Printed Page:-04 Subject Code:- BCSBS0301

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: III - THEORY EXAMINATION (2024 - 20256)

Subject: Computational Statistics

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 Standard normal distribution, the value of mode is? [CO1] 1

(a) 2

(b) 1

(c) 0

(d) not fixed

1-b. The shape of the Normal Curve is ? [CO1] 1

(a) Bell Shaped

(b) Flat

(c) Circular

(d) Spiked

1-c. Which of the following is an assumption of one-way ANOVA comparing samples from three or more experimental treatments? [CO2] 1

(a) All the response variables within the k populations follow a normal distributions.

(b) The samples associated with each population are randomly selected and are independent from all other samples.

(c) The response variable within each of the k populations have equal variances.

(d) All of the above.

1-d. The error deviations within the SSE statistic measure distances: ? [CO2] 1

(a) within groups

(b) between groups

(c) both (a) and (b)

(d) none of the above

1-e. Which of the following techniques would perform better for reducing dimensions of a data set? [CO3] 1

(a) Removing columns which have too many missing values

(b) Removing columns which have high variance in data

(c) Removing columns with dissimilar data trends

(d) None of the above

1-f. It is not necessary to have a target variable for applying dimensionality reduction algorithms? [CO3] 1

(a) TRUE

(b) FALSE

(c) None of the above

(d) Cant determined

1-g. An interdependence technique primarily used for data reduction, summarization, and the discovery of underlying constructs or latent dimensions is called as: ? [CO4] 1

(a) Discriminant analysis

(b) Multiple regression analysis

(c) Factor analysis

(d) Partial regression analysis

1-h. Which of the following statements is true? [CO4] 1

(a) The correlation matrix will have 1s in the diagonals in PCA and less than 1s in EFA

(b) The correlation matrix will have 1s in the diagonals both in PCA and EFA

(c) The correlation matrix will have less than 1s in the diagonals both in PCA and EFA

(d) The correlation matrix will have 1s in the diagonals in EFA and less than 1 in PCA

1-i. Sentiment Analysis is an example of:? [CO5] 1

1.Regression
2.Classification
3.Clustering
4.Reinforcement Learning

(a) 1 and 3

(b) 1, 2 and 4

(c) 1, 2, 3 and 4

(d) 1 and 2

1-j. Movie Recommendation systems are an example of:? [CO5] 1

1.Classification

2.Clustering
3.Reinforcement Learning
4.Regression

(a) 2 Only
(b) 1 and 2
(c) 2 and 3
(d) 1, 2, 3 and 4

2. Attempt all parts:-

2.a.	Given a normal distribution with x=60, ꭓ75 and σ=9 , find the corresponding z-score? [CO1]	2
2.b.	What are the Assumptions for ANOVA? [CO2]	2
2.c.	Comment whether PCA can be used to reduce the dimensionality of the non-linear dataset? [CO3]	2
2.d.	Define DBSCAN in detail? [CO4]	2
2.e.	Why do you prefer Euclidean distance over Manhattan distance in the K means Algorithm? [CO5]	2

SECTION-B 30

3. Answer any five of the following:-

3-a.	What is Squared Mahalanobis distance and If X follows $N_p(\mu,\Sigma)$, then Squared Mahalanobis distance follows to which distribution? [CO1]	6
3-b.	Define Multivariate analysis also write down the applications of multivariate analysis? [CO1]	6
3-c.	Define canonical correlation and coefficient of determination? [CO2]	6
3-d.	Why was the concept of ANOVA introduced and write down the applications of analysis of variance? [CO2]	6
3.e.	Discuss the concept of Supervised learning with example? [CO3]	6
3.f.	Factor Loading is a data reduction method designed to explain the correlations. Explain the statement? [CO4]	6
3.g.	Differentiate between the different types of machine learning models? [CO5]	6

SECTION-C 50

4. Answer any one of the following:-

4-a. Define likelihood function for Normal distribution.? [CO1] 10

Let X follows If X distributed as $N_2(\mu,\Sigma)$,
where

ꭓ 0 and $\sum = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$

Find the distribution of $Z = \begin{bmatrix} X - Y \\ Y \end{bmatrix}$

4-b. Define Multivariate analysis and How does multivariate analysis differ from 10

univariate and bivariate analysis? What is the process of conducting multivariate analysis? [CO1]

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

5-a. The following figures relate to the production in kg of three varieties I, II, III of wheat shown in 12 plots:? [CO2] 10

Variety I: 14 16 18
Variety II: 14 13 15 22
Variety III:18 16 19 19 20
Is there any significant difference in the production of three varieties?
Given the tabulated value of F(2,9) at 5% level of significance is 4.26.

5-b. Define Discriminant analysis and what is the purpose of discriminant analysis? [CO2] 10

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

6-a. Given data = { 2, 3, 4, 5, 6, 7 ; 1, 5, 3, 6, 7, 8 }. 10
Compute the principal component using PCA Algorithm? [CO3]

6-b. What do you understand by Dimensionality reduction? Discuss the applications of Dimensionality reduction? [CO3] 10

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

7-a. Define the following terms: ? [CO4] 10
a) Observed variable
b) Latent variable
c) Communality
d) Factor Loading
e) Score Matrix

7-b. What is the minimum eigenvalue used for extracting factors? What is it known as? [CO4] 10

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

8-a. What is K means Clustering Algorithm ,explain in detail ,also discuss its applications in brief? [CO5] 10

8-b. What do you understand by Cluster Analysis. Explain it by an example? Discuss the distance measuring method used in clustering? [CO5] 10