

# INDIAN STATISTICAL INSTITUTE

## MS QMS

### TEST ON MULTIVARIATE DATA ANALYSIS

Date: 26 April 2023

Time: 3 hours

Maximum Marks: 50

*Answer as many questions as you can. The maximum you can score is 50 marks*

1. A popular ice cream parlor is looking to create a new dessert by combining their ice cream with a sauce/syrup and topping it off with delicious decorations. To determine customer preferences, the parlor has conducted a conjoint analysis and collected the data, which is presented below

Case No	Ice Cream	Sauce/Syrup	Toppings	Aggregate Score
1	Vanilla	Honey	Dry Fruits	6.50
2	Vanilla	Caramel	Cherries	6.00
3	Vanilla	Chocolate	Roasted Cashew nuts	5.80
4	Vanilla	Pineapple	Lychees	5.50
5	Chocolate	Honey	Cherries	6.00
6	Chocolate	Caramel	Dry Fruits	4.50
7	Chocolate	Chocolate	Lychees	2.50
8	Chocolate	Pineapple	Roasted Cashew nuts	6.80
9	Strawberry	Honey	Roasted Cashew nuts	9.30
10	Strawberry	Caramel	Lychees	6.00
11	Strawberry	Chocolate	Dry Fruits	7.00
12	Strawberry	Pineapple	Cherries	9.50
13	Mango	Honey	Lychees	6.00
14	Mango	Caramel	Roasted Cashew nuts	7.30
15	Mango	Chocolate	Cherries	6.50
16	Mango	Pineapple	Dry Fruits	8.00

- Please analyze the data and calculate the part worth utilities and importance scores.
- Based on your analysis, which ice cream dessert is the most preferred and which is the second most preferred?
- Kindly estimate the expected scores for the best and second-best ice cream dessert?

[15]

2.

- Explain the differences between ridge and lasso regression? When to use ridge regression and lasso regression, provide one example?

- b. Explain the concept of regularization and how it applies to ridge and lasso regression?  
How to choose tuning parameter  $\lambda$  in ridge and lasso regression?
- c. Explain the steps of implementing ridge and lasso regression in a real world problem?
- d. The given equations (I) and (II) represent the ridge and lasso models for predicting the response  $y$  using the explanatory features  $x_1, x_2$  &  $x_3$ . Compare the performance of both models using the test data given below and provide your insight
- $$y = 36.12 + 6.60x_1 - 3.74x_2 + 1.85x_3 \quad (i)$$
- $$y = 36.12 + 9.99x_1 - 0.60x_2 + 1.77x_3 \quad (ii)$$

Case	$x_1$	$x_2$	$x_3$	$y$	Case	$x_1$	$x_2$	$x_3$	$y$
1	1.09	-0.89	-0.61	50.2	5	-0.16	-0.45	0.19	35.0
2	1.09	-0.86	0.19	48.5	6	-0.16	0.02	1.86	38.5
3	1.09	-0.89	1.86	44.5	7	-1.4	1.82	-0.87	17.0
4	-0.16	-0.07	-0.87	31.5	8	-1.4	1.44	0.8	29.5

[15]

3.

- a. Explain the differences between linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA)? When to use LDA and QDA, provide one example?
- b. Explain how LDA can be used for classification? How do LDA assign observations to different classes? How to evaluate the performance of LDA model? What metrics can be used?
- c. The train and test error matrices for LDA and QDA models trained on a specific dataset are provided. Evaluate and compare the performance of these models using various performance metrics, and provide your insight

	LDA – Train		LDA – Test		QDA – Train		QDA – Train	
	Predicted		Predicted		Predicted		Predicted	
Actual	0	1	0	1	0	1	0	1
0	175	12	80	4	168	19	77	7
1	13	486	3	207	8	491	2	208

[15]

4.

- a. What are similarities and differences between Multi-Dimensional Scaling (MDS) and Principal Component Analysis (PCA)? How to choose appropriate number of dimensions for MDS?
- b. How to interpret the results of MDS and Correspondence Analysis (CA)? What do the axes and distances/angles represent?
- c. How to evaluate the goodness of fit of MDS and CA models?
- d. Describe the steps involved in conducting MDS and CA analyses?

[12]