BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: MTECH **SEMESTER : II**
BRANCH: CSE, AI & ML **SESSION : SP/2025**

SUBJECT: CS632 DATA ANALYSIS AND INTERPRETATION

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	If the mean of 5 numbers is 17. The difference between two consecutive numbers is 3. The third number is the same as the mean value. Find the difference between the sum of the last two numbers and the first two numbers.	[2]	1	3
Q.1(b)	31st to 40th-rank students are standing around the table. The marks obtained in mathematics in the final semester are 95 by 8 out of 10 and the remaining got 90. Find the standard deviation of their marks.	[3]	1	3
Q.1(c)	A study was performed to test whether cars get better mileage on premium gas than on regular gas. Each of 10 cars was first filled with regular or premium gas, decided by a coin toss, and the mileage for the tank was recorded. The mileage was recorded again for the same cars using other kind of gasoline. Determine whether cars get significantly better mileage with premium gas. Mileage with regular gas: 16,20,21,22,23,22,27,25,27,28 Mileage with premium gas: 19, 22,24,24,25,25,26,26,28,32 The t-distribution is given as:	[2+3+2]	1	5

One Tail	0.05	0.025
Two Tails	0.1	0.05
df		
1	6.3138	12.7065
2	2.9200	4.3026
3	2.3534	3.1824
4	2.1319	2.7764
5	2.0150	2.5706
6	1.9432	2.4469
7	1.8946	2.3646
8	1.8595	2.3060
9	1.8331	2.2621
10	1.8124	2.2282

			CO	BL
Q.2(a)	A video game developer is testing a new game on three different groups. Each group represents a different target market for the game. The developer collects scores from a random sample from each group. The results are shown in Table. Find at the 10% significance level, are the scores among the different groups different?	[7]	1	4

Group A	Group B	Group C
101	151	101
108	149	109
98	160	198
107	112	186
111	126	160

F Distribution Table at α = 0.1 or 10%

n \ m	1	2
1	40.0337	49.5
2	8.5357	9
3	5.5344	5.4624
4	4.5346	4.3246
5	4.0458	3.7797

PTO

Q.2(b)	Discuss the formula for Ridge Regression by explaining each term involved in the formula.	[3]	2	2
Q.3(a)	In a machine learning model, if bias is significantly higher than variance, which of the following statements is correct and Why? (A) The model is overfitting the data. (B) The model is underfitting the data. (C) The model has a perfect fit.	[2]	2	5
Q.3(b)	How classes are being discriminated in LDA write its optimization function. And explain all components of optimization function. Write your answer in 2-3 lines only.	[3]	3	2
Q.3(c)	A dataset contains the following information about whether students play cricket based on the weather: Weather Play Cricket Sunny No Overcast Yes Rainy Yes Sunny No Rainy Yes Overcast Yes (a)Calculate the entropy of the root node. (b)Calculate the information gain if the data is split based on "Weather." (c)Based on the information gain, decide whether "Weather" is a good attribute for splitting.	[5]	3	3
Q.4(a)	Compare and Contrast Bagging and Boosting in terms of data sampling, approach, model combination, handling overfitting with the help of a diagram. Write 4 points only or show above parameters in your figure.	[4]	4	2
Q.4(b)	Give 2 lines comments on benefit of multi layer architecture over single layer architectures. How different activation functions helps in transforming data in to useful forms. Discuss any two important activation with their equations.	[4]	4	2
Q.5(a)	Identify give problems which algorithm is suitable for them and why? (i) A supermarket analyzes transaction data to find that customers who buy bread also often buy butter. (ii) A robot learns to navigate a maze by receiving a reward for reaching the exit and penalties for hitting walls. (iii)A company groups its customers into different segments based on purchasing behavior without having any predefined labels.	[1+2+2]	4,5	5
Q.5(b)	Apply K(=2)-Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to two iterations and show the clusters. Initially choose first two objects as initial centroids.	[5]	2	3

::::::25/04/2025::::::E