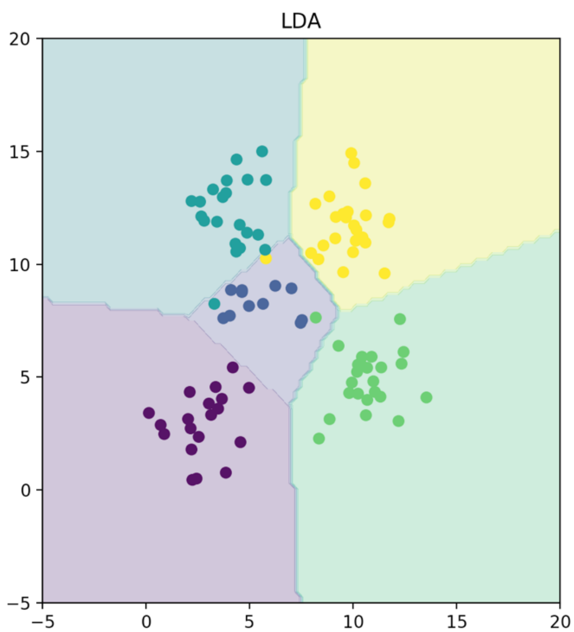
Joseph Cascioli
Jobin Joseph
Benjamin Riexinger
CSE 474 assignment 2

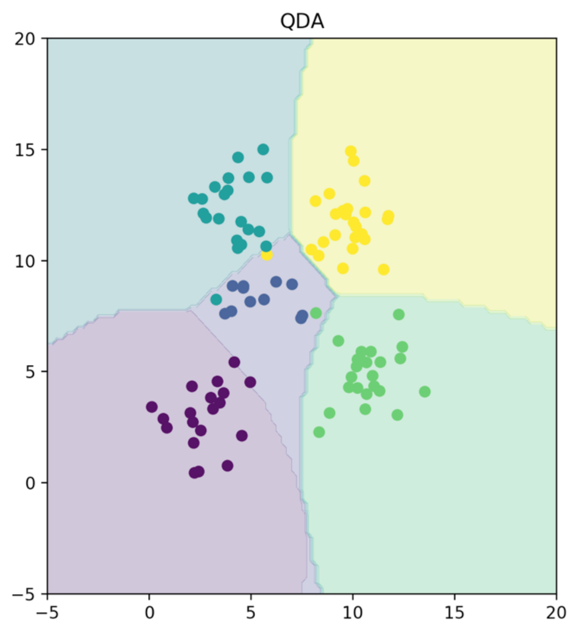
Group 11

1)

For problem 1 we were tasked with calculating Linear Discriminant Analysis and Quadratic Discriminant Analysis (LDA and QDA respectively). Using the provided data set we were able to acquire and accuracy percentage of 97 and 96 respectively. Despite using the same data, the QDA had a lower accuracy because the co-variance is calculated for each class instead of for the entire data set as in LDA. This can be seen visual in the following plots. If looking at the LDA plot you will notice that the boundary lines are more jagged and linear whereas in the QDA plot the lines are more curved and smoother.

These techniques are pictured below (LDA then QDA):





2)

For problem 2 we worked with Linear Regression and calculating the MSE. We ran the data set with the following results:

MSE without intercept for Test 106775.36156
MSE with intercept for Test 3707.84018163
MSE without intercept for Training 19099.4468446
MSE with intercept for Training 2187.16029493

As with problem one we can see that our results are better with an intercept. This is represented with a lower MSE.

