BRAINWARE UNIVERSITY

Term End Examination 2022
Programme – B.Tech.(CSE)-2017/B.Tech.(CSE)-2018/B.Tech.(CSE)-2019
Course Name – Data Analytics
Course Code - BCSE701/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) A fair six-sided die is rolled twice. Calculate What is the probability of getting 2 on the first roll and not getting 4 on the second roll?

    a) 1/36   b) 1/18

    c) 5/36   d) 1/6

(ii) Suppose you have trained a logistic regression classifier and it outputs a new example x with a prediction $h_\theta(x) = 0.2$. This determine

    a) our estimate for $P(y=1 \mid x)$ b) our estimate for $P(y=0 \mid x)$

    c) All of these   d) None

(iii) For Â 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

(iv) Suppose you are using a bagging based algorithm say a random forest in model building. Select which of the following can be true? 1.Number of tree should be as large as possible 2.You will have interpretability after using random forest

    a) 1   b) 2

    c) 1 and 2   d) none of these

(v) Categorize, The process of adjusting the weight is known as

    a) activation   b) synchronization

    c) learning   d) none of the mentioned

(vi) Select Which of the following mentioned standard probability density functions is applicable to discrete random variables?

    a) gaussian distribution   b) poisson distribution

    c) rayleigh distribution   d) exponential distribution

(vii) The denominator (bottom) of the z-score formula is defined as

    a) the standard deviation   b) b.the difference between a score and the mean

    c) the range   d) the mean

(viii) 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

(ix) Measure,A windows failover cluster can support up to how many nodes

a) 12

b) 14

c) 16

d) 18

(x) Explain how can you prevent a clustering algorithm from getting stuck in bad local optima?

a) set the same seed value for each run

b) use multiple random initialization

c) both set the same seed value for each run and use multiple random initialization

d) none of these

(xi) Select which of the following is a windows failover cluster quorum mode?

a) node majority

b) no majority: read only

c) file read majority

d) none of the mentioned

(xii) 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

(xiii) To register a watch on a z node data, write, what commands you need to use to access the current content or metadata.

a) stat

b) put

c) receive

d) gets

(xiv) Write which of the following specifies the required minimum number of observations for each column pair in order to have a valid result?

a) min_periods

b) max_periods

c) minimum_periods

d) all of the mentioned

(xv) You run gradient descent for 15 iterations with a=0.3 and compute J (theta) after each iteration. You find that the value of J (Theta) decreases quickly and then levels off. Based on this, select which of the following conclusions seems most plausible?

a) rather than using the current value of a, use a larger value of a (say a=1.0)

b) rather than using the current value of a, use a smaller value of a (say a=0.1)

c) a=0.3 is an effective choice of learning rate

d) none of these

Group-B
(Short Answer Type Questions)

3 x 5=15

2. Explain data cleansing .

(3)

3. Explain, How can the initial number of clusters for k-means algorithm be estimated?

(3)

4. Explain in detail about the probability distribution.

(3)

5. Analyze the different types of Hypothesis testing.

(3)

OR

Explain What is the role activation function in neural network?

(3)

6. evaluate the characteristics of big data.

(3)

OR

compare the Type I and Type II errors in Statistics.

(3)

Group-C
(Long Answer Type Questions)

5 x 6=30

7. write a few problems that data analyst usually encounter during performing the analysis.

(5)

8. Describe ANOVA.

(5)

9. Discuss about the applications of clustering.

(5)

10. Classify about overfitting and underfitting and how to tackle them?

(5)

11. Analyze the design principles of neural network.

(5)

OR

Analyze why SVMs are often more accurate than logistic regression with examples.   (5)

12. Assess what are the best practices in big data analytics.   (5)

OR

Evaluate the techniques used in big data analytics.   (5)

****