mais próximos
- b) agrupamento hierárquico e dendrograma
- c) *k*-médias com *k*=25
- d) *naive Bayes*
- e) máquinas de vetor de suporte