Reg. No.																

Question Paper Code	13710

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2025

Third Semester

Computer Science and Business Systems

20BSMA305 - COMPUTATIONAL STATISTICS

Regulations - 2020

Duration: 3 Hours Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

Marks K – Level CO

1. For $\sum = \begin{pmatrix} 3 & 6 & 8 \\ 6 & 9 & 2 \\ 8 & 2 & 5 \end{pmatrix}$, what is Σ_{13}? *1 K2 CO1*

   (a) 2 (b) 5 (c) 8 (d) 6

2. The mean for conditional distribution of X_1 given X_2 is *1 K2 CO1*

   (a) $\mu_2 + \rho\sigma_1 \left(\frac{y-\mu_1}{\sigma_2}\right)$ (b) $\mu_1 + \rho\sigma_1 \left(\frac{y-\mu_2}{\sigma_2}\right)$ (c) $\mu_1 + \rho\sigma_2 \left(\frac{y-\mu_2}{\sigma_1}\right)$ (d) $\mu_1 - \rho\sigma_1 \left(\frac{y-\mu_2}{\sigma_2}\right)$

3. The variance inflation factor is given by *1 K1 CO2*

   (a) $\frac{1}{1-R^2}$ (b) $\frac{1}{1+R^2}$ (c) $\frac{2}{1-R^2}$ (d) $\frac{1}{R^2-1}$

4. The estimate of β is *1 K1 CO2*

   (a) $\hat{\beta} = (X^T X)^{-1} Y^T Y$ (b) $\hat{\beta} = (X^T X)^{-1} X^T X$

   (c) $\hat{\beta} = (X^T X)^{-1} X^T Y$ (d) $\hat{\beta} = (X^T Y)^{-1} X^T Y$

5. Multicollinearity of dependent variables in MANOVA can prevented by removing one of the strongly correlated pairs if the r between them is equal or greater than *1 K1 CO3*

   (a) 1 (b) 0.9 (c) 0 .8 (d) 0.7

6. Linear discriminant analysis is used in *1 K1 CO3*

   (a) Unsupervised learning (b) Supervised learning

   (c) Reinforcement learning (d) None of the above

7. H-plot is used to determine *1 K1 CO4*

   (a) The size of the dataset (b) The accuracy of the data

   (c) The number of principal components to retain (d) The standard deviation of the data

8. Which of the following is a measure of the amount of variance explained by a principal component in PCA? *1 K1 CO4*

   (a) Eigen value (b) Covariance (c) Correlation (d) Mean absolute deviation

9. Which clustering algorithm is based on the concept of centroids? *1 K1 CO5*

   (a) K-Means (b) DBSCAN (c) Agglomerative (d) Mean-Shift

10. Which clustering algorithm is based on the concept of minimizing the within-cluster variance? *1 K1 CO5*

    (a) K-Means (b) DBSCAN (c) Agglomerative (d) Mean-Shift

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11. Let X ~ $N\left(\begin{pmatrix} 5 \\ 10 \end{pmatrix}, \begin{pmatrix} 16 & 12 \\ 12 & 36 \end{pmatrix}\right)$ such that X = $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$. Find the mean of the distribution. *2 K2 CO1*

12. Let X= $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ with μ = $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$ respectively. Prove that the random variable Y = X_1+ X_2 has normal distribution with mean 3 and variance 7. 2 K2 CO1

13. Let X~$N_5(\mu, \Sigma)$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$. 2 K1 CO1

14. Define Multi collinearity and the method used to detect it. 2 K1 CO2

15. Write the formula to find the residual of Multivariate linear regression? 2 K1 CO2

16. What is multivariate linear regression and give an example? 2 K1 CO2

17. Calculate the APER for the confusion matrix $\begin{pmatrix} 4 & 3 \\ 3 & 2 \end{pmatrix}$ for the population π_1 and π_2 with 13 and 10 samples respectively. Hence find the percentage of APER. 2 K2 CO3

18. Define Fisher's Discriminant for several populations. 2 K1 CO3

19. What is the formula for proportion of total population variance due to k^{th} principle component? 2 K1 CO4

20. Convert the covariance matrix Σ= $\begin{pmatrix} 1 & 16 \\ 16 & 225 \end{pmatrix}$ to correlation matrix. 2 K2 CO4

21. Define Cluster Analysis. 2 K1 CO5

22. Define k-mean clustering. 2 K1 CO5

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23. a) Let X = $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ be a normal random vector with the following mean vector and covariance matrix μ = $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, Cov = $\begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$.Let also A = $\begin{pmatrix} 2 & 1 \\ -1 & 1 \\ 1 & 3 \end{pmatrix}$, b = $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$, 11 K3 CO1

Y= $\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$ = AX+b

(a) Find P(X_2 > 0).

(b) Find the expected value vector of Y, μ_y = E_y.

(c) Find the covariance matrix of Y.

(d) Find P($Y_2 \leq 2$).

OR

b) Let X be distributed as $N_3(\mu, \Sigma)$ where $\mu = (1, -1, 2)$ and $\Sigma = \begin{pmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ 1 & 0 & 2 \end{pmatrix}$ which of the random variables are independent? Explain. 11 K3 CO1

(i) X_2 and X_3.

(ii) (X_1, X_3) & X_2.

(iii)(X_2, X_3) & X_1.

(iv) X_1 and $X_1 + 2X_2 - 3X_3$.

