KUMARAGURU college of technology
character is life

Register Number:……..…

M.TECH DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

First Semester

DATA SCIENCE

P18MAT1107**:** Mathematics for Data Science

COURSE OUTCOMES

CO1: Check linear dependency of vectors and identify Eigen values, Eigen vectors and derivative of a matrix, which will form the basis for Principal Component Analysis.

CO2: Apply the concept of probability and random variables, which will help in learning Bayesian classifiers.

CO3: Apply the concepts of two-dimensional random variables, central limit theorem and multivariate normal distribution, which lay the foundation for Machine Learning.

CO4: Fit curves to given data, analyse the correlation and regression and find the maximum likelihood estimate Fit curves to given data, analyse the correlation and regression and find the maximum likelihood estimate

CO5: Learn and apply multivariate analysis necessary for Principal Component Analysis.

CO6: Determine the extreme values of functions without constraint, and with equality Constraints

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

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. Examine the two statements carefully and select the answer using the codes given below: CO3 [K_4]

Assertion (A): If the value of x for normal distribution is 95, the mean of the normal distribution is 65 and the standard deviation is 25, then the value of the standardized normal variate is 1.2.

Reason (R): Standard normal variate $Z = \frac{X - \mu}{\sigma}$

a) Both A and R are individually true, and R is the correct explanation of A

b) Both A and R are individually true, but R is not the correct explanation of A

c) A is true but R is false

d) A is false but R is true.

2. Which of the following can be probability density functions? CO2 [K_3]

(1) $f(x) = 3x^2,\ 0 < x < 1$

(2) $f(x) = 4x^3,\ 1 < x < 2$

(3) $f(x) = x,\ \ 0 < x < \sqrt{2}$

(4) $f(x) = 6x(1-x),\ 0 < x < 1$

a) 1,2,3 b) 1,3,4

c) 2,3,4 d) 1,2,4

3. Eigenvalues of a Variance-Covariance matrix is always CO5 [K_1]

a) Non-zero b) Non-negative

c) Positive d) Negative

4. Match List I with List II and select the answer using the codes given below. Consider the random experiment of selecting a card from a standard pack of cards. CO2 [K_3]

List I	List II
A. King	1. $\frac{4}{13}$
B. Spade	2. $\frac{2}{13}$
C. King or queen	3. $\frac{1}{4}$
D. King or spade	4. $\frac{1}{13}$

a) A – 4, B - 3, C - 2, D – 1 b) A – 1, B - 2, C - 3, D - 4

c) A – 4, B - 3, C - 1, D – 2 d) A – 1, B - 2, C - 4, D - 3

5. The steps involved in finding the extreme values of a function f(x) are. CO6 [K_2]
1) Identify how many variables are involved in that function
2) Check if any constraint is given
3) Apply the necessary conditions to find the basic variable values.
4) Use sufficient conditions to evaluate the extremum

a) 1-3-2-4 b) 2-1-4-3

c) 1-2-3-4 d) 2-1-3-4

6. Match list I with list II for a given matrix of order ‘n’. CO1 [K_1]

List I	List II
A. Sum of the eigenvalues	i) Determinant of A
B. Product of the eigenvalues	ii) Real values
C. Eigenvalues of the unit matrix	iii) Trace of the matrix A
D. Eigenvalues of real symmetric matrix	iv) Equal

a) A- iii; B- iv; C - ii; D -i b) A - ii; B - i; C - iv; D - iii

c) A - ii; B - iii; C - i; D -iv d) A - iii; B -i; C - iv; D - ii

7. If $r = 1$ and $b_{xy} = 2, SD_x = 2$, values of SD_y & b_{yx} are CO4 [K3]

a) 0.5, 1 b) 1, 1

c) 1, 1.5 d) 1, 0.5

8. Which of the following statements is correct? CO4 [K1]

1. Maximum likelihood estimators are surely Unbiased.
2. Maximum likelihood estimators are consistent.
3. Maximum likelihood estimators are most efficient.
4. Maximum likelihood estimators are sufficient if any exist.

a) 1, 2, 3 b) 2, 3, 4

c) 1, 3, 4 d) 1, 2, 4

9. Principal Component Analysis is done to CO5 [K2]

a) Find the mean values of the various variables in the data b) Reduce the dimensionality of the data

c) Find the pairwise correlation between the variables d) Obtain the covariance matrix

10. Which of the following statements are correct? CO3 [K1]

1. F (-∞,∞)=1
2. $P[x_1 < X < x_2, y_1 < Y < y_2] = \int_0^{\infty} \int_{y_1}^{y_2} f(x,y)dydx$
3. $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)dydx$
4. $P[X = x_i / Y = y_j] = \frac{P[X = x_i \cup Y = y_j]}{P[Y = y_j]}$

a) 1,2,3 b) 2, 3, 4

c) 1,3,4 d) 1, 2, 4

PART B (10 x 2 = 20 Marks)

