KUMARAGURU
college of technology
character is life

Register Number:……..…

M.TECH DEGREE EXAMINATIONS: DEC 2022

(Regulation 2018)

First Semester

DATA SCIENCE

P18MAT1107**:** Mathematics for Data Science

(Statistical Tables Required)

COURSE OUTCOMES

CO1: Check linear dependency of vectors and identify eigenvalues, eigenvectors 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

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: CO1 [K_3]

   Assertion (A): The vectors $(2,-1,0), (4,1,1), (8,-1,1)$ are linearly dependent.

   Reason (R):The value of the determinant $\begin{vmatrix} 2 & -1 & 0 \\ 4 & 1 & 1 \\ 8 & -1 & 1 \end{vmatrix} = 0$

   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. If two of the eigenvalues of $A = \begin{pmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{pmatrix}$ are 3 and 15, then $|A| =$ CO1 [K_3]

   a) 18
   b) 45
   c) 1
   d) 0

3. Three coins are tossed together. Which of the following probabilities are equal to 3/8? CO2 [K_3]
   1. The probability that there are exactly two heads.
   2. The probability that there are exactly two tails.
   3. The probability that there is exactly one head.
   4. The probability that there is exactly one tail.

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

4. Match List I with list II CO2 [K3]

Given $P(A) = \frac{1}{3}, P(B) = \frac{3}{4}, P(A \cup B) = \frac{11}{12}$

List I	List II
A. $P(\overline{A})$	i. 1/2
B. $P(A \cap B)$	ii. 1/6
C. $P(A\|B)$	iii. 2/3
D. $P(B\|A)$	iv. 2/9

a) A-ii, B-iii, C-iv, D-i b) A-iii, B-ii, C-i, D-iv
c) A-ii, B-iii, C-i, D-iv d) A-iii, B-ii, C-iv, D-i

5. Let $X_1, X_2, ..., X_n$ be Poisson variates with parameter λ = 2. Let CO3 [K2]

$S_n = X_1 + X_2 + ... + X_n$ where n = 75. Then S_n follows normal distribution with mean and variance respectively

a) 75, 75 b) 75, $75\sqrt{2}$
c) 150,150 d) 150,$150\sqrt{2}$

6. The joint p.d.f. of (X, Y) is CO3 [K2]

x \ y	0	1	2	3
0	$\frac{1}{21}$	$\frac{3}{14}$	$\frac{1}{7}$	$\frac{1}{84}$
1	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{14}$	0
2	$\frac{1}{21}$	$\frac{1}{28}$	0	0

Then $P(Y = 0)=$

a) 15/21 b) 35/84
c) 5/21 d) 15/84

7. Which of the following is not a property of maximum likelihood estimators? CO4 [K1]

a) Sufficiency b) Efficiency
c) Consistency d) Unbiasedness

8. Write the correct sequence for fitting a straight line $y = ax + b$ to a given set of x and y values, by the method of least squares. CO4 [K2]

1. Substitute the values in the normal equations.
2. Find $\sum x$, $\sum y$, $\sum x^2$, $\sum y^2$, $\sum xy$
3. Solve to find a and b.
4. Substitute these values in $y = ax + b$

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

9. Principal component analysis is used to CO5 [K1]

a) Find correlation b) Find covariance matrix
c) Reduce dimensionality of data d) To analyse the data statistically

10. The stationary points of the function $y = 4x^4 - x^2 + 5$ are CO6 [K3]

a) $0, 1, -4$ b) $0, -1, 4$

c) $0, \pm\sqrt{\frac{1}{2}}$ d) $0, \pm\sqrt{\frac{1}{8}}$

PART B (10 x 2 = 20 Marks)

11. If $A = \begin{pmatrix} 5 \\ -2 \\ 3 \end{pmatrix}, B = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$, find $A.B$ and $A \otimes B$. CO1 [K2]

12. A company has two plants to manufacture scooters. Plant I manufactures 80% of the scooters and plant II manufactures the remaining 20%. The probability that a manufactured scooter is of good quality is 0.85 in Plant I and and 0.65 in plant II. What is the probability that a randomly chosen scooter is of good quality? CO2 [K3]

13. A is known to hit the target in 2 out of 5 shots whereas B is known to hit the target in 3 out of 4 shots. Find the probability of the target being hit when both try independently? CO2 [K3]

14. State Central limit theorem for (i) mean and (ii) sum of n independent and identically distributed random variables. CO3 [K1]

15. Let (Y_1, Y_2) follow bivariate normal distribution, where $Y_1 = 2X_1 + X_2, Y_2 = X_1 - X_2$, X_1 and X_2 are independent standard normal variates. Find the covariance between Y_1 and Y_2. CO3 [K3]

16. It is required to fit a normal distribution to a given frequency distribution with the following features: Total frequency=100, mean = 165.5, standard deviation = 15.26. Determine the theoretical frequency of the class $120 \leq X \leq 130$. CO3 [K3]

