TEAM MEMBERS:

PARTH JITHENDRA OZA
SAI KIRAN PUTTA

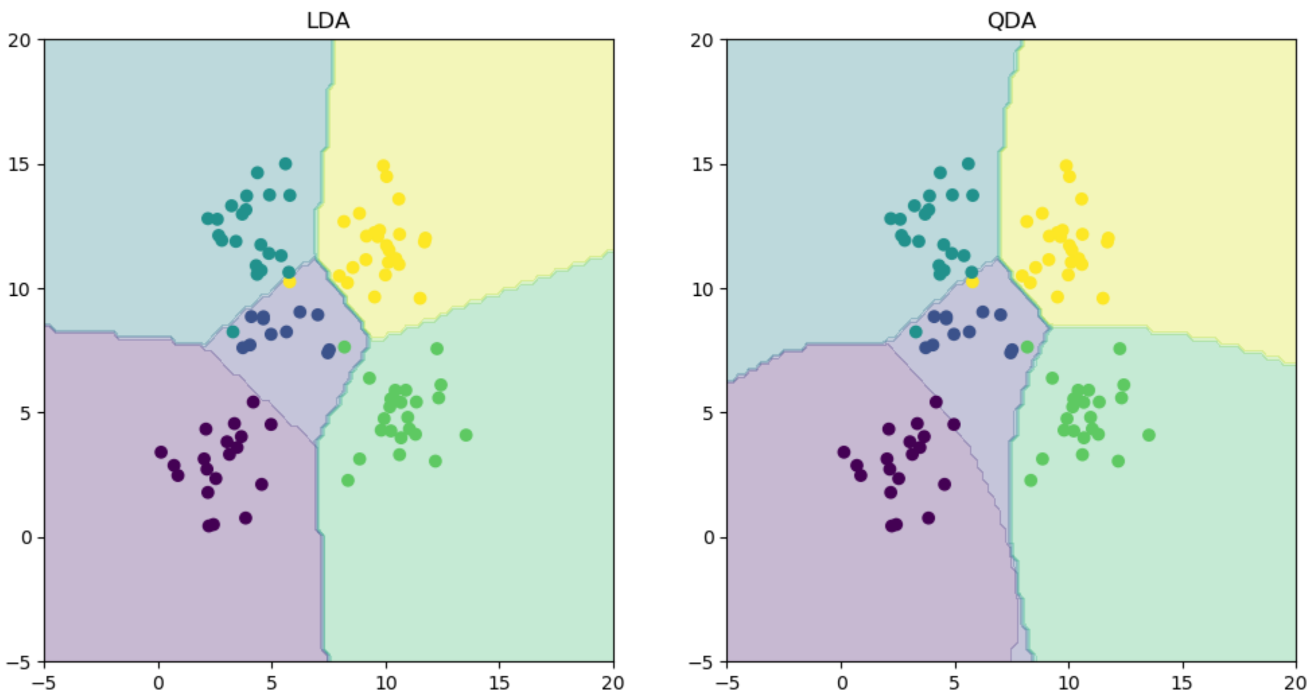
PROBLEM – 1

Following is the accuracy for LDA and QDA.

```
LDA Accuracy = 0.97
QDA Accuracy = 0.96
```

LDA has an accuracy of 97% and QDA has an accuracy of 96%

Following is the result we get by plotting the decision boundary for both LDA and QDA.



Inference:
As expected, the decision boundary for LDA is linear and the decision boundary for QDA is non-linear. There is a difference in the decision boundary because of the mathematical formulation of QDA and LDA. Where in we are aspiring for a quadratic decision boundary for QDA and a linear decision boundary for LDA.

Probably, the two points that sit on the decision boundary (Purple and Violet) is the reason for difference in the accuracy of LDA and QDA. Though QDA fits the data better in general, for this data since the two points sit on the decision boundary of QDA, the accuracy is off by 1% for QDA for this dataset.

