

LAHORE UNIVERSITY OF MANAGEMENT SCIENCES
Department of Electrical Engineering
EE514/CS535 Machine Learning
Quiz 03 Solutions

Name: _____

Campus ID: _____

Total Marks: 10

Time Duration: 12 minutes

Question 1 (4 marks)

True or False: (Circle your answer)

- (a) The accuracy of classification always increases with an increase in the number of inputs/features.
T/F

Solution: F: Not necessarily.

- (b) The “Intrinsic” dimensionality of the data may be smaller than the actual size of the data.
T/F

Solution: T: Self-explanatory.

- (c) In feature selection, we select the features in the subset that either improves classification accuracy or maintain the same accuracy. T/F

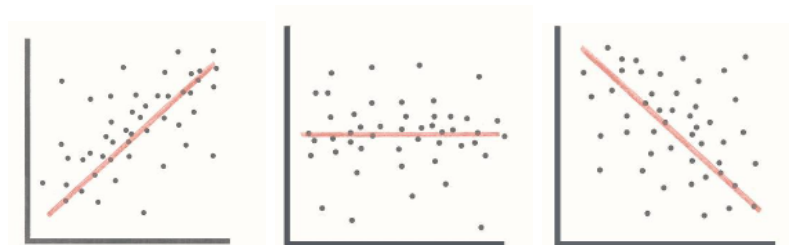
Solution: T: Self-explanatory.

- (d) Dimensionality reduction using principal component analysis maximizes discriminatory information (i.e., maximizes the separation between classes to enable better classification). T/F

Solution: F: It does not account for discriminatory Information.

Question 2 (1 mark)

You are provided with the following correlation graphs between features $[x^{(1)}, x^{(2)}, x^{(3)}]$ and true labels as [left, middle, right] plots respectively. Choose the top 2 features and explain your choice.



Solution: x_1 and x_3 (strong correlation)

Question 3 (3 marks)

You are provided with the following feature value table. Calculate the covariance between x and y .

x	y
-1	-3
3	5
1	1

Solution: Mean of $x = 1$

Mean of $y = 1$

Covariance : $1/n \sum_{i=1}^n (x_i - 1) * (y_i - 1) = \frac{16}{3}$

Question 4 (2 marks)

Define the difference between feature selection and feature extraction (either in words or mathematically).

Solution: Dimensionality reduction refers to the reduction in the number of features such that the learning ability of the classifier is enhanced. Feature selection refers to selecting a subset of the existing features, where we transform existing features to obtain a set of new features using some mapping function in feature extraction. Mathematically, given d features $\mathbf{x} = [x_1, x_2, \dots, x_d]$, we want to select k out of d features in as a) feature selection: $\mathbf{z} = [x_{i_1}, x_{i_2}, \dots, x_{i_k}]$, and (b) feature extraction: $\mathbf{z} = [z_1, z_2, \dots, z_k] = f(\mathbf{x})$.