3)

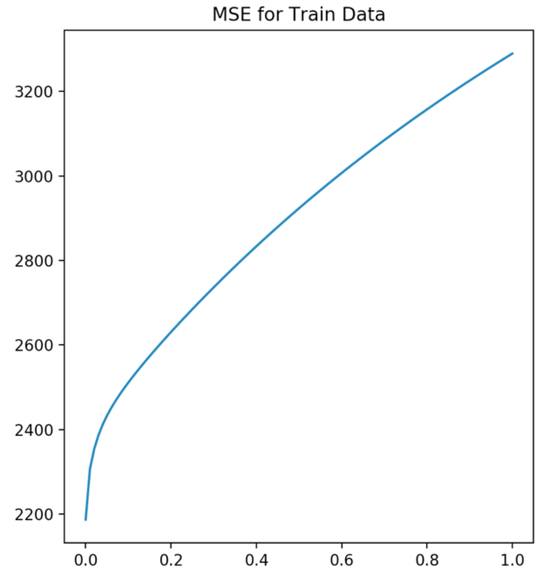
For problem 3 we experimented with Ridge Regression. We ran the regression using MSE with the following results:

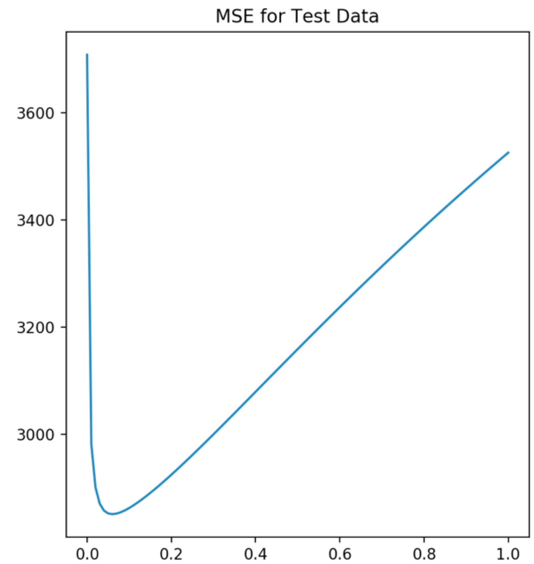
OLE Regression with intercept
MSE with intercept for Training 2187.16029493
MSE with intercept for Test 3707.84018163

Ridge Regression with intercept
MSE for Train Data 2451.52849064
MSE for Test Data 2851.33021344

The Ridge regression is better than the OLE regression with the intercept because the test data MSE is lower in the Ridge regression with intercept.

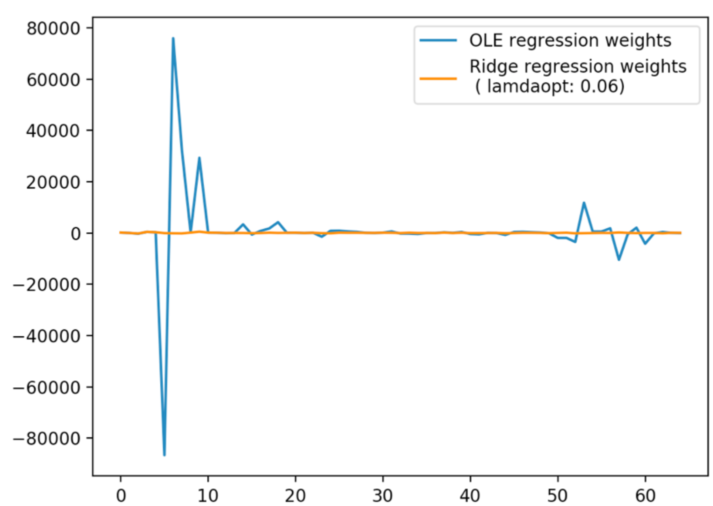
With the following two plots we can see the error of this method.





We can see with this plot comparison of OLE Regression vs Ridge Regression. We also see that our lambda optimal occurs when the error is lowest on the test data—it occurred at .06, leading to minimum MSE in test data. Here is a snippet of the Lambda/Error table to show the minimum:

Lambda: 0.03	Test Error: 2871.00639587	Training Error: 2386.83461891
Lambda: 0.04	Test Error: 2857.94197829	Training Error: 2412.13399602
Lambda: 0.05	Test Error: 2852.57954052	Training Error: 2433.18197531
Lambda: 0.06	Test Error: 2851.20831764	Training Error: 2451.9288123
Lambda: 0.07	Test Error: 2852.35011103	Training Error: 2468.07867326
Lambda: 0.08	Test Error: 2854.87964918	Training Error: 2483.36630198
Lambda: 0.09	Test Error: 2858.44435887	Training Error: 2497.74024352