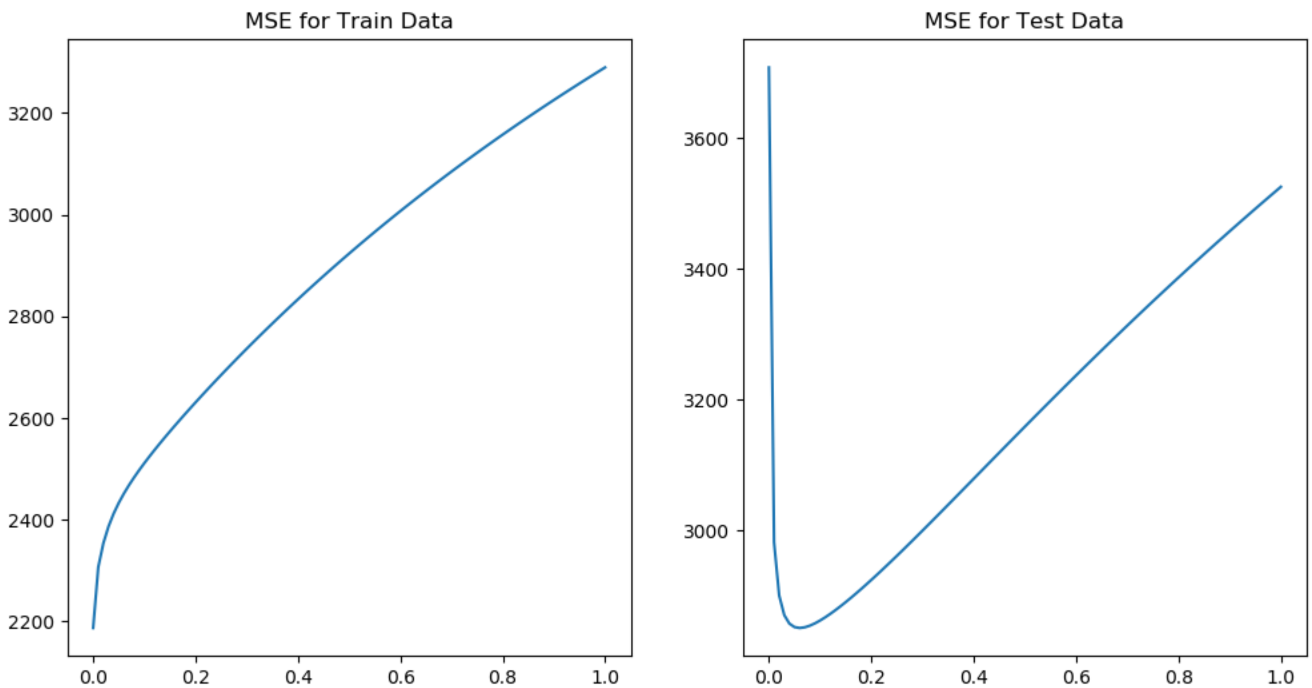
PROBLEM – 2

Following are the Mean Square Errors that are obtained,

```
Mean Square Error for Train without Bias: 19099.4468446
Mean Square Error for Train with Bias: 2187.16029493
Mean Square Error for Test without Bias: 106775.361553
Mean Square Error for Test with Bias: 3707.8401812
```

Inference:
Like expected the MSE for training is less than Test for both with and without bias. And the MSE for test with bias is less than without bias since the line fits much better with bias term.

PROBLEM – 3
Following is the MSE comparison for Train and Test dataset.



Inference: The optimal lambda for this dataset is 0.06
The left plot is the MSE for the train set. Increase in the lambda value indicates more penalization of learning. Hence as the lambda increases the fit is not proper(underfit) resulting in high MSE.

The right plot is the MSE for test set. The optimal value of lambda for this dataset is 0.06 for which the test set MSE is lowest. And since with the increase in lambda resulted in poor fit to training data, the corresponding test error is high as expected for the remaining part of the graph.

