BRAINWARE UNIVERSITY

Term End Examination 2023-2024
Programme – B.Tech.(CSE)-2020
Course Name – Data Analytics
Course Code - PEC-702A
(Semester VII)

Full Marks : 60 **Time : 2:30 Hours**

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question) 1 x 15=15

1. *Choose the correct alternative from the following :*

(i) For t distribution, increasing the sample size, the effect will be apply on

a) degrees of freedom b) the t-ratio
c) standard error of the means d) all of these

(ii) Choose which of the following would have a constant input in each epoch of training a deep learning model

a) weight between input and hidden layer b) weight between hidden and output layer
c) biases of all hidden layer neurons d) activation function of output layer

(iii) Select which of the following data mining technique is used to uncover patterns in data?

a) data bagging b) data booting
c) data merging d) data dredging

(iv) Choose what is the name of the model in figure below

a) rosenblatt perceptron model b) mcculloch-pitts model
c) widrow's adaline model d) none of the mentioned

(v) Select which of the following algorithm would you take into the consideration in your final model building on the basis of performance? Suppose you have given the following graph which shows the ROC curve for two different classification algorithms such as random forest (Red) and logistic regression (Blue)

a) random forest b) logistic regression
c) both random forest & logistic regression d) none of these

(vi) In random forest or gradient boosting algorithms, features can be of any type. For example, it can be a continuous feature or a categorical feature. Select which of the following option is true when you consider these types of features

a) only random forest algorithm handles real valued attributes by discretizing them b) only gradient boosting algorithm handles real valued attributes by discretizing them

c) both algorithms can handle real valued attributes by discretizing them
d) none of these

(vii) Select, The resources owned by WSFC node include _____
a) destination address
b) SQL server browser
c) one file share resource, if the FILESTREAM feature is installed
d) none of the mentioned

(viii) Select an exciting new feature in SQL Server 2014 is the support for the deployment of a failover cluster instance (FCI) with _____
a) cluster shared volumes (CSV)
b) in memory database
c) column oriented database
d) all of the mentioned

(ix) Select which of the following argument is used to set importance values
a) scale
b) set
c) value
d) all of the mentioned

(x) Select which of the following has a design policy of using zoo keeper only for transient data
a) hive
b) imphala
c) hbase
d) oozie

(xi) Select what is the primary function of tRNA (transfer RNA) in the process of protein synthesis
a) Transcribing genetic information from DNA
b) Carrying amino acids to the ribosome
c) Providing a template for protein synthesis
d) Forming the ribosomal structure

(xii) Select which of the following is one of the largest boost subclass in boosting
a) variance boosting
b) gradient boosting
c) mean boosting
d) all of the mentioned

(xiii) Select Which of the following is the most important language for data science
a) Java
b) Ruby
c) R
d) none of the mentioned

(xiv) If a test was generally very easy, except for a few students who had very low scores, then the distribution of scores would be defined as _____
a) positively skewed
b) negatively skewed
c) not skewed at all
d) normal

(xv) select which of the following is the advantage/s of decision trees?
a) possible scenarios can be added
b) use a white box model, if given result is provided by a model
c) use a white box model, if given result is provided by a model
d) all of the mentioned

Group-B

(Short Answer Type Questions) 3 x 5=15

2. Describe cleansing and what are the best ways to practice data cleansing? (3)
3. Write the best practices in big data analytics? (3)
4. Write is the difference between data mining and data analysis? (3)
5. Write how Hadoop is related to big data? (3)
6. Justify why SVM is so fast. (3)

OR

Justify What are the best practices in big data analytics. (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Explain in detail about the probability distribution. (5)
8. Explain What is the role activation function in neural network? (5)
9. Describe the term Data Wrangling in Data Analytics. (5)
10. What is the role activation function in neural network? explain (5)
11. Jystify "SVMs often more accurate than logistic regression" (5)
12. Explain Why is checking for multicollinearity important in multiple regression models? (5)

OR

Analyze What are some common techniques for checking the assumptions of linear regression models? (5)

****