**Assignment 5**

1. What are the reasons for feature scaling?

Ans:

Gradient Descent based and Distance/Similarity metric algorithms suffers a lot when features are of different scales. As these algorithms involve in calculating distances and minimizing it.

There will be convergence issue in those algos if features are of different scales.

Hence Feature scaling is performed.

1. What is the difference between Feature Selection under Feature Engineering? Can you perform feature selection using regularization, if yes then how?

Ans:

Feature selection involves selecting certain important features using metrics like correlation matrix (HeatMap), Random forest Feature selection, Tree based classifiers.

Whereas Feature engineering is a broader term which involves features selection, Feature Extraction (Dimensionality Reduction) and creating new features

Yes, Feature selection can be performed by using LASSO regularization. It uses L1 regularization, it adds a penalty equal to the absolute value of the magnitude of coefficients.

It will set coefficients of few features to zero, there by eliminating them. In this way Feature selection can be performed using LASSO regularization.

1. Suppose you are working on a Machine Learning problem, your training accuracy is lower than the testing accuracy, what can be the reason for this?

Ans:

This maybe due to underfitting. Can be improved by

-Reducing any constraints on the model (Regularization)

-Adding more training data

-Adding more features

-Increasing model complexity

1. You are training a machine learning model, your training and the testing accuracy are decreasing, what can be the reason for this?

Ans:

This maybe due to underfitting. Can be improved by

-Reducing any constraints on the model (Regularization)

-Adding more training data

-Adding more features

-Increasing model complexity

1. What solutions you can provide for optimal bias-variance levels in a machine learning problem?

Ans:

When the model fails to learn properly it is high bias case (underfit), and when it learns too much and noise in the data it is High variance case (overfit). Both the cases are not desirable.

There should be a trade-off between these two. This can be observed when both train loss and test loss are relatively low. This case is also known as “Low Bias and Low Variance” case.