

-- CONHECIMENTOS ESPECÍFICOS --

Considerando a função real na forma $g(x) = x e^{-x}$, na qual $x \in \mathbb{R}$, julgue os itens subsequentes.

- 51 $\int_0^{+\infty} g(x) dx > 2$.
- 52 Para $x > 0$, é correto afirmar que $g(\ln x) = \ln x - x$.
- 53 $\lim_{x \rightarrow +\infty} g(x) = 0$.
- 54 $\left. \frac{dg(x)}{dx} \right|_{x=0} = g(0) + 1$.
- 55 Para todo $x \in \mathbb{R}$, $\frac{d^2 g(x)}{dx^2} < 0$.

Com respeito à função

$$S(x) = \sum_{i=1}^n (a_i - x)^2,$$

na qual $a_i \in \mathbb{R}$ e $x \in \mathbb{R}$, julgue os próximos itens.

- 56 O valor mínimo global da função $S(x)$ é $S(b)$, em que $b = \frac{1}{n} \sum_{i=1}^n a_i$.
- 57 Definindo-se os vetores $\vec{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ e $\vec{u}_x = x \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, ambos de dimensões $n \times 1$, a função $S(x)$ pode ser escrita na forma de um produto vetorial como $S(x) = (\vec{a} - \vec{u}_x) \times (\vec{a} - \vec{u}_x)$.

Com respeito à matriz $A = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$, julgue os itens a seguir.

- 58 $A^{-1} = \begin{pmatrix} 0,5 & 0 \\ 0 & 0,5 \end{pmatrix}$ representa a matriz inversa de A .
- 59 Os autovalores de A são -2 e $+2$.
- 60 Denotando-se a matriz transposta de A como A^T , é correto afirmar que $A^T = A$.

		X		
		0	1	2
Y	0	0,0	0,0	0,1
	1	0,1	0,5	0,1
	2	0,1	0,0	0,1

Considerando o quadro precedente, que mostra a distribuição conjunta de um par de variáveis aleatórias discretas (X, Y) , julgue os seguintes itens.

- 61 $P(Y = 1 | X = 0) = \frac{1}{7}$.
- 62 A covariância entre as variáveis X e Y é positiva.
- 63 A variância da distribuição de X é maior do que a de Y .
- 64 As variáveis X e Y possuem a mesma esperança.

No que se refere à variável aleatória V , que segue uma distribuição contínua, tal que $P(V > v) = \exp(-v)$, se $v \geq 0$, e $P(V > v) = 0$, se $v < 0$, julgue os próximos itens.

- 65 $P(V > 1 | V > 2) = 1$.
- 66 A esperança e a variância de V são iguais a 1.
- 67 $P(V = 0) = \exp(0)$.

Uma amostra aleatória simples de tamanho $n = 4$, denotada por X_1, X_2, X_3, X_4 , foi retirada de uma população cuja função de distribuição de probabilidade é representada pela expressão $P(X = x) = \pi^x(1 - \pi)^{1-x}$, na qual x pode assumir os valores 0 ou 1 e π é o parâmetro desconhecido que denota uma probabilidade.

A partir das informações anteriores, e considerando a estimação do parâmetro π e o teste da hipótese nula $H_0: \pi = 0,5$ contra a hipótese alternativa $H_1: \pi \neq 0,5$, bem como sabendo que os valores observados na amostra foram 0,0,0,1, julgue os itens a seguir.

- 68 A estimativa de máxima verossimilhança da probabilidade π é igual a 0,75.
- 69 Sob a hipótese nula, a variância populacional é igual a 0,25.
- 70 Mantendo-se os mesmos valores 0,0,0,1 observados na amostra, o intervalo simétrico de 95% de confiança para π deve apresentar amplitude superior àquela proporcionada pelo intervalo simétrico de 99% de confiança para esse mesmo parâmetro.

A respeito das bibliotecas NumPy e Pandas, julgue os itens a seguir.

- 71 A classe `numpy.polyd()` permite a criação de *arrays* multidimensionais.
- 72 Se um conjunto de dados for armazenado em um *array* `numpy`, então, por meio dos seus métodos pré-definidos, será possível obter dados de resumo estatístico desse conjunto de dados.
- 73 A biblioteca Pandas apresenta os dados em uma estrutura de `DataFrame`, composta por linhas e colunas.
- 74 O método `describe()` da biblioteca Pandas retorna as linhas superiores e inferiores do `DataFrame`.
- 75 A biblioteca `numpy` permite realizar operações matemáticas entre *arrays* de diferentes dimensões usando o mecanismo de *broadcast*.

No que se refere a técnicas, ferramentas e algoritmos de aprendizado não supervisionado, julgue os itens a seguir.

