.

Printed Page:-3 **Subject Code:- BBT0612**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VI - THEORY EXAMINATION (2025 - 2026)

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:-

(1.1) Which ANN model is best suited for image recognition? CO1, K2 1

(a) CNN
(b) RNN
(c) MLP
(d) GAN

(1.2) Which ML model is used for spam email classification? CO1, K2 1

(a) Naive Bayes
(b) Decision Tree
(c) SVM
(d) CNN

(1.3) Which function is used to plot graphs in base R? CO2, K2 1

(a) plot()
(b) hist()
(c) graph()
(d) ggplot()

(1.4) Which of these denotes matrix multiplication? CO2, K2 1

(a) %*
(b) *
(c) %/%
(d) ^

(1.5) What does the posterior probability depend on in Bayes' theorem? CO3, K2 1

(a) Prior only
(b) Likelihood only
(c) Prior and Likelihood
(d) None

(1.6) Which is NOT a measure of central tendency? CO3, K2 1

(a) Mean
(b) Standard Deviation
(c) Median
(d) Mode

(1.7) Why is R preferred over other programming languages in bioinformatics? CO4, K2 1

(a) Supports high-speed hardware control
(b) Has specialized packages for biological data analysis
(c) Offers better gaming graphics
(d) Enables operating system development

(1.8) Which R package supports epidemiological modeling including SIR models? CO4, K2 1

(a) GWASTools
(b) EpiModel
(c) genetics
(d) DESeq2

(1.9) Which method is used to compare group means while controlling for another variable? CO5, K2 1

(a) T-test
(b) ANOVA
(c) ANCOVA
(d) Regression

(1.10) Chi-square test is used to analyze relationships between: CO5, K2 1

(a) Numerical variables
(b) Categorical variables
(c) Time series data
(d) Regression models

2. Attempt all parts:-

(2.1) What is the learning rate in training ANN? CO1, K2 2

(2.2) What function is used to determine the length of a vector in R? CO2, K2 2

(2.3) Explain the significance of the prior probability in Bayesian functions. CO3, K2 2

(2.4) Illustrate how the Bioconductor package in R can be used for biological dataset processing. CO4, K2 2

(2.5) Apply ANCOVA to a dataset comparing two teaching methods while adjusting for prior test scores. CO5, K3 2

SECTION-B 30

Answer any one of the following:-

(3.1)	Write a short note on feedforward neural networks. CO1, K2	6
(3.2)	What is Recurrent Neural Network? Explain its working. CO1, K2	6

Answer any one of the following:-

(3.3)	What are the key differences between vectors and lists in R? CO2, K2	6
(3.4)	What is a factor in R and how is it used in categorical data analysis? CO2, K2	6

Answer any one of the following:-

(3.5)	Explain how logistic regression can be used for binary classification. CO3, K2	6
(3.6)	Interpret the standard normal distribution and its z-scores. CO3, K2	6

Answer any one of the following:-

(3.7)	Perform a correlation test using R between temperature and microbial activity and interpret the result. CO4, K3	6
(3.8)	Cluster a gene expression dataset using k-means in R and interpret the grouping. CO4, K3	6

Answer any one of the following:-

(3.9)	What is the role of entropy or Gini index in splitting decision tree nodes? CO5, K2	6
(3.10)	How would you choose between linear and nonlinear models for a dataset? CO5, K2	6

SECTION-C 50

4. Answer any one of the following:-

(4.1)	What are the primary characteristics of supervised machine learning? CO1, K2	10
(4.2)	Define machine learning and list its major characteristics. CO1, K2	10

5. Answer any one of the following:-

(5.1)	Explain how arrays are defined in R and how they differ from matrices. CO2, K3	10
(5.2)	Describe how string manipulation functions like paste(), strsplit(), and substr() work in R. CO2, K3	10

6. Answer any one of the following:-

(6.1)	Compare and contrast the mean and median in skewed distributions. CO3, K3	10
(6.2)	Explain the concept of odds and odds ratio in logistic regression. CO3, K3	10

7. Answer any one of the following:-

(7.1)	Execute a clinical trial analysis using R, including steps for randomization and power testing. CO4, K3	10
(7.2)	Simulate disease spread using Epi Model in R to model epidemiological patterns. CO4, K3	10

8. Answer any one of the following:-

(8.1)	Use Chi-square and decision tree together to select relevant features before classification. CO5, K3	10
(8.2)	Fit data using both linear and nonlinear models and compare. CO5, K3	10