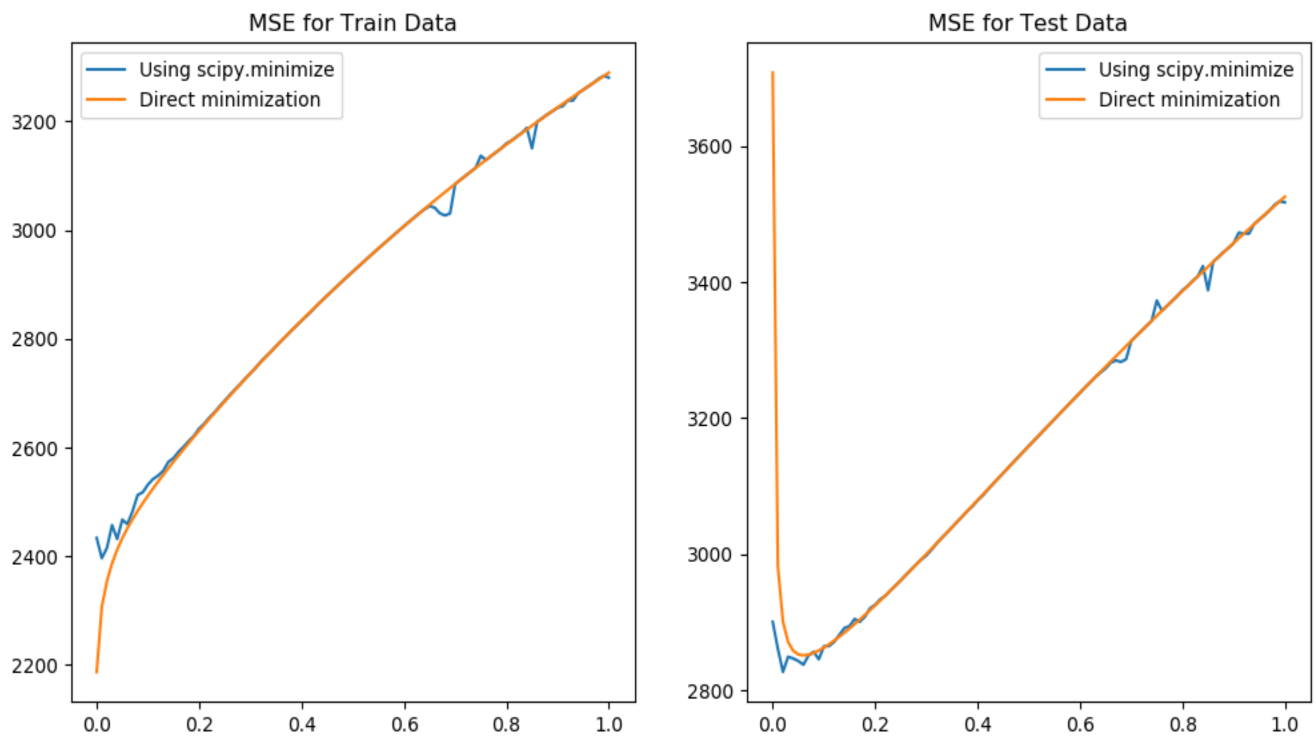
Comparison of OLE weights and Ridge weights (for lambda = 0.06)

```
OLE weights         Ridge Weights
(148.1548759957809, 150.4595980697856)
(1.274852203365299, 4.807768987793224)
(-293.38352240162203, -202.9061146819866)
(414.72544841350464, 421.71945760092547)
(272.0891344660213, 279.4510728848136)
(-86639.4571325779, -52.29708232685175)
(75914.46802997589, -128.5941890721565)
(32341.622823297977, -167.50057028162132)
(221.1012148818154, 145.74068095550138)
(29299.551196813583, 496.306041227377)
(125.23036027501803, 129.94845774987414)
(94.41108336453908, 88.3043807646896)
(-93.86286326643312, 11.290676886252186)
(-33.72827995982698, 1.8853253064293938)
(3353.1977123767138, -2.583641568928556)
(-621.0963073372841, -66.89445480936837)
(791.7365328744054, -20.619399553241237)
(1767.7603888482554, 113.39301453676626)
(4191.674056112766, 17.9908682671458)
(119.43812093093584, 52.502359626481166)
(76.61034003720124, 109.6876551252123)
(-15.200129292559723, -10.727796288278483)
(82.2424593586129, 71.67974828666965)
(-1456.662083948846, -69.30906365623127)
(827.3867025491782, -124.0343729317674)
(869.2909521837719, 102.63981795252107)
(586.23449510106, 72.64220588231626)
(427.0267265938455, 79.24754013481567)
(90.24676901056955, 38.483192149999425)
(-17.887622413851204, 32.980094455627906)
(141.69677383009184, 92.09539121793594)
(582.8193837780273, 68.97936153759315)
(-234.0375101728423, -24.417009136952174)
(-256.0714520220645, 101.85387966574453)
(-385.1774005157786, 1.3912266905651904)
(-33.41767353095929, 20.85757155232517)
(-10.735006613573205, -29.65490134454234)
(257.1071888642182, 130.41115985990402)
(59.95545961431344, -16.751087956501692)
(383.72804197794176, 87.51340343600694)
(-404.1583898694662, -45.64238361747604)
(-514.2864343129404, -30.9228849939975)
(38.363664007076295, -10.071397806905608)
(-44.61028889998852, 31.133348961433853)
```

Inference: For most of the features, as expected the weights of Ridge are low as compared to weights of OLE since they are penalized.

