Register Number:……………..

KUMARAGURU
college of technology
character is life

M.E / M.TECH DEGREE EXAMINATIONS: NOV/ DEC 2024

(Regulation 2024)

First Semester

DATA SCIENCE

24MAI505: Mathematics for Data Science

COURSE OUTCOMES

CO1: Apply the properties of vector spaces, subspaces, and linear transformations to solve problems related to matrix operations and eigenvalue computation.

CO2: Apply concepts of probabilities, expected values, and Baye's theorem to discrete and continuous random variables in real-world scenarios.

CO3: Apply marginal and conditional distributions and apply the Central Limit Theorem and normal distribution to interpret statistical data.

CO4: Evaluate the reliability of regression models using correlation and regression techniques by fitting curves to data using the method of least squares.

CO5: Apply the multivariate concepts, including principal component analysis (PCA), to reduce data dimensionality and compute covariance and correlation matrices.

CO6: Apply Lagrange multipliers to solve constrained and unconstrained optimization problems to determine optimal solutions in engineering contexts.

Time: Three Hours **Maximum Marks: 100**

PART A (4*20 = 80 Marks)

1. a) Find eigenvalues and eigenvectors of the matrix — 8 — CO1 — [K3]

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

Scenario: In a data compression project for a company that handles large image datasets, Singular Value Decomposition (SVD) is being utilized to reduce the dimensionality of matrices while retaining critical information.

b) Given a simplified version of an image data matrix $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$, perform the singular value decomposition expressing it in the form $A = U\Sigma V^T$. — 10 — CO1 — [K4]

c) Explain the importance of SVD in this scenario. — 2 — CO1 — [K4]

2. a) The contents of urns I, II, III are as follows: 1 white, 2 black and 3 red balls; 2 white, 1 black and 1 red balls; 4 white, 5 black and 3 red balls respectively. One urn is — 6 — CO2 — [K3]

chosen at random and one ball is drawn from it. It happens to be white. What is the probability that it come from urns I, II, or III?

b) A discrete random variable has the following probability distribution 8 CO2 [K3]

x:	0	1	2	3	4	5	6	7	8
$P(x)$:	a	$3a$	$5a$	$7a$	$9a$	$11a$	$13a$	$15a$	$17a$

Find (i) the value of a (ii) $P(2 \le X < 6)$

(iii) $P(X > 3)$ (iv) distribution function of X.

c) If a random variable X has density function 6 CO2 [K3]

$$f(x) = \begin{cases} \frac{x}{6} & \text{for } x = 1,2,3 \\ 0 & \text{otherwise} \end{cases}$$

find mean, variance, S.D. of X.

3. a) If X and Y are two random variables having joint density function. 8 CO3 [K3]

$$f(x, y) = \begin{cases} \frac{1}{8}(6 - x - y): & 0 < x < 2,\ 2 < y < 4 \\ 0 & : otherwise \end{cases}$$

(i) Find the marginal densities of X and Y.

(ii) Check whether X and Y are independent

(iii) Find $P(X < 1 \cap Y < 3)$, $P(X < 1|Y < 3)$

b) 2. Suppose the heights of men of a certain country are normally distributed with average 68 inches and standard deviation 2.5, find the percentage of men who are 6 CO3 [K4]

(i) between 66 inches and 71 inches in height

(ii) approximately 6 feet tall (ie, between 71.5 inches and 72.5 inches)

c) The lifetime of a particular variety of electric bulbs may be considered as a random variable with mean 1200 hours and standard deviation 250 hours. Using central limit theorem find the probability that the average life time of 60 bulbs exceeds 1250 hours. 6 CO3 [K3]

4. a) Find the mean matrix, covariance matrix, standard deviation matrix and correlation matrix for two random variables X_1 and X_2 whose joint mass function is given by 10 CO5 [K3]

$X_1 \ddots X_2$	0	1
-1	0.24	0.06
0	0.16	0.14
1	0.40	0.0

b) Solve the NLPP: optimize $f(x,y,z) = 4x^2 + 2\,y^2 + z^2 - 4xy$ subject to the constraints $x + y + z = 15,\quad 2x - y + 2z = 20,$ 10 CO6 [K3]

$$x, y, z \geq 0$$

Answer any ONE Question

PART B (1* 20 = 20 Marks)

5. a) Find the MLE of λ in the Poisson distribution. 5 CO4 [K3]

b) Calculate the rank coefficient of correlation for the following data: 7 CO4 [K3]

X:68	64	75	50	64	80	75	40	55	64
Y:62	58	68	45	81	60	68	48	50	70

c) Fit a second-degree curve to the data: 8 CO4 [K3]

x:	1	2	3	4	5	6	7	8	9
y:	2	6	7	8	10	11	11	10	9

OR

6. a) From the following data, obtain the two regression equations: 8 CO4 [K3]

Sales(X):	91	97	108	121	67	124	51	73	111	57
Purchase(Y): 71	75	69	97	70	91	39	61	80	47

b) Fit a straight line to the following data and find the value of y when x = 3.5 7 CO4 [K3]

x:	1	2	3	4	5	6
y:	0.5	2.5	2.0	4.0	3.5	6.0

c) Obtain the MLE of θ in $f(x,\theta) = (1+\theta)x^{\theta}$, $0 < x < 1$. 5 CO4 [K3]

****