17. Find the rank correlation between X and Y given: CO4 [K3]

X:	1	2	3	4	5
Y:	4	5	3	1	2

18. Find the maximum likelihood estimate of the parameter θ in the p.d.f $f(x;\theta) = \theta\, e^{\theta x}, 0 \leq x < \infty, \theta > 0$, of the exponential distribution. CO4 [K4]

19. Given the covariance matrix $\Sigma = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{pmatrix}$, find the standard deviation matrix $V^{\frac{1}{2}}$. CO5 [K3]

20. An engineering firm is designing a bridge and wants to minimize the amount of steel used in the construction. The steel used in the construction is modelled by the function $S(x) = x^3 - 6x^2 + 9x + 26$, where x is the width of the bridge in meters. What is the minimum amount of steel required to build the bridge? CO6 [K4]

PART C (10 x 5 = 50 Marks)

21. Determine the eigenvalues and eigenvectors of $A = \begin{pmatrix} 3 & -1 & 1 \\ -1 & 5 & -1 \\ 1 & -1 & 3 \end{pmatrix}$ CO1 [K3]

22. 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 chosen at random 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]

23. A discrete random variable has the following probability distribution CO2 [K3]

x :	0	1	2	3	4	5
$P(x)$:	a	3a	5a	7a	9a	11a

Find (i) the value of 'a' (ii) $P(2 \leq X < 4)$ (iii) distribution function of X.
(iv) Mean of X

24. The joint probability function of (X, Y) is given by $P(x, y) = k(2x + 3y)$, $x = 0,1,2$; $y = 1,2,3$. Find CO3 [K3]
(a) k. (b) The marginal distributions of X and Y. (c) $E(X)$ and $Var(X)$.

25. Let $X_1, X_2, ..., X_n$ be independent identically distributed random variables with mean=2 and $\sigma^2 = \frac{1}{4}$. Find $P(192 < S_n < 210)$ where $S_n = X_1 + X_2 + ... + X_{100}$ — CO3 [K4]

26. Fit a parabola, by the method of least squares, to the following data, and estimate y when $x = 6$. — CO4 [K3]

x:	1	2	3	4	5
y:	5	12	26	60	97

27. In a random sample from a normal distribution $N(\mu, \sigma^2)$, find the maximum likelihood estimate for μ when σ^2 is known — CO4 [K4]

28. If X_1, X_2 have the joint pmf $p(x_1, x_2) = \frac{x_1+2x_2}{18}$, $x_1 = 1,2;\ x_2 = 1,2$, find — CO5 [K3]
    (i) the marginal distributions of X and Y
    (ii) Mean vector (iii) Variance-covariance matrix

29. Solve the NLPP using Lagrange's method: — CO6 [K4]
    Maximize $z = 4x_1 - x_1^2 + 8x_2 - x_2^2$
    subject to $x_1 + x_2 = 2, x_1 \geq 0, x_2 \geq 0$

30. Determine the maximum or minimum point of — CO6 [K4]
    $f(x) = 4x_1^2 + 3x_2^2 + x_3^2 - 6x_1x_2 + x_1x_3 - \frac{x_1}{2} - 2x_2 + 15.$

Answer any TWO Questions
PART D (2 x 10 = 20 Marks)

31. (a) Calculate Karl Pearson's coefficient of correlation between price (X in Rupees) and supply (Y in Kg.) of a commodity from the following data: — 5 CO4 [K3]

X:	17	18	19	20	21	22
Y:	38	37	38	33	32	33

    (b) The mean yield for one-acre plots is 662 kgs with S.D 32. Assuming normal distribution, how many one-acre plots in a batch of 1000 plots would you estimate to yield. (i) Over 700 kgs (ii) below 650 kgs? — 5 CO3 [K4]

32. (a) For the continuous distribution with p.d.f $f(x) = k(x - x^2),\ 0 \leq x \leq 1$, find k, mean and standard deviation. — 5 CO2 [K3]
    (b) Obtain the principal components of the standardised variables $Z = (z_1, z_2)$ from the variance-covariance matrix $\Sigma = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}$. Which is the first principal component? What proportion of the total population variance does it explain? — 5 CO5 [K4]

33. (a) In a partially destroyed laboratory record of an analysis of a correlation data, the following results only are legible. — 5 CO4 [K4]
    Variance of $X = 9$
    Regression equations are $8x - 10y + 66 = 0,\ 40x - 18y = 214$
    Find (a) The mean values of X and Y
    (b) The coefficient of correlation between X and Y.
    (c) The standard deviation of Y

    (b) The two dimensional random variable (X, Y) has the joint p.d.f, — 5 CO3 [K4]

$$f(x, y) = \begin{cases} x + y, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

    Compute (i) the marginal density functions $f_X(x)$ and $f_Y(y)$ (ii) the conditional density functions $f(y|x)$ and $f(x|y)$.

****