

Marks: 20

CSPC41 Machine Learning Test I (4<sup>th</sup> Sep 2023)

Time: 50 min sharp

1. What information is conveyed by various boundaries of a Box Plot? What are the formulas to decide its boundaries, lines are decided and how are outliers identified? 5

2. A random sample of 400 families of KKR is found to have avg income of Rs 3025. Assume that population std dev is Rs 280. An agency claims that the avg monthly income of families of KKR is greater than Rs 3000. Conduct an appropriate hypothesis test at 95% confidence interval to check the validity of the agency. Design both hypotheses, mention which test is applied and then compute the validity. 5

3. Consider the following data for PCA based reduction. Compute its Eigen values and Eigen vectors. 5

Feature	Point 1	Point 2	Point 3	Point 4
X1	5	10	2	4
X2	8	12	6	7

4. Consider  $X = \{2, 4, 6, 10, 12\}$  and  $Y = \{5, 7, 15, 18, 25\}$ . Step by step, estimate its coefficients ( $b_0$  &  $b_1$ ) for simple linear regression using Method 2 (using  $SS_{xy}$  and  $SS_{xx}$ ). 5