Subject Code:- ABT0612

Max. Marks: 100

20

1

1

1

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VI - THEORY EXAMINATION (2022-2023)

Subject: Probability and Statistics using R in Biotechnology

Time: 3 Hours

Printed Page:-04

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

1. Attempt all parts:-

- 1-a. The actual processing occurs in which of the following layer? (CO1)
 - (a) Input Layer
 - (b) Hidden Layer
 - (c) Output Layer
 - (d) none of above

1-b. Machine learning is a subset of_____ (CO1)

- (a) Deep Learning
- (b) Data Learning
- (c) Artificial Intelligence
- (d) none
- 1-c. List entries contains_____(CO2)
 - (a) Numbers
 - (b) Characters
 - (c) Both numbers and characters

(d) None

1-d. Which of the following is the correct syntax for assigning a value to a variable in 1 R? (CO2)

1

1

1

1

1

JUNE

- (a) var = 10
- (b) 10 = var
- (c) var == 10
- (d) var := 10
- 1-e. For normal distribution, which factors are equal. (CO3)
 - (a) Mean
 - (b) Median
 - (c) Mode
 - (d) All of the above
- 1-f. The range of normal distribution is ___ (CO3)
 - (a) 0 to n
 - (b) 0 to infinite
 - (c) -1 to +1
 - (d) plus infinity to minus infinity
- 1-g. A time series has _____ components. (CO4)
 - (a) 1
 - (b) 2
 - (c) 3
 - (d) 4

1-h. In time series seasonal variations can occur within a period of: (CO4)

- (a) four years
 - (b) three years
- (c) less than one year
- (d) none
- 1-i. In the plural sense, statistics means: (CO5)
 - (a) Numerical Data
 - (b) Methods
 - (c) Population data
 - (d) Sample data
- 1-j. Protein and DNA attain their 3 dimensional reactive configuration contact with 1

(CO5)

- (a) water
- (b) lipids
- (c) zymogens
- (d) sugars

2. Attempt all parts:-

2.a.	List down Types of Machine Learning and Types of ANN. (CO1)	2
2.b.	Differentiate between vector, List, Matrix, and Data frame. (CO2)	2
2.c.	A class consists of 50 students, out of which 30 are girls. The mean of marks scored by girls in a test is 73 (out of 100), and that of boys is 71. Determine the mean score of the whole class. (CO3)	2
2.d.	Mention the components of the time series. (CO4)	2
2.e.	Explain, how technology helps in biological process. (CO5)	2
	SECTION B	30
3. Answe	er any <u>five</u> of the following:-	
З-а.	Difference between Supervised and Unsupervised Learning. (CO1)	6
3-b.	Define Semi-supervised Machine Learning. (CO1)	6
З-с.	Write down the advantages and disadvantages of R. (CO2)	6
3-d.	Explain functions in R. (CO2)	6
3.e.	Differentiate between Regression and Classification. Why Linear Regression not suitable for classification? (CO3)	6
3.f.	Explain ANOVA. (CO4)	6
3.g.	Explain, what biological problem has AI solved. (CO5)	6
	SECTION C	50
4. Answe	er any <u>one</u> of the following:-	
4-a.	Differentiate single layer and multi layer feedforward network. Sketch the architecture of 10/16/5 multiplayer feed forward neural network. (CO1)	10
4-b.	(a) What is neuron and explain history of neural network. (b) What is hidden layer, explain with example. (CO1)	10
5. Answe	er any <u>one</u> of the following:-	
5-a.	Define Array. How to create and access array. Explain types of array in R. (CO2)	10
5-b.	Explain R and R studio. Explain its features and its applications. (CO2)	10
6. Answer any <u>one</u> of the following:-		

.

- 6-a. Explain Linear Regression. Find linear regression equation for the following 10 two sets of data: x(2,4,6,8) and y(3,7,5,10) respectively. (CO3)
- 6-b. Explain :(i) Classification (ii) Regression (iii) AUC and ROC curve (CO3)

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

- 7-a. Define Decision Tree. In class we used decision trees and ensemble methods 10 for classification, but we can use them for regression as well (i.e. learning a function from features to real values). Let's imagine that our data has 3 binary features A, B, C, which take values 0/1, and we want to learn a function which counts the number of features which have value 1. Draw the decision tree which represents this function. How many leaf nodes does it have. (CO4)
- 7-b. Explain Random forest. How does it works? Write down advantages and 10 disadvantages of it. (CO4)

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

8-a. Explain the application of R in biological processes. (CO5)

202-23

8-b. Explain the DNA technology. (CO5)

10