- 76 A redução de dimensionalidade é uma técnica que reduz a quantidade de atributos que descrevem um objeto, mantendo a integridade dos dados originais.
- 77 A regra de associação é uma técnica que busca relações de co-ocorrência entre objetos de uma base de dados.
- 78 A compressão de atributos é uma técnica de redução de dimensionalidade na qual atributos irrelevantes ou redundantes são identificados e desconsiderados.
- 79 O PCA é um procedimento estatístico que converte um conjunto de objetos com atributos possivelmente correlacionados em um conjunto de objetos com atributos linearmente descorrelacionados.
- 80 O algoritmo *k-means* seleciona objetos reais de uma base de dados como centroide do grupo para realizar o agrupamento de objetos semelhantes.
- 81 O algoritmo *k-means* é utilizado para realizar o agrupamento de dados e opera por meio de refinamento iterativo.

Julgue os próximos itens, relativos a aprendizado de máquina.

- 82** As técnicas de regressão utilizam um conjunto finito de hipóteses para, a partir dos atributos previsores, determinar a categoria de um objeto do conjunto de dados analisado.
- 83** Os algoritmos SVM (*support vector machines*) realizam apenas tarefas de regressão.
- 84** A técnica de agrupamento é um tipo de aprendizado não supervisionado em que o algoritmo identifica padrões em um conjunto de dados de entrada sem ter recebido qualquer *feedback* prévio.
- 85** O aprendizado por reforço é um tipo de aprendizagem de máquina que tem por objetivo prever o resultado de um atributo alvo exclusivamente por meio de reforço no treinamento do modelo.

A respeito de redes neurais, julgue os itens que se seguem.

- 86** As conexões entre as camadas de uma rede neural do tipo MLP são de natureza *feedforward*.
- 87** Nas redes neurais completamente conectadas, todos os neurônios de uma camada estão conectados aos neurônios da camada seguinte, no entanto, não é possível que as saídas das camadas posteriores alimentem a entrada de camadas anteriores.
- 88** As redes neurais têm a capacidade de adaptar seus pesos sinápticos considerando as mudanças de padrão dos dados de entrada.
- 89** *Backpropagation* propaga o erro da camada de saída para as camadas intermediárias de uma rede neural a fim de que estas possam modificar seus pesos de forma a minimizar o erro médio.
- 90** Apesar das CNN serem redes neurais profundas, com várias camadas ocultas, elas são dispensadas nos algoritmos de reconhecimento de imagens.

No que se refere à biblioteca Pandas, julgue o próximo item.

- 91**

```
import pandas as pd
a = [10, 78, 23]
myvar = pd.Series(a, index = ["x", "y", "z"])
print(myvar.describe())
```

Caso o código em Python precedente seja executado, então será apresentado em tela o que se segue.

```
count 4.000000
mean 38.000000
std 6.097091
min 10.000000
25% 16.500000
50% 23.000000
75% 50.500000
max 78.000000
dtype: float64
```

A respeito da biblioteca Numpy, julgue os itens a seguir.

- 92** O resultado da execução do código Python a seguir resultará em valores diferentes para as variáveis *c* e *d*.

```
import numpy as np
a = np.array([[1,2],[3,4]])
b = np.array([[4,3],[2,1]])
c= np.dot(a,b)
print(c)
d=a*b
print(d)
```

- 93**

```
import numpy as np
a = np.array([(1, 2), (3, 4), (5, 6)],
dtype=[('x', 'i4'), ('y', 'i4')])
print(a)
print(a.shape)
```

O resultado da execução do código Python apresentado anteriormente é o que se segue.

```
[[1 2]
 [3 4]
 [5 6]
 ['x','i4']
 ['y','i4']]
(3, 2)
```

Espaço livre

Com relação às bibliotecas Matplotlib, Seaborn, Streamlit e Scipy, julgue os itens subsequentes.

94 O resultado da execução do código a seguir será 3.4535923699999997.

```
from scipy import constants
print(constants.pi)
```

95 O módulo `pyplot` da biblioteca Matplotlib é um conjunto de funções de estilo de comando que permitem que gráficos sejam gerados.

96 A função `histogram()` do Seaborn permite que se gere um histograma de um *array* unidimensional.

97

```
from scipy import optimize
x = [-2,-1.6, -1.2, -0.8, -0.4, 0, 0.4, 0.8, 1.2, 1.6, 2.0]
def f(x):
    return x*x-1
ra = optimize.root(f, 0.5)
print("ra %s" % (ra.x))
```

A execução do código Python apresentado anteriormente resultará no seguinte.

```
ra [1.]
```

98 Streamlit é uma biblioteca utilizada para criar interfaces *web*, como, por exemplo, um *dashboard* com vários gráficos interativos.

Julgue o item a seguir, a respeito de TensorFlow.

99 *Pipelines*, em TensorFlow, são um conjunto de componentes padrão que podem ser utilizados como parte de um *script* de treinamento de *machine learning*.

A respeito da biblioteca Leaflet para R, julgue o próximo item.

