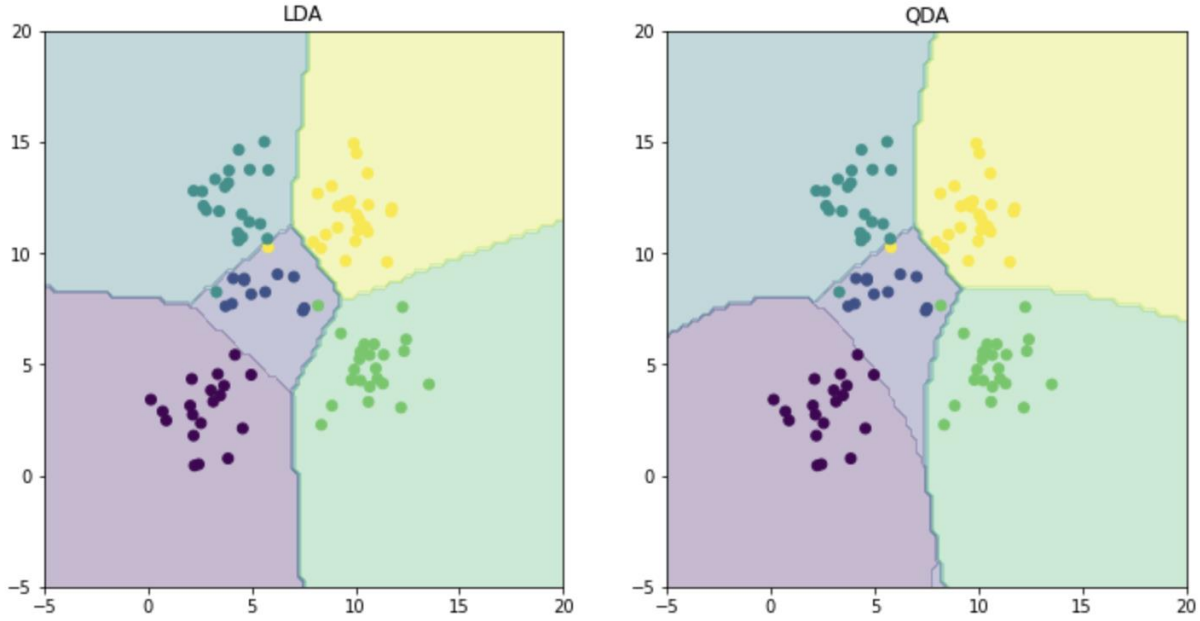
Machine Learning Assignment 1

First Name:	**Viswa Nihar**
Last Name:	**Nukala**
UBID:	**50414392**
First Name:	**Nithin Sastry**
Last Name:	**Tellapuri**
UBID:	**50365532**
First Name:	**Amartya**
Last Name:	**Banerjee**
UBID:	**50425019**

Problem 1

Accuracy of LDA: 97.0

Accuracy of QDA: 96.0



There is a difference between both the graphs as for QDA, we learn the data by dividing the data into different classes and finding the covariance in each class, while in LDA, we use only one covariance for the whole data and hence the classification is a little linear. Hence in QDA, the plot is curved.

Problem 2

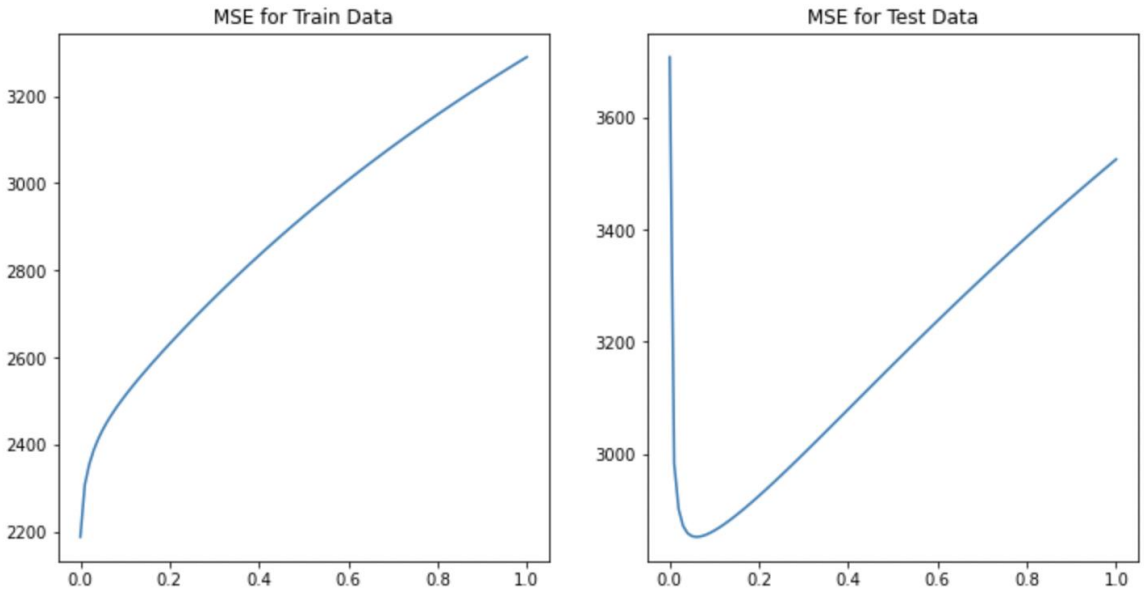
MSE without intercept [106775.36152001]

MSE with intercept [3707.84018118]

MSE with intercept is better as the data is predicted more accurately. The data is more normalized rather than making the data restricted to be only linearly dependent without the intercept.

