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. Neural networks learn from the processing numerous _____.(CO1) 1

(a) Labelled example

(b) Unlabelled example

(c) Both

(d) none of above

1-b. In which of the following type of learning networks are trained to provide the correct output using several example units? (CO1) 1

(a) Supervised Learning

(b) Unsupervised Learning

(c) Reinforcement Learning

(d) All

1-c. Which of the following statements is true about factors in R? (CO2) 1

(a) Factors are used to represent continuous variables

(b) Factors are used to represent categorical variables

(c) Factors are always stored as integers

(d) Factors can be used in mathematical calculations

1-d. Data Frames are created using the? (CO2) 1

(a) frame() function

(b) data.frame() function

(c) data() function

(d) frame.data() function

1-e. Find the mean, mode and median of the given sets of data: 5,8,12,17,12,14,6,8, 12, and 10 ? (CO3) 1

(a) 11,12,10

(b) 10,12,13

(c) 11,12,13

(d) 10,12,11

1-f. Mean, median and mode are equals for _____ (CO3) 1

(a) chi square distribution

(b) Poisson distribution

(c) Binomial distribution

(d) Normal distribution

1-g. Decision trees can handle_____ (CO4) 1

(a) high dimensional data

(b) low dimensional data

(c) medium dimensional data

(d) none

1-h. Which of the following is not weakness of Decision Tree ? (CO4) 1

(a) able to generate understandable rules

(b) can be computationally expensive to train

(c) less appropriate for estimation tasks

(d) prone to errors in classification problems with many class

1-i. Diagrams are for _____ (CO5) 1

(a) The use of exports

(b) Better quantitative picture.

(c) better mental appeal.

(d) the use of imports.

1-j. The structure of DNA____ (CO5) 1

(a) Triple helix

(b) Single helix

(c) circular helix

(d) Double helix

2. Attempt all parts:-

2.a. Discuss Classification.(CO1) 2

2.b. Define the data structures in R that is used to perform statistical analyses and create graphs. (CO2) 2

2.c. Identify the mode for the following data set: 21, 19, 62, 21, 66, 28, 66, 48, 79, 59, 28, 62, 63, 63, 48, 66, 59, 66, 94, 79, 19 94 (CO3) 2

2.d. List down some popular algorithms used for deriving Decision Trees and their attribute selection measures. (CO4) 2

2.e. Define the analytics element in biotechnology. (CO5) 2

SECTION B **30**

3. Answer any <u>five</u> of the following:-

3-a. Difference between single layered feedforward and Multilayer feedforward network. (CO1) 6

3-b. Difference between Classification and Regression with examples. (CO1) 6

3-c. Explain R and Python. Explain applications of R. (CO2) 6

3-d. Explain vector, list, matrix. (CO2) 6

3.e. Explain Linear Regression. How we can calculate m-valu and c-value in Linear Regression. (CO3) 6

3.f. Differentiate between random forest and decision tree. (CO4) 6

3.g. Illustrate the reason, why technology used in medical science. (CO5) 6

SECTION C **50**

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

4-a. (a)How AI differ from Machine learning? (b) Discuss the difference between training set and testing set. (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. Answer any <u>one</u> of the following:-

5-a. Explain R and R studio. Write down advantages and disadvantages of R. (CO2) 10

.

5-b. write a short note on list. Explain how to access the elements from list, manipulate the elements and merge the list along with programs. (CO2) 10

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

6-a. Define Logistic Regression.Explain the terminology used for Logistic Regression. Explain Sigmoid Function. (CO3) 10

6-b. Explain Logistic Regression.Explain the terminologies used in Logistic Regression. Derive the equation: p=(1/[(e^-mx+c)+1]) (CO3) 10

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

7-a. Explain Chi-square test. Explain the example for chi square test. When to use chi-square test? (CO4) 10

7-b. Society of 1000 residents has four neighbourhoods, P, Q, R, and S. A sample is taken of 650 residents of the community, having doctors, engineers, and teachers. The null hypothesis states that each one's neighbourhood is independent of their professional division. The categories are: Doctors[P= 90, Q=60, R=104, S=95, total=349], Engineers[P=30,Q=50,R=51,S=20, total=151], Teachers[P=30,Q=40,R=45,S=35, total=150]. Total for P is 150, Q=150, R is 200, S is 150. (CO4) 10

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

8-a. Explain the advanatages and disadvanatages of DNA technology. (CO5) 10

8-b. Explain the relationship between the technology and biology. (CO5) 10