EMC

Exam E20-065

Advanced Analytics Specialist Exam for Data Scientists

Version: 7.0

[Total Questions: 66]

Question No : 1

What is an important simu-lation design consideration?

A. Ensure model Inputs align with reality
B. Use different seed values to regenerate results
C. For rare event models, minimize number of trials
D. A complex model is better than a simple model

Answer: A

Question No : 2

How is the relative value of a node visualized in a sunburst?

A. Color
B. Area
C. Gradient
D. Position

Answer: A

Question No : 3

What describes how nodes in a social network are similar to each other in characteristics?

A. Community clustering
B. Modularity
C. Homophily
D. Strongly tied network

Answer: C

Question No : 4

What is a characteristic of stemming?

A. Reduces words of variant forms to their base forms based on a set of heuristics
B. Can be performed by calling the stemming!) function on a lemma in NLTK
C. Can be performed by calling the stemming() function on a synset in NLTK
D. Reduces words of variant forms to their base forms based on a dictionary

Answer: A

Question No : 5

Given an input vector of features, a Random Forests model performs a classification task and ends in a tie.

How does the model handle this outcome?

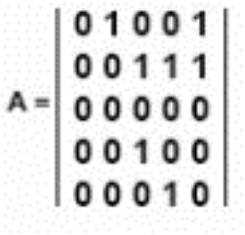
A. The model will be rebuilt
B. A winner is chosen at random
C. The tree that caused the tie is discarded
D. One more tree is added to the forest

Answer: B

Question No : 6

Assuming the node index starts at 1, what is the out-degree of node 3 in the adjacency matrix shown?

Refer to the exhibit.



A. 0
B. 1
C. 2
D. 3

Answer: A

Question No : 7

Consider dataset that resides in HDFS. Which tool natively provides the capability to run a Random Forests model against this data?

A. Mahout
B. Pig
C. Hive
D. HBase

Answer: A

Question No : 8

What is a typical use of a UDF in Pig?

A. Creating functionality outside of what is provided by the built-in functions
B. Providing Functional access to user-defined data in HDFS

C. Providing advanced analytics to Hadoop
D. Providing an interface from Pig to Microsoft Excel for easier data manipulation

Answer: A

Question No : 9

You conduct a TFIDF analysis on 3 documents containing raw text and derive TFIDF ("data", document y) = 1.908. You know that the term "data" only appears in document 2.

What is the TF of “data" in document 2?

A. 2 based on the following reasoning:
TFIDF = TF1DF = 1 908
You then know that IDF will equal LOG (32)=0.954
Therefore, TFIDF=TF*0.954 = 1.908
TF will then round to 2
B. 4 based on the following reasoning:
TFIDF = TF1DF = 1.908
You then know that IDF will equal LOG (3/1)=0.477
Therefore, TFIDF=TF'0 477 = 1.908
TF will then round to 4
C. 6 based on the following reasoning:
TFIDF = TF1DF = 1.908
You then know that IDF will equal 3/1=3
Therefore, TFIDF=TF/3 = 1.908
TF will then round to 6
D. 11 based on the following reasoning:
TFIDF = TF1DF = 1908
You then know that IDF will equal LOG(3/2)=0.176
Therefore, TFIDF=TF"0.176 = 1.908
TF will then round to 11

Answer: B

Question No : 10

What is the maximum number of edges in an undirected graph of 10 nodes?