11. Find the inner product and outer product of the vectors $[1 \quad 3 \quad -5]$ *and* $[4 \quad -2 \quad -1]$. CO1 [K2]

12. Compute the Jacobian matrix of the vector function CO1 [K2]

$u = 2x - y + 3z, v = 2x - y - z, w = 2x - y + z$

13. A box contains 4 bad and 6 good tubes. Two are drawn from the box at a time. One of them is tested and found to be good. What is the probability that the other one is also good? CO2 [K3]

14. If a random variable X takes the value 1, 2, 3, 4 such that $2P(X = 1) = 3P(X = 2) = P(X = 3) = 5P(X = 4)$, find the probability distribution of X. CO2 [K2]

15. The joint probability mass function of (X,Y) is given by $P(x, y) = k(2x + 3y)$; CO3 [K2]

$x = 0,1,2,\ y = 1,2,3$. Find the value of k.

16. Let X_1 and X_2 be independent standard normal random variables with zero mean. Let $Y_1 = 2X_1 + X_2$ & $Y_2 = X_1 - X_2$. Find $\text{Cov}(Y_1, Y_2)$. CO3 [K3]

17. Calculate the rank correlation coefficient between marks in the selection test (X) and the proficiency test (Y) of 5 recruits. CO4 [K3]

X	12	15	17	20	22
Y	34	22	18	25	30

18. Find the MLE of θ, given $f(x,\theta) = \theta x^{\theta-1}, 0 < x < 1, 0 < \theta < \infty$. CO4 [K3]

19. Explain a random vector. CO5 [K1]

20. Examine the function $f(x) = x^3 - 15x^2 + 10x + 100$ for extreme points. CO6 [K3]

PART C (10 x 5 = 50 Marks)

21. Find the Eigenvalues and Eigenvectors of the matrix $A = \begin{bmatrix} 2 & -3 & 1 \\ 3 & 1 & 3 \\ 5 & 2 & -4 \end{bmatrix}$. CO1 [K4]

22. Show that the vectors $X1 = (1, 2, 1)$; $X2 = (4, 1, 2)$; $X3 = (6, 5, 4)$; $X4 = (-3, 8, 1)$ are linearly dependent. Hence find the relationship between them. CO1 [K4]

23. The contents of urns I, II, and III are: 1 white, 2 black, and 3 red balls; 2 white, 1 black, and 1 red ball; 4 white, 5 black, and 3 red balls respectively. One urn is chosen randomly, and two balls are drawn from it. They happen to be white and red. What is the probability that they come from urns I, II, or III? CO2 [K4]

24. Find the mean, variance of the random variable X which has the following density function CO2 [K3]

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

25. In a test of 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average life of 2040hrs and SD of 60 hrs. Estimate the number of bulbs likely to burn for CO3 [K4]
(i) more than 2150 hrs
(ii) less than 1950hrs
(iii) more than 1920 hrs but less than 2160hrs.

26. Fit a second-degree curve to the data: CO4 [K3]

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

27. Compute the mean vector and covariance matrix for two random variables X_1 and X_2 whose joint mass function is given by CO5 [K4]

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

28. Calculate the population principal component of the random variables X_1, X_2 whose covariance matrix is $\Sigma = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix}$. CO5 [K4]

29. A rectangular box, open at the top is to have volume of 32 cubic meters. Find the dimension of the box requiring least material for its construction. CO6 [K3]

30. Solve the following NLPP: CO6 [K3]

Optimize $f(x, y, z) = x^2 + y^2 + z^2$ subject to the constraints

$$x + y + 3z = 2, \quad 5x + 2y + z = 5,$$

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

Answer any TWO Questions

PART D (2 x 10 = 20 Marks)

31. A discrete random variable has the following probability distribution 10 CO2 [K4]

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) $P(2 \leq X < 6)$

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

32. a) Two fair dice are tossed 600 items. Let X denote the number of times a total of 7 occurs. Using Central Limit Theorem, find $P(90 \leq X \leq 110)$. 5 CO3 [K4]

b) Given $f_{xy}(x, y) = \begin{cases} Cx(x - y), 0 < x < 2, -x < y < x \\ 0, \quad otherwise \end{cases}$. 5 CO3 [K4]

i) Evaluate C ii) Find $f_x(x)$ iii) Find $f_{\frac{y}{x}}\left(\frac{y}{x}\right)$

33. Obtain the equations of two lines of regression for the following data using method of least square and also estimate the value of Y when X=55. 10 CO4 [K4]

X:	60	34	40	50	45	40	22	43	42	64
Y:	75	32	33	40	45	33	12	30	34	51

****