24. a) Given the mean vector and covariance matrix of Y,X_1,X_2 11 K3 CO2

μ=$\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}$=$\begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix}$, Σ = $\begin{pmatrix} 10 & 1 & -1 \\ 1 & 7 & 3 \\ -1 & 3 & 2 \end{pmatrix}$.

Determine the following

(i) The best linear predictor β_0+$\beta_1 X_1$+$\beta_1 X_2$.

(ii) Mean square Error.

(iii) Multiple correlation coefficient.

(iv) Also verify that mean square error equals σ_{yy}(1-ρ_y^2(x)).

OR

b) Find y value for the following data by least square method: *11 K3 CO2*

Y	-2.7	5.5	3.5	11.5	5.7
X_1	4	6	5	8	3
X_1	5	8	7	4	2

25. a) Consider the two data sets $X_1 = \begin{pmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{pmatrix}$ & $X_2 = \begin{pmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{pmatrix}$ and for which *11 K3 CO3*

$\overline{X_1} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \overline{X_2} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$.

Spooled $= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.

a) Calculate the Linear Discriminant function.

b) Classify the observation $x_0{}' = (2, 7)$ as population π_1 or π_2 using rule with equal prior and equal cost.

c) Classify the observation $x_0{}' = (1, 8)$ as population π_1 or π_2 using rule with equal prior probabilities $p_1 = \frac{1}{3}$, $p_2 = \frac{2}{3}$ and cost of misclassification $C(1/2) = 40, C(2/1) = 55$.

OR

b) Construct a MANOVA table for the following *11 K3 CO3*

$$\begin{pmatrix} \begin{pmatrix} 9 \\ 3 \end{pmatrix} & \begin{pmatrix} 6 \\ 2 \end{pmatrix} & \begin{pmatrix} 9 \\ 7 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 4 \end{pmatrix} & \begin{pmatrix} 2 \\ 0 \end{pmatrix} & \\ \begin{pmatrix} 3 \\ 8 \end{pmatrix} & \begin{pmatrix} 1 \\ 9 \end{pmatrix} & \begin{pmatrix} 2 \\ 7 \end{pmatrix} \end{pmatrix}$$

26 a) Construct the Principle Components for the following sample data: *11 K3 CO4*

Features	s_1	s_2	s_3	s_4
X_1	4	8	13	7
X_2	11	4	5	14

OR

b) If $\Sigma = \begin{pmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{pmatrix}$ is a population covariance matrix and L= $\begin{pmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{pmatrix}$ is a factor loaded matrix then build its specific variance matrix Ψ and the communality of the variables X_1, X_2, X_3 and X_4. *11 K3 CO4*

27. a) Construct the clusters using Complete Linkage procedure. Use Euclidean distance and draw the Dendogram. *11 K3 CO5*

Points	A	B	C	D	E
X	2	6	2	2	5
Y	5	5	4	2	4

OR

b) Consider the matrix of distance D= $\begin{pmatrix} 0 & & & & \\ 4 & 0 & & & \\ 6 & 9 & 0 & & \\ 1 & 7 & 10 & 0 & \\ 6 & 3 & 5 & 8 & 0 \end{pmatrix}$. Construct the Clusters for five items using the single, complete and average linkage hierarchical procedure. Also draw the Dendrogram. *11 K3 CO5*

28. a) (i) If X $\sim N_3(\mu, \Sigma)$, $\mu = \begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix}$ $and\ \Sigma = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{pmatrix}$, then find *6 K3 CO4*

(i) P(X_1 > 6).

(ii)P($5X_2 + 4X_3 > 70$).

(iii) P($4X_1 - 3X_2 + 5X_3 < 80$).

(ii) Consider the matrix of distance D= $\begin{pmatrix} 0 & & & & \\ 4 & 0 & & & \\ 6 & 9 & 0 & & \\ 1 & 7 & 10 & 0 & \\ 6 & 3 & 5 & 8 & 0 \end{pmatrix}$. Construct the Clusters for five items using the Average linkage hierarchical procedure. Also draw the Dendrogram. *5 K3 CO5*

OR

b) (i) In a consumer preference study, a random sample of customers were asked to rate several attributes of a new product R = $\begin{pmatrix} 1 & 0.02 & 0.96 & 0.42 & 0.01 \\ 0.02 & 1 & 0.13 & 0.71 & 0.85 \\ 0.96 & 0.13 & 1 & 0.5 & 0.11 \\ 0.42 & 0.71 & 0.5 & 1 & 0.79 \\ 0.01 & 0.85 & 0.11 & 0.79 & 1 \end{pmatrix}$ be the correlation matrix. The eigenvalues more than unity are 2.85 and 1.81. *6 K3 CO4*

$F_1 = (0.56\ \ 0.78\ \ 0.65\ \ 0.94\ \ 0.80)^T$

$F_2 = (0.82\ \ -0.53\ \ 0.75\ \ -0.10\ \ -0.54)^T$ are the loading factors. Construct the residual matrix.

(ii) Consider the matrix of distance D= $\begin{pmatrix} 0 & 9 & 3 & 6 & 11 \\ 9 & 0 & 7 & 5 & 10 \\ 3 & 7 & 0 & 9 & 2 \\ 6 & 5 & 9 & 0 & 8 \\ 11 & 10 & 2 & 8 & 0 \end{pmatrix}$. Construct the Clusters for five items using the Single linkage hierarchical procedure. Also draw the Dendrogram. *5 K3 CO5*