LAHORE UNIVERSITY OF MANAGEMENT SCIENCES
Department of Electrical Engineering

EE212 Mathematical Foundations of Machine Learning and Data Science
Quiz 06

Total Marks: 10 **Time Duration:** 45 minutes

Question 1 (2 marks)

You have developed a regression model for predicting a scalar outcome y from a feature vector x of dimension 20, using a collection of $N = 600$ data points. The mean of the outcome variable y across the given data is 1.85, and its standard deviation is 0.32. For training-test split, you use 5-fold cross-validation that splits the data into 5 parts or folds. If you obtain the following RMS test errors (based on forming a model based on the data excluding fold i, and testing it on fold i).
How would you expect your model to do on new, unseen (but similar) data? Provide a brief explanation.

Fold Excluded	RMS test error
1	0.13
2	0.11
3	0.09
4	0.13
5	0.12

Question 2 (8 marks)

Consider the data given below:

x	y
0	2
1	3
2	5
3	4
4	6

The data can be modeled to follow a linear model of $y_i = ax_i + b$, where a and b are unknown parameters.

(a) [2 marks] Model the above equation as $y = A\Theta$ and find A. Θ is a 2 x 1 matrix containing a and b, and A is generated using x.

(b) [4 marks] Evaluate a and b.
Information: The inverse of a 2 x 2 *matrix* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *is:* $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

(c) [2 marks] Mark the data points on a graph and sketch the line $y = ax + b$ on the same graph.