Total Marks: 15   Quiz II   Duration: 45 min

Instructions

1. Write your name and roll number.
2. Write answers legibly.
3. Mark the final answer.
4. **You MUST show all the relevant steps of derivations/calculations. Marks will be deducted if you skip important/essential steps.**

This question paper has five questions with equal marks

Q1. We have a multiple linear regression problem involving six predictor variables. We performed auxiliary regressions for each of the predictors to identify multicollinearity. The R^2 value for the auxiliary regression for the predictor x_2 is equal to **X**/100. Should we retain this predictor (x_2) in our model? Show your calculations, and explain your answer in one or two sentences. **X** is the last two digits of your roll number. For example, if your roll number is 190106050, then **X** = 50.

Q2. What is the mathematical relation between R^2 and the adjusted R^2?

Q3. Distances between five data points (D1 to D5) are shown in the following distance matrix. In this matrix, an entry d_{ij} is the distance between data point i and data point j. We are performing hierarchical clustering of this data. Calculate the distance between the 6th and 7th clusters. Use the principle of Complete Linkage for calculating the distance between clusters. Show all steps of your calculation, and briefly (one or two sentences) justify your answer.

	D1	D2	D3	D4	D5
D1	0	10	12	6	18
D2	10	0	10	24	2
D3	12	10	0	12	10
D4	6	24	12	0	20
D5	18	2	10	20	0

Q4. We measured the expression of several genes in three experimental conditions. The data for two genes are shown here. Calculate the Euclidean distance between these two genes. You have to use your roll number for this question. Write your roll number in your answer.

Your Roll Number: XYZABCPQR

190106 019
XYZ   PQR

XYZ are the first three digits
PQR are the last three digits

Gene 1	Gene 2
X	P
Y	Q
Z	R

Q5. y is the dependent variable for the following data, and x is the predictor. We fit a cubic polynomial to this data using regression. For regression, we are using the linear algebra-based method. In this method, we use the relationship $B = \left(X^T X \right)^{-1} X^T Y$. Here, B is the vector with unknown parameters/coefficients of the polynomial. X is a matrix, and Y is a vector. What is the sum of all elements in the first row of the X matrix? M is the last digit of your roll number. For example, if your roll number is 190106050, then M = 0. Show all steps of your calculations.

x	M	3	4	5	6	7	8
y	25	61	120	200	340	500	701