Printed Page:-4 **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 (2025 - 2026)

Subject: Computational Statistic

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 inverse of the covariance matrix is known as: (CO1,K1) 1

- (a) Correlation matrix
- (b) Precision matrix
- (c) Identity matrix
- (d) Hessian matrix

1-b. If $X \sim N_2(\mu,\Sigma)$, where $\mu=[2 \;\; 2]'$ and $\Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}$ then the variance of X_2 is (CO1,K3) 1

- (a) 4
- (b) 2
- (c) 3
- (d) 5

1-c. The maximum number of discriminant functions you can obtain is:(CO2,K2) 1

- (a) Number of predictors
- (b) Number of observations
- (c) Number of categorical variables
- (d) Number of groups minus one

1-d. The following is a key benefit of using LDA. (CO2,K2) 1

- (a) Reduces file size
- (b) Enhances noise
- (c) Improves classification performance

(d) Increases contrast

1-e. If the covariance matrix is $\Sigma = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$, then the sum of eigenvalues:(CO3,K2) 1

(a) 5
(b) 6
(c) 4
(d) 2

1-f. PCA can have the following no. of principal components:(CO3,K2) 1

(a) One less than the number of original features
(b) Equal to the number of original features
(c) More than the number of original features
(d) None Of the these

1-g. The total of all eigen values will equal to:(CO4,K2) 1

(a) 0
(b) Covariance
(c) Variance
(d) The number of variables in the analysis

1-h. The factor transformation method is commonly used to make factor loadings interpretable by rotating the factors.(CO4,K2) 1

(a) Principal Component Analysis (PCA)
(b) Varimax rotation
(c) Hierarchical Cluster Analysis
(d) Cronbach's Alpha

1-i. Agglomerative clustering is an example of :(CO5,K1) 1

(a) Hierarchical
(b) Distance-based clustering
(c) Both a and b
(d) K means

1-j. The following type of clustering is suitable for finding clusters of varying shapes and sizes, and does not require specifying the number of clusters beforehand.(CO5,K2) 1

(a) K-Means Clustering
(b) Fuzzy Clustering
(c) DBSCAN
(d) None of them

2. Attempt all parts:-

2.a. Write down the Condition that must satisfy covariance matrix Σ in a multivariate normal distribution (CO1,K2) 2

2.b. Write down the main purpose of discriminant analysis in statistics. (CO2,K2) 2

2.c. Name two methods to decide how many principal components to retain in PCA.(CO3,K1) 2

2.d. Given: Communality for $X_1 = 0.60$, for $X_2 = 0.75$ 2
Find Uniqueness for both variables. (CO4, K3)

2.e. Define overlapping clustering.(CO5,K1) 2

SECTION-B 30

3. Attempt all parts:-

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

3.a.(i) Let X follows $N_2(\mu\Sigma)$ where 6

$$\mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$$

Find the mean and variance of $Y=(3X_1-X_2)$ (CO1,K3)

3.a.(ii) Define Multivariate normal distribution and also explain the quadratic form in Multivariate normal distribution. (CO1,K2) 6

3.b. Answer any one of the following:-

3.b.(i) Define discriminant analysis and Give the reasons to use discriminant analysis rather than regression analysis. (CO2,K2) 6

3.b.(ii) Suppose you have two groups (C and D) and two variables (X and Y): 6

Group	X (mean)	Y (mean)
C	8	12
D	5	9

and $$\Sigma = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}$$

Calculate the linear discriminant function coefficients. (CO2,K3)

3.c. Answer any one of the following:-

3.c.(i) Define PCA and Discuss the real life examples of PCA. (CO3,K2) 6

3.c.(ii) Differentiate between Unsupervised learning and Supervised learning with example. (CO3,K4) 6

3.d. Answer any one of the following:-

3.d.(i) Explain the types of factor rotation in factor analysis.(CO4,K2) 6

3.d.(ii) Define factor analysis and also discuss the need of factor analysis.(CO4,K2) 6

3.e. Answer any one of the following:-

3.e.(i) Differentiate between Partitional and Hierarchical clustering.(CO5,K2) 6

3.e.(ii) Explain the steps of DBSCAN algorithm.(CO5,K2) 6

SECTION-C 50

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

4-a. If X distributed as $N_3(\mu, \Sigma)$, where (CO1,K3) 10

$$\mu = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

Check the following random vectors are independent or not.

i. (X_1, X_3) and
ii. (X_1, X_2)
iii. (X_2, X_3)
iv. (X_1, X_3) and X_2
v. X_1 and $(-X_1 + 3X_2 - 2X_3)$

4-b. If X distributed as $N_3(\mu, \Sigma)$, where (CO1,K3) 10

$$\mu = \begin{bmatrix} 5 \\ 3 \\ 7 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{bmatrix}$$

then find the conditional distribution of $(X_1/X_2 = 1, X_3 = 10)$

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

5-a. Let there are two classes C_1 and C_2 with the following 2D sample: (CO2,K3) 10
C_1:[2,3],[3,3] and C_2:[2,1],[3,2]
Compute the class means and within-class scatter S_w.

5-b. Let there are two classes with 1D data points: (CO2,K3) 10
C_1:x_1=2,x_2=3,x_3=3,x_4=5
C_2 :,x_5=6,x_6=7
Compute the LDA projection vector w (in 1D, it's just the direction).

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

6-a. Explain the terms : 1) Eigen vectors 2)Eigen values 3)Orthogonality 4)Principal Components .(CO3,K2) 10

6-b. Given data : X= { 2, 3, 4, 5, 6, 7 }; Y ={1, 5, 3, 6, 7, 8 }. 10
Compute the Eigenvalues, Eigenvectors and first principal component direction using PCA Algorithm. (CO3,K3)

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

7-a. Define Factor Analysis Model and Discuss the method of determining number of factors in factor analysis.(CO4,K2) 10

7-b. Differentiate between Exploratory Factor analysis and Confirmatory Factor analysis. (CO4,K4) 10

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

8-a. Differentiate between classification and clustering in detail with examples. (CO5,K2) 10

8-b. The given dataset of 8 data points in a two-dimensional space :, 10
A_1: (2, 3),A_2: (3, 2),A_3: (5, 6),A_4: (6, 5), A_5: (8, 6),A_6: (9, 7),A_7: (1, 1) ,A_8: (2, 2)
Perform K-Means clustering with K = 2, starting with the initial cluster centroids at (1, 1) and (2, 2). (CO5,K3)