Roll Number:

Thapar Institute of Engineering and Technology, Patiala

Department of Computer Science and Engineering

END SEMESTER EXAMINATION

B. E. (Second Year): Semester-III (2020/21) (CSBS) **Course Code: UCT304**

January 27, 2021 **Name: COMPUTATIONAL STATISTICS**

Time: 2 Hours **Wednesday, 11.00 – 1.00 Hrs**

M. Marks: 50 **Name of Faculty: Dr. Karamjeet Singh**

Note: Attempt any 5 questions.
Assume missing data, if any, suitably.

Q-1 Derive MLE's for Multivariate statistics having two different parameters $\mu, \sum$. Let, $X_i \cong (X_{i1}, X_{i2}, \dots, X_{in})^T$ be the i.i.d random variables from MVN$(\mu, \sum)$. Also, find MLE for $\mu, \sum$. (7)

Q-2 Define the following term with mathematical formulations and matrix form:-

(a) Multivariate Linear Regression (b) Multivariate Normal Distribution (c) Conditional Distribution (d) Multiple Linear Regression (e) Quadratic form (10)

Q-3 The marks obtained by a student are dependent on her/his study time. Given the study time in minutes and marks out of 2000, find the relationship between study time and marks using the concept of linear regression. Also, predict the marks for a student if he/she studied for 790 minutes. (5)

S. No	Study Time (Min)	Marks Obtained
1	350	520
2	1070	1600
3	630	1000
4	890	850
5	940	1350
6	500	490

Q-4 Apply K-mean algorithm on the following dataset with assumptions k=3 (i.e. 3 clusters), iterations=2, and initial clusters centres are $C_1(2), C_2(16)$ *and* $C_3(38)$. (7)

Data: 2,4,6,3,31,12,15,16,38,35,14,21,23,25,30.

Q-5 Consider the following vector and matrix notations $X \simeq N_3(\mu, \Sigma)$; $\mu = \begin{pmatrix} 5 \\ 3 \\ 7 \end{pmatrix}$, (7)

$\Sigma = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{pmatrix}$, Find $P(X_1 > 8)$ (ii) Find $P(4X_1 - 3X_2 + 5X_3 < 63)$

Q-6 Consider the following table which shows that the weight X_1 in pounds (lb), height X_2 in inch (in) and ages X_3 in years for boys residing in Patiala city is provided to us. (7)

Weight (X_1)	64	71	53	67	55	58	77	57	56	51	76	68
Height (X_2)	57	59	49	62	51	50	55	48	52	42	61	57
Age (X_3)	8	10	6	11	8	7	10	9	10	6	12	9

(a) Find the required regression equation for above problem.

(b) Estimate the weight of a boy who is 9 years old and 54 in tall.

Q-7 (a) Define Multicollinearity by taking the example of dataset of your choice (it must contain 5 entries). (3)

(b) Compute Residual and Residual squared for the following data: (4)

Y	68	66	68	65	69	66	68	65	71	67	68	70
Fitted value (Y_{est})	66.79	65.84	67.74	66.31	68.22	65.36	69.17	67.27	68.22	67.74	68.69	69.65