

32. a. What are the some common tuning parameter for classification and regression trees in R, and how they affect the model? 12 2 5 4

(OR)

b. What is the difference between the bagging function and the random forest function in R? 12 2 5 1

\* \* \* \* \*

Reg. No.

**B.Tech. DEGREE EXAMINATION, MAY 2023**  
Fourth to Seventh Semester

18CSO106T – DATA ANALYSIS USING OPEN SOURCE TOOL  
(For the candidates admitted during the academic year 2018-2019 to 2021-2022)

Note:

- (i) **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40<sup>th</sup> minute.  
(ii) **Part - B & Part - C** should be answered in answer booklet.

Time: 3 hours

Max. Marks: 100

**PART – A (20 × 1 = 20 Marks)**

Answer **ALL** Questions

- |  | Marks | BL | CO | PO |
|--|-------|----|----|----|
| 1. What is the characteristics of processed data?<br>(A) Data is not ready for analysis (B) Difficult to use for data analysis<br>(C) All steps should be noted (D) Few steps may be avoid   | 1     | 1  | 1  | 1  |
| 2. What is the length of b? $b \leftarrow 2!7$<br>(A) 4 (B) 7<br>(C) 6 (D) 9   | 1     | 1  | 2  | 1  |
| 3. The most convenient way to use R is at a graphics work station running a _____ system.<br>(A) Windowing (B) Running<br>(C) Interfacing (D) Matrix   | 1     | 1  | 1  | 1  |
| 4. Functionality of R is divided into a number of _____.<br>(A) Functions (B) Domains<br>(C) Packages (D) Files  | 1     | 1  | 1  | 1  |
| 5. Amongst which of the following is /are the true about regression analysis?<br>(A) Describes associations within the data (B) Modeling relationship within the data<br>(C) Answering yes/no questions about the data (D) Clustering the data | 1     | 1  | 2  | 1  |
| 6. The process of quantifying's data is referred to as _____.<br>(A) Decoding (B) Structure<br>(C) Enumeration (D) Coding  | 1     | 1  | 2  | 2  |
| 7. _____ are used when we want to visually examine the relationship between to quantitative variables<br>(A) Bar graph (B) Scatter plot<br>(C) Line graph (D) Pie chart  | 1     | 1  | 2  | 2  |
| 8. _____ is a simple approach to supervised learning, it assumes that the dependence of you $X_1, X_2, \dots, X_P$ is linear.<br>(A) Linear regression (B) Logistic regression<br>(C) Gradient descent (D) Greedy algorithms                   | 1     | 2  | 2  | 1  |

9. The parameter  $\beta_0$  is termed as intercept term and the parameter  $\beta$ , is termed as slope parameter. These parameter are usually called as  
 (A) Regressionists (B) Coefficients  
 (C) Regressive (D) Regression coefficients
10. When hypothesis tests and confidence limits are to be used, the residuals are assumed to follow the \_\_\_\_\_ distribution.  
 (A) Formal (B) Mutual  
 (C) Normal (D) Abnormal
11. \_\_\_\_\_ is an incredibly powerfull tool for analyzing data.  
 (A) Linear regression (B) Logistic regression  
 (C) Gradient descent (D) Greedy algorithms
12. Which of the following is used where the target variable is of categorical nature?  
 (A) Keras (B) Ktime  
 (C) Logistic regression (D) MxNet
13. Which of the following class in bootstrap is used to create a dropdown menu?  
 (A) .drop down (B) .select  
 (C) .select-list (D) .select-array
14. Which plugin is used to cycle through elements, like a slideshow?  
 (A) Carousel plugin (B) Modal plugin  
 (C) Tooltip plugin (D) Mat plugin
15. Which of the following class applies hover colour to a specific row or a cell?  
 (A) Warning (B) Active  
 (C) Success (D) Danger
16. Ridge regression takes \_\_\_\_\_ value of variables.  
 (A) Squared value of variables (B) Absolute value of variables  
 (C) Cube value of variables (D) Root value of variables
17. To do ridge and lasso regression in R which library we will used?  
 (A) Gplot (B) Glmnet  
 (C) Caret (D) Dplyr
18. A \_\_\_\_\_ is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences including cache event outcome, resource costs, and utility.  
 (A) Decision tree (B) Graphs  
 (C) Tree (D) Neural network
19. End nodes are represented by  
 (A) Disks (B) Squares  
 (C) Circles (D) Triangles

20. Which of the following ensemble model helps in reducing variance?  
 (A) Boosting (B) Bagging  
 (C) Stacking (D) Voting

**PART – B (5 × 4 = 20 Marks)**

Answer ANY FIVE Questions

21. Differentiate data, information and knowledge.
22. What are the different types of built-in functions in R?
23. What are the different types of machine learning? What is the training set and test set in machine learning?
24. Discuss about the linear discriminant analysis for  $P > 1$ .
25. What do you mean by leque-one cross validation?
26. Discuss in detail about the bootstrap in R.
27. How can decisions trees be used to solve the regression problem?

**PART – C (5 × 12 = 60 Marks)**

Answer ALL Questions

28. a. Explain different types of data type in details with suitable example.
- (OR)
- b. Write steps to install the R programming. Which function is used to install the package in R?
29. a. What do you mean by simple regression? Write libraries used for simple linear regression in R.
- (OR)
- b. Examine the purpose of using regression modeling in data analysis.
30. a. Explain logistic regression in machine learning with their types and proper equation.
- (OR)
- b. How classification model are evaluated? Explain confusion matrix with suitable example.
31. a. What is the difference between forward and backward stepwise selection? Explain in brief.
- (OR)
- b. What is the Lasso method and how does it differ from ridge regression?