Worksheet Set 1 Machine Learning Assignment

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 modal

7. D) Regularization

8. A) Cross validation

9. A) TPR and FPR

10. B) False

11. B) Apply PCA to project high dimensional data

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

13. When we use regression models to train some data, there is a good chance that the model will overfit (high variance & low bias) the given training dataset. Regularization helps sort this overfitting problem by restricting the degree of freedom of a given equation i.e. simply reducing the no. of degrees of a polynomial function by reducing their corresponding weights. In Linear regression we do not want coefficients as a small change in weight can make a large difference for the dependent variable (the target). So, regularization constraints the weights of such features to avoid overfitting. To regularize the model, we apply some penalty like learning rate to our model that techniques are:

    A) LASSO

    B) RIDGE

    C) ELASTICNET (LESS POPULAR)

14. Particular Algorithms used for Regularization are:

A) **Lasso (Least Absolute Shrinkage & Selection Operator):** In this it finds relation between features & labels, nullify that features that doesn't have relation to labels. It also known as L1 form. $\lambda = \Sigma |\beta j|$

B) Ridge: It gives very less important to features that doesn't have relation with the label. And also known as L2 Form. $\lambda = \Sigma |\beta j|^2$.

C) **Elasticnet:** In this it linearly combines the L1 and L2 penalties of the lasso and ridge methods.

15. When we train our model, it produces the best fit line. This line gives us the predicted value for the testing dataset. So, the error term or the residual is the sum of all the mode of difference between the Actual target value & predicted value by the model. For Linear Regression, we can use Root Mean Square Error (RMSE) or Mean Square Error (MSE).