Printed Page:- **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 (20 ..... - 20.....)

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. The neural network is also known as _____. CO1, K1 1

(a) Machine Learning
(b) Artificial Neural Network
(c) Artificial System
(d) none of the above

1-b. What is the term known as on which the machine learning algorithms build a model based on sample data? CO1, K1 1

(a) Data Training
(b) Training Data
(c) Transfer Data
(d) none of above

1-c. What is the output of the following code in R? x <- c(1, 2, 3, 4, 5) mean(x) CO2, K2 1

(a) 3
(b) 3.5
(c) 4
(d) 5

1-d. Which of the following functions in R can be used to create a boxplot? CO2, K2 1

(a) boxplot()

.

(b) hist()
(c) plot()
(d) lines()

1-e. Which of the following of a random variable is not a measure of central tendency? CO3, K2 1

(a) mean
(b) variance
(c) mode
(d) median

1-f. The Mean of a constant 'x' is CO3, K2 1

(a) 0
(b) 1
(c) 0.5
(d) x

1-g. _____is the measure of uncertainty of a random variable, it characterise the impurity of an arbitrary collection of examples. CO4, K2 1

(a) Information Gain
(b) Gini Index
(c) Entropy
(d) None

1-h. Choose a disadvantage of decision trees among the following. CO4, K2 1

(a) Decision trees are robust to outliers.
(b) Factor analysis
(c) Decision trees are prone to overfit
(d) All of the above

1-i. The first step of investigation is ____ CO5, K2 1

(a) collection of data
(b) presentation of data.
(c) analysisi of data.
(d) explanation of data.

1-j. R is an_____programming language? CO5, K2 1

(a) closed source
(b) GPL
(c) Open source
(d) Definite source

2. Attempt all parts:-

2.a. Difference between labelled and unlabelled data. CO1, K2 2

2.b. Define similarity between R and Python. CO2, K2 2

.

2.c.	Write down the equation of linear regression. CO3, K2	2
2.d.	Explain the difference between the CART and ID3 Algorithms. CO4, K2	2
2.e.	Relate biology and technology together. CO5, K2	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	Discuss the applications of Machine Learning. CO1, K2	6
3-b.	Discuss Reinforcement Learning with example. CO1, K2	6
3-c.	Define vectors and how you can create vector in R? CO2, K2	6
3-d.	Define functions. Explain components of R functions. CO2, K2	6
3.e.	Explain Bayesian Function or Bayes theorem. When can we use Bayes theorem? CO3, K2	6
3.f.	Compare Linear Regression and Decision Tree.	6
3.g.	Explain genetic and non genetic technology.	6

SECTION-C 50

4. Answer any one of the following:-

4-a.	Design a neural model with output Y. Also Calculate the net input to the output Y having inputs 0.6, 0.4 and 0.8 with weights 0.3, 0.7 and -0.5 respectively. CO1, K3	10
4-b.	Explain machine learning? Discuss about learning and machine learning. Choose various types of machine learning. CO1, K2	10

5. Answer any one of the following:-

5-a.	Write down the basic syntax of R. Also application of R. CO2, K2	10
5-b.	Define operators. Explain different types of operators in R. CO2, K2	10

6. Answer any one of the following:-

6-a.	Explain confusion matrix. Why do we calculate confusion matrix. Explain the important terminology for confusion matrix. CO3, K2	10
6-b.	Find the R-squared for the following data: Xi(11,10,2,8,4,20,1,9,5) and Yi(90,45,19,35,25,80) CO3, K3	10

7. Answer any one of the following:-

7-a.	Explain ANCOVA and ANOVA. How they are different? CO4, K2	10
7-b.	Explain Time series Analysis. Also differentiate between ARMA and ARIMA. CO4, K2	10

8. Answer any one of the following:-

8-a.	Explain the advantages of R language over other languages in biotechnology. CO5, K2	10
8-b.	Explain, what happens when biology becomes technology. CO5, K2	10