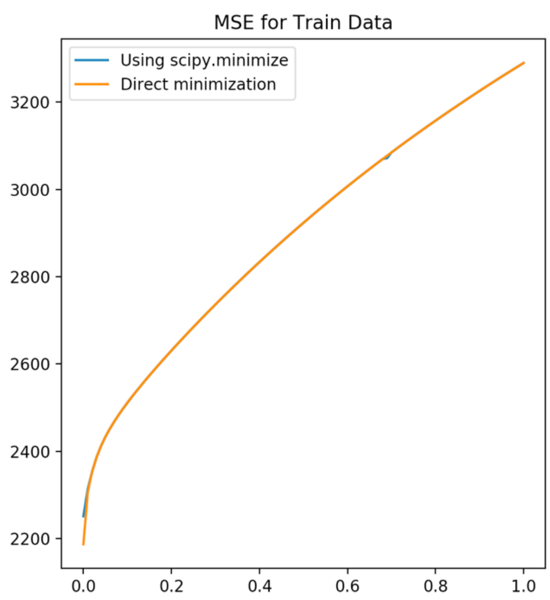
We can see that ridge regression is a better than OLE giving us the best results so far.

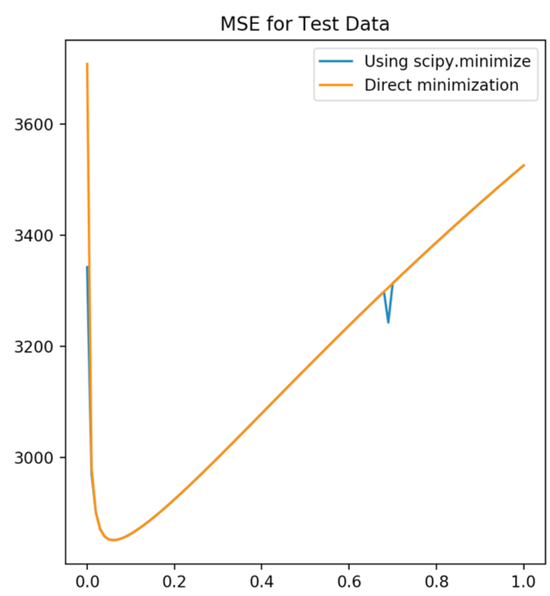
4)

For problem 4 we used gradient decent for Ridge Regression Learning. We obtained the following values:

MSE for Train Data 2240.44909363

MSE for Test Data 2842.69772813





We can derive based on how the graphs overlay almost identically that the Gradient Descent is the same as regular Ridge Regression sharing the optimal lambda value of .06. Comparisons between graphs are smoother with max iterations of 100.

5)

For problem 5 we're using Non-Linear Regression.

```
MSE for Train Data 3866.88344945
MSE for Test Data 3845.03473017
```