100 Leaflet é uma biblioteca R para criar e manipular uma hierarquia de classes ou árvore de herança, mostrando os relacionamentos entre classes na programação orientada a objetos.

```
library('dplyr')
lista <- data.frame(
  id = c(10,11,12,13),
  nome = c('pedro','mauro','maria','joana'),
  sexo = c('M','M','F','F'),
  uf = c('PA','RJ','AL',NA),
  row.names=c('r1','r2','r3','r4')
)
```

Considerando o código precedente, julgue os itens a seguir, a respeito da biblioteca `dplyr` para R.

101 Considerando-se o *dataframe* `lista` do código R apresentado, ao se executar o comando `lista %>% select(c(1,3))`, o resultado será o seguinte.

```
id nome
r1 10 pedro
r2 11 mauro
r3 12 maria
r4 13 joana
```

102 Considerando-se o *dataframe* `lista` do código R apresentado, ao se executar o comando `dim(lista)`, o resultado será `[1] 4 4`.

Tendo como referência a linguagem R, julgue os itens subsequentes.

103 O resultado da execução do código R a seguir será 2.

```
x <- 10
y = 12
z = x+y
print(z)
```

104 O resultado da execução do código R a seguir será `[1] -2`.

```
B <- matrix(1:4, ncol = 2)
det(B)
```

105 O resultado da execução do código R a seguir será 6.

```
i = 1
n = 5
while i <= n:
  i = i + 1
print(i)
```

106 Para fazer comentários na linguagem R, é necessário acrescentar os caracteres `//` antes do que se deseja comentar.

107

```
A=matrix(c(1,2,3,4,5,6), ncol=2, nrow=3)
B=matrix(c(6,5,4,3,2,1), nrow=3, ncol=2)
C=A*B
```

O resultado da execução do código R precedente será o seguinte.

```
dim(C)
[1] 2 2
```

Espaço livre

Julgue o item a seguir, a respeito de documentação em R.

- 108** O R Markdown é a junção da linguagem Markdown com o poder de códigos em R. Em arquivos do tipo `.rmd`, é possível adicionar textos, códigos e resultados de códigos.

Julgue os itens a seguir, relativos ao Hadoop e ao Spark.

- 109** Hadoop é uma plataforma de código aberto que oferece uma solução tecnológica confiável e escalável para armazenamento e análise de grande quantidade de dados.
- 110** Spark é uma arquitetura de sistema operacional para trabalhar com dados organizados de forma hierárquica, utilizando-se o módulo Spark SQL, que permite consultar arquivos pelo seu conteúdo.

A respeito de modelagem de dados, julgue os itens a seguir.

- 111** A modelagem conceitual representa o negócio sob a perspectiva dos dados, e, em projetos de TI, o principal objetivo de um modelo de dados conceitual é fornecer uma visão geral dos requisitos de informação envolvidos no projeto.
- 112** O modelo em estrela e o modelo em floco de neve são os principais modelos de dados relacionais.
- 113** A integridade referencial é uma parte essencial de qualquer banco de dados relacional, que, entre outros benefícios, facilita as consultas.
- 114** A modelagem dimensional é uma técnica de modelagem de banco de dados que usa um modelo de dados baseado em dimensões para organizar os dados, e as tabelas de dimensões são sempre as maiores tabelas do modelo dimensional.
- 115** Um banco de dados que se encontra na terceira forma normal obrigatoriamente está também na segunda e na primeira formas normais.

Em relação às linguagens de banco de dados SQL, DDL e DML, julgue os itens que se seguem.

- 116** Na DDL, que é uma linguagem declarativa, é descrito o que se deseja fazer, em vez de como fazê-lo, como comandos para definir tabelas e *procedures*, tal qual exemplificado a seguir.

```
CREATE TABLE <tablename> (col1 int, col2 int, col3 int)
```

- 117** A DML é uma linguagem que interage com os objetos do banco de dados, em vez de interagir com os dados.
- 118** Somente são possíveis os seguintes quatro tipos de JOIN em SQL, segundo o padrão ANSI: INNER JOIN, LEFT JOIN, RIGHT JOIN e CROSS JOIN.

No que se refere a banco de dados NoSQL e *data lake*, julgue os próximos itens.

- 119** Os bancos relacionais permitem uma escalabilidade mais barata e menos trabalhosa que os bancos não relacionais, pois não exigem máquinas extremamente poderosas e sua facilidade de manutenção possibilita um número menor de profissionais necessários.
- 120** Considere-se que uma empresa de comércio eletrônico deseje armazenar todos os dados relacionados às interações dos clientes em seu *site* — registros de visitas, cliques, compras, avaliações, comentários, entre outros. Nessa situação, se a empresa decidir não estruturar e organizar esses dados antecipadamente, será melhor armazenar esses dados em uma solução *data lake* do que em um banco relacional.

Espaço livre