Printed Page:- 04 **Subject Code:- ABT0612**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VI - THEORY EXAMINATION (2023 - 2024)

Subject: Probability and Statistics using R in Biotechnology

Time: 3 Hours **Max. Marks: 100**

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. *This Question paper comprises of* ***three Sections -A, B, & C.*** *It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.*

2. *Maximum marks for each question are indicated on right -hand side of each question.*

3. *Illustrate your answers with neat sketches wherever necessary.*

4. *Assume suitable data if necessary.*

5. *Preferably, write the answers in sequential order.*

6. *No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.*

SECTION-A 20

1. Attempt all parts:-

1-a. In which ANN, loops are allowed? (CO1) 1

(a) Feedforward

(b) Feedbackward

(c) Both

(d) none of above

1-b. How many types of Machine Learning Techniques? (CO1) 1

(a) 3

(b) 5

(c) 7

(d) 8

1-c. How could be the matrix constructed by using the following R code? m <- matrix(1:6, nrow = 2, ncol = 3) (CO2) 1

(a) row-wise

(b) column-wise

(c) any manner

(d) data insufficient

1-d. What is the function to set row names for a matrix? (CO2) 1

(a) names()

(b) rownames()

(c) row.names()

(d) column name cannot be set for a matrix

1-e. X in linear regression is _____ variable (CO3) 1

(a) Dependent

(b) Independent

(c) Both

(d) None

1-f. Which of the following of a random variable is a measure of central tendency? (CO3) 1

(a) mean

(b) variance

(c) standard deviation

(d) range

1-g. Which of the following is the cyclic behavior of time series? (CO4) 1

(a) level

(b) trend

(c) seasonality

(d) none

1-h. A time series is a set of recorded data (CO4) 1

(a) periodicaaly

(b) all time or space interval

(c) at successive point of time

(d) All of the above

1-i. The science of collecting,organising, presenting , analyzing and interpreting data to assist in making more effective decisions is called: (CO5) 1

(a) Statistic

(b) Parameter

(c) Population

(d) Statistics

1-j. Methods of organizing, summarizing and presenting data in an informative way are called: (CO5) 1

(a) Descriptive statistics

(b) Inferential Statistics

(c) Theoritical statistics

(d) Applied statistics

2. Attempt all parts:-

2.a. Explain different types of Activation Function. (CO1) 2

2.b. Give any five features of R.(CO2) 2

2.c.	Find the mean of first ten whole numbers. (CO3)	2
2.d.	List down the attribute selection measures used by the ID3 algorithm to construct a Decision Tree. (CO4)	2
2.e.	Define genetic engineering. (CO5)	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Discuss the applications of ANN. (CO1)	6
3-b.	Discuss Recurrent Neural Network. (CO1)	6
3-c.	Define variables and constant. Define the rules for variable with valid and invalid variables. (CO2)	6
3-d.	Explain loops in R. (CO2)	6
3.e.	Differentiate between symmetric and non symmetric function. (CO3)	6
3.f.	Explain chi-square test. When to use it? (CO4)	6
3.g.	Is body itself a biotechnology? Explain. (CO5)	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Illustrate BNN and ANN. Draw their structures with appropriate comparison. (CO1)	10
4-b.	Illustrate the practical machine learning problem. (CO1)	10

5. Answer any one of the following:-

5-a.	Define if and if-else statements. Write the syntax, flowchart and example for it. (CO2)	10
5-b.	Define Matrix in R. explain the properties of matrix. Create a 4*2 matrix in R. (CO2)	10

6. Answer any one of the following:-

6-a.	Suppose we are trying to create a model that can predict the result for the disease that is either a person has that disease or not. So, the confusion matrix concluded as: the two-class classifier, which has two predictions "Yes" and "NO." Here, Yes defines that patient has the disease, and No defines that patient does not has that disease. The classifier has made a total of 100 predictions. Out of 100 predictions, 89 are true predictions, and 11 are incorrect predictions. The model has given prediction "yes" for 32 times, and "No" for 68 times. Whereas the actual "Yes" was 27, and actual "No" was 73 times. Calculate : (i) Accuracy (ii) Recall (iii) Precision (CO3)	10
6-b.	Explain classification. Also explain ROC and AUC curve. (CO3)	10

7. Answer any one of the following:-

7-a.	Explain Random Forest. For what applications are random forests used? Explain How does a Random Forest Work? (CO4)	10

7-b.	Nadir is testing an octahedral die to see if it is unbiased. The results are given as: Score (1,2,3,4,5,6,7,8) and Frequency (7,10,11,9,12,10,14,7) resp. Test the hypothesis that the die is fair. (CO4)	10

8. Answer any <u>one</u> of the following:-

8-a.	Illustrate your opinion, which programming you will prefer in your research with reason. (CO5)	10
8-b.	Describe the programming languages used for biostatistics. (CO5)	10