P vs. Lambda = 0 & Lambda = optimal for testing

p = 0
Zero Lamda Testing: 6286.404791
Optimal Lambda Testing: 6286.88196

p = 1
Zero Lamda Testing: 3845.034730
Optimal Lambda Testing: 3895.85646

p = 2
Zero Lamda Testing: 3907.128099
Optimal Lambda Testing: 3895.58405

p = 3
Zero Lamda Testing: 3887.975538
Optimal Lambda Testing: 3895.58271

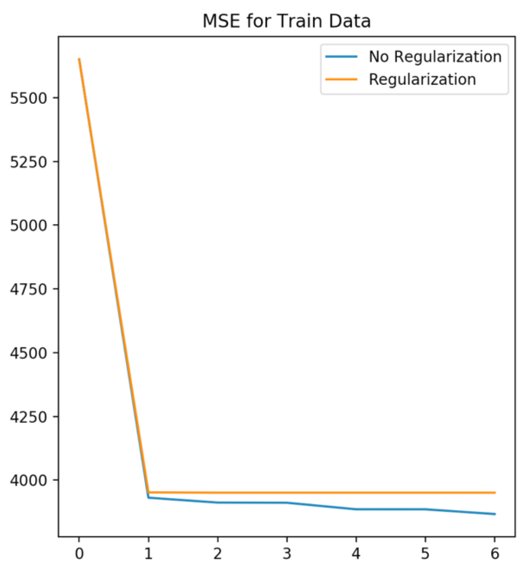
p = 4
Zero Lamda Testing: 4443.327891
Optimal Lambda Testing: 3895.58266828

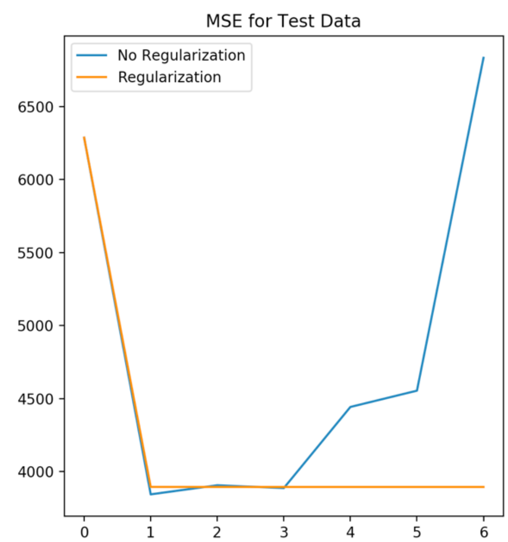
p = 5
Zero Lamda Testing: 4554.8303774
Optimal Lambda Testing: 3895.5826687

p = 6
Zero Lamda Testing: 6833.459148
Optimal Lambda Testing: 3895.58266872

The optimal value for p when lambda is 0 for Testing is 1. While the optimal value for p when lambda is optimal for testing is 4.

```
p = 0
Zero Lamda Training: 5650.7105389
Optimal Lambda Training: 5650.71190703

p = 1
Zero Lamda Training: 3930.91540732
Optimal Lambda Training: 3951.83912356

p = 2
Zero Lamda Training: 3911.8396712
Optimal Lambda Training: 3950.68731238

p = 3
Zero Lamda Training: 3911.18866493
Optimal Lambda Training: 3950.68253152

p = 4
Zero Lamda Training: 3885.47306811
Optimal Lambda Training: 3950.6823368

p = 5
Zero Lamda Training: 3885.4071574
Optimal Lambda Training: 3950.68233518

p = 6
Zero Lamda Training: 3866.88344945
Optimal Lambda Training: 3950.68233514

The optimal value for p while lambda value is zero or optimal is 6.
```





Non linear regression performs better with regularization than without regularization. As seen above MSE for training data decreases with no regularization but stays the same with regularization. On the test data however when no regularization is performed the testing error decreases for p =1,2,3. After p = 3, the testing error increases because of over fitting problem. Performing regularization can reduce the over fitting problem.

6)
For question 6:

<u>OLE Regression</u>
```
MSE without intercept for Test 106775.361458
MSE with intercept for Test 3707.84018102
MSE without intercept for Training 19099.4468446
MSE with intercept for Training 2187.16029493
```

<u>Ridge Regression</u>
```
MSE with intercept on testing data: 2851.33021344
MSE with intercept on training data: 2451.52849064
MSE without intercept on testing data: 38923.3438457
MSE without intercept on training data: 21704.7493572
```

Optimal Lambda for Ridge Regression: 0.06

<u>Non linear Regression</u>

Testing – optimal p value at lambda 0 is p = 1
Testing – optimal p value at lambda optimal is p = 4
Training – optimal p value at either lambdas is p = 6

We have observed that Ridge Regression works better for Test data while OLE regression is better for Training. The Optimal value for Lambda is 0.06 for Ridge Regression. So we can conclude that, Ridge Regression is the best for Training and Test data.