PROBLEM – 4

Following is the comparative graph that is obtained which compares MSE for Train data and Test data



Inference: By visual inspection the MSE obtained by Gradient Descent is less.

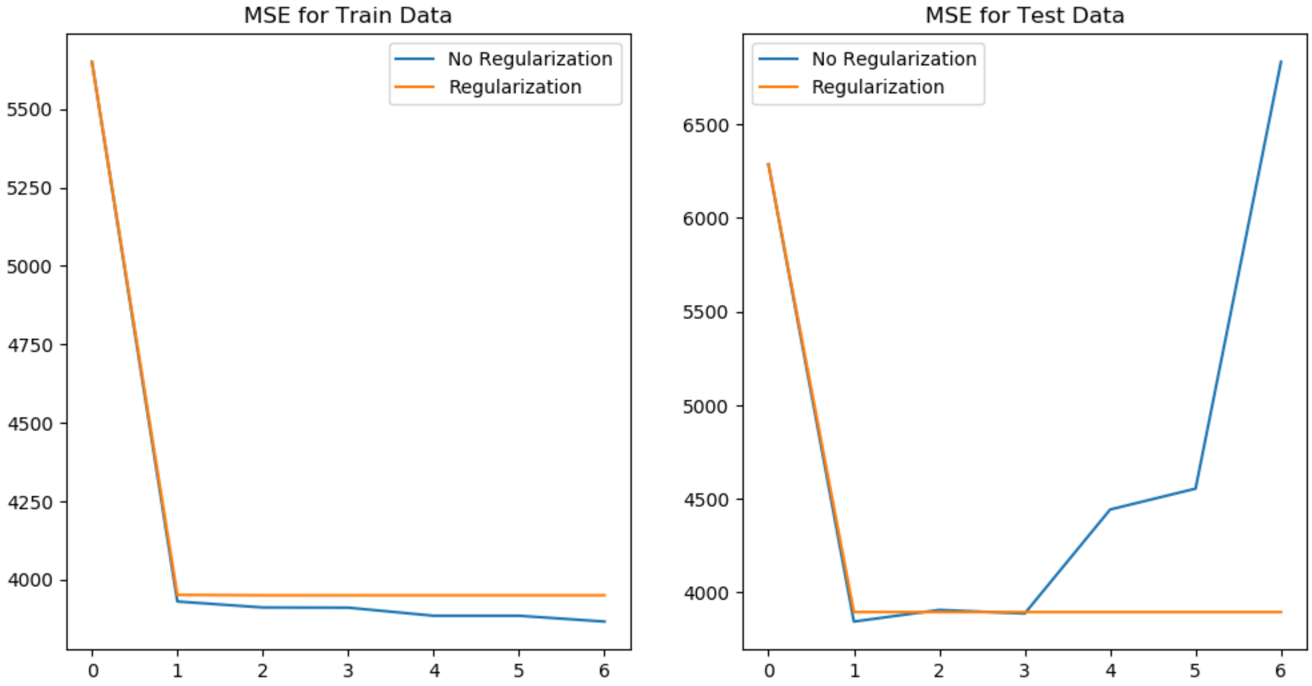
```
Minimum MSE with out Gradient Descent: [ 2851.33021344]
Minimum MSE with Gradient Descent: [ 2826.9525654]

Minimum Lambda obtained with out Gradient Descent: 0.06
Minimum Lambda obtained with Gradient Descent: 0.02
```

Inference: Like we observed above, the MSE by using Gradient Descent is less (2826.9525)
And the corresponding lambda value is 0.02.

PROBLEM – 5

Following is the graph of MSE of both train and test data with and without regularization.



Inference: As expected with out using any regularization, the MSE on training decreases but it is overfitting. We know this since the corresponding MSE values for test data is increasing.
While MSE for both train and test when regularization is used, almost remains constant.

PROBLEM – 6

Overall, we are trying to get minimum MSE on the test set without overfitting.
From all the results above we can see that this is possible with using Gradient Descent for optimization with regularization parameter Lambda as 0.02