Problem 3

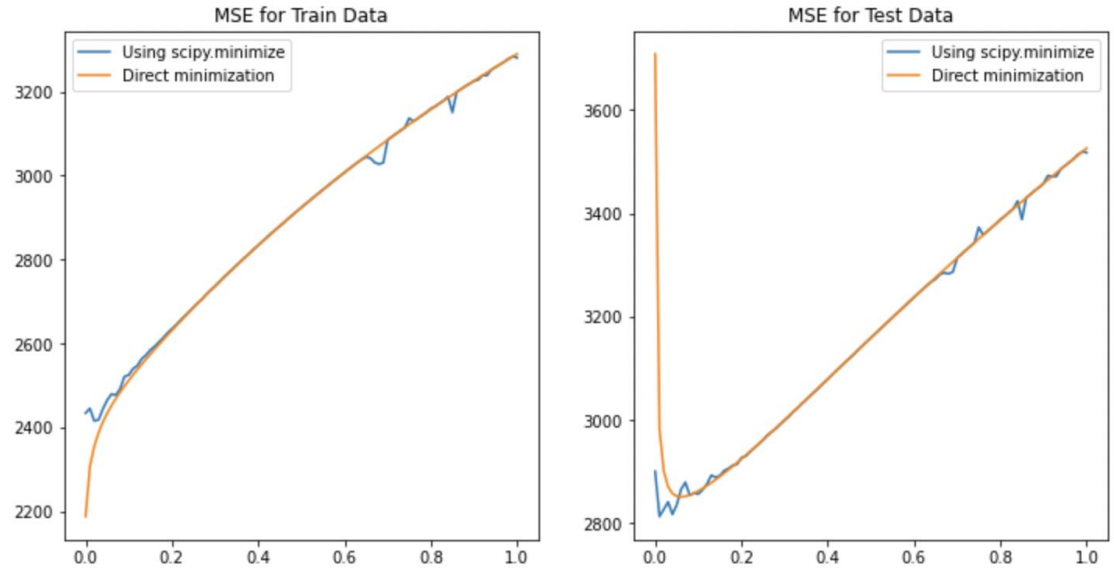
Following is the graph for the MSE for test and train data with different lambdas:



We can see that the MSE is least for **lambda = 0.06** because at this value the MSE is least. The weight values in both the cases are hugely different by large magnitudes and Ridge Regression clearly show which parameters are more dependent on the output.

Problem 4

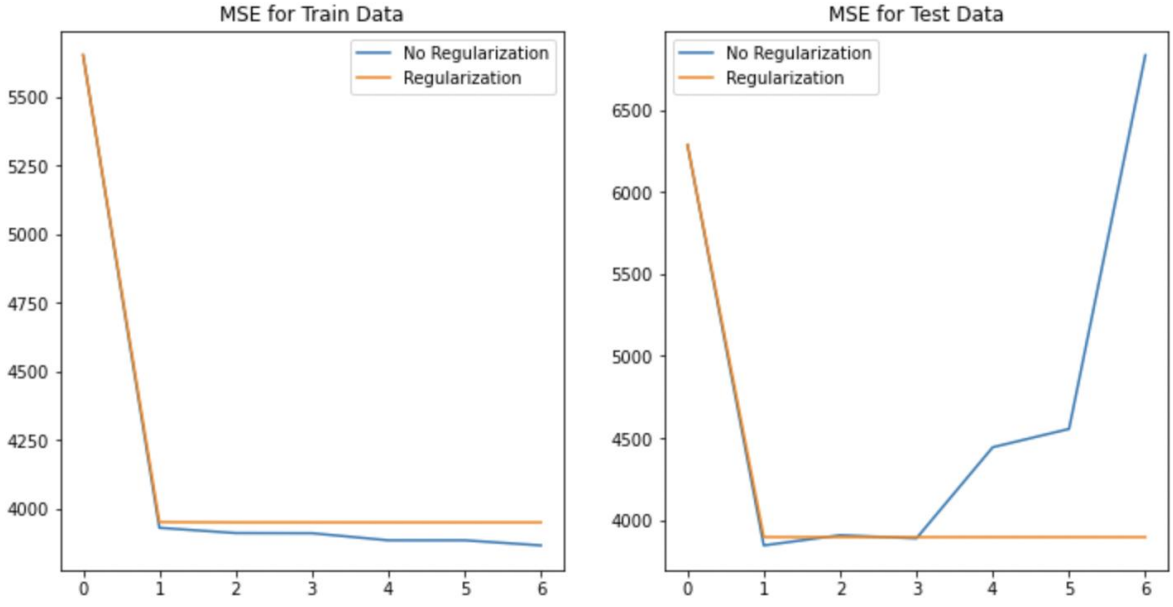
Following is the graph for the MSE for test and train data with different lambdas:



We can see that the MSE is least for **lambda = 0.01** because at this value the MSE is least. The weight values in both the cases are quite different by large magnitudes and Ridge Regression clearly show which parameters are more dependent on the output.

Problem 5

The following are graphs for **lambda = 0.06:**



The difference when there is no regularization (lambda = 0) and with regularization (lambda = 0.06) is that the maximum square error increases even after exponentializing the data.

The optimal value of p is 4 as MSE reaches minimum after this value and stays constant even if we increase the value of p.

Problem 6

The training error is the same in the OLE Linear Regression, Ridge Regression, Gradient Descent, but it is a little higher in the Non-linear Regression's optimal value. This means that the data was not trained that good in the last approach.

When we come to testing data, the error was least in Ridge Regression with Gradient Descent. In other cases, it is a little higher. Even though we will need to multiple computations in gradient descent, if we choose the right first w, then we will have the optimal lambda in least number of iterations and the best one. In Non-linear, we will need too many computations and an optimal lambda value. In a few cases, non-linear regression will be better.

Hence Ridge Regression using gradient descent is a better approach due to the l2 normalization and optimal weights.