MACHINE LEARNING ASSIGNMENT SOLUTIONS

1) A- Least Square Error

2) A - Linear regression is sensitive to outliers.

3) B – Negative.

4) B – Correlation.

5) C - Low bias and high variance.

6) B - Predictive model.

7) D – Regularization.

8) D - SMOTE.

9) A - TPR and FPR.

10) B – False.

11) A - Construction bag of words from a email.

12) A - We don't have to choose the learning rate.
B - It becomes slow when number of features is very large.
C - We need to iterate.

13) Regularization is used to know whether the model is overfitted or not. It is used to reduce the variance of the model without a substantial increase in the bias. The tuning parameter lamda is used if more the lamda used the model will become bias. So selecting lamda using cross validation method. CV error should be minimum.

14) Lasso (L1) and Ridge (L2) are the algorithms used in Regularization. In lasso feature column having very less effect on prediction of label is eliminated from feature and in ridge a feature column with less effect on prediction of label its importance is reduced while predicting label.

15) The term error used in Linear Regression is the distance difference between the actual data point and predicted line. It should be minimum.