Report for
CSE 474/574 Introduction to Machine Learning
Programming Assignment 2
Classification and Regression

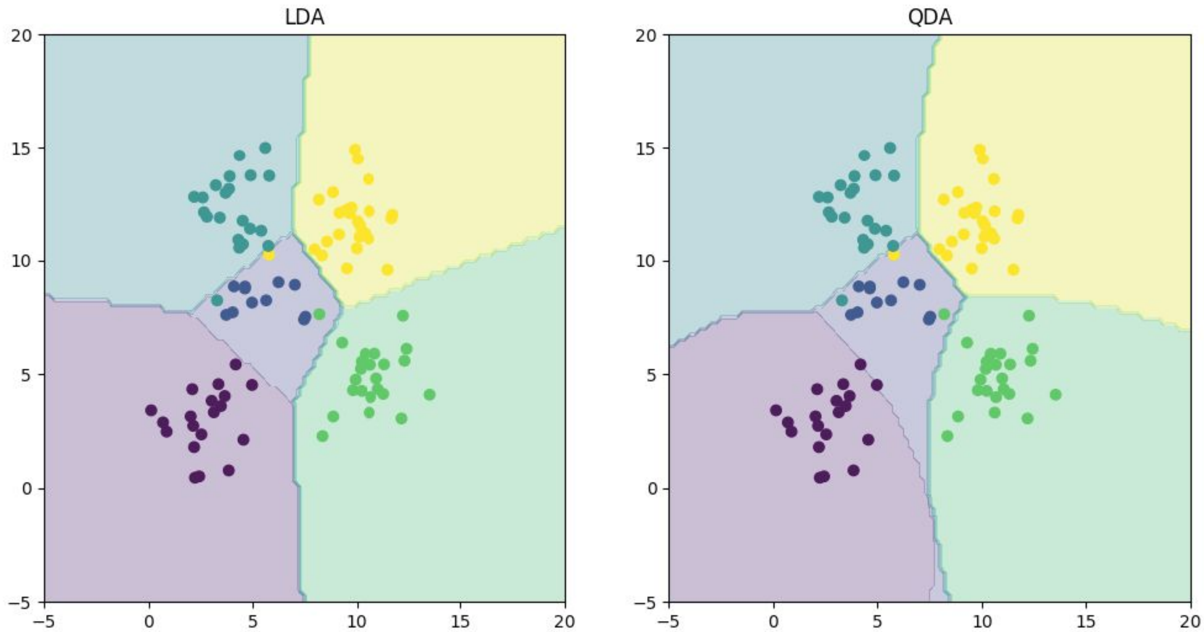
Course Number: CSE 474

Student Name:
Steven Mundt
Qinxin Tian
Fengyu Wu

1.

The LDA Accuracy = 97%
The QDA Accuracy = 96%

LDA and QDA Plots:



The boundaries for LDA QDA are different. As you can see that LDA can only learn linear boundaries while QDA can learn quadratic boundaries. This makes QDA more flexible and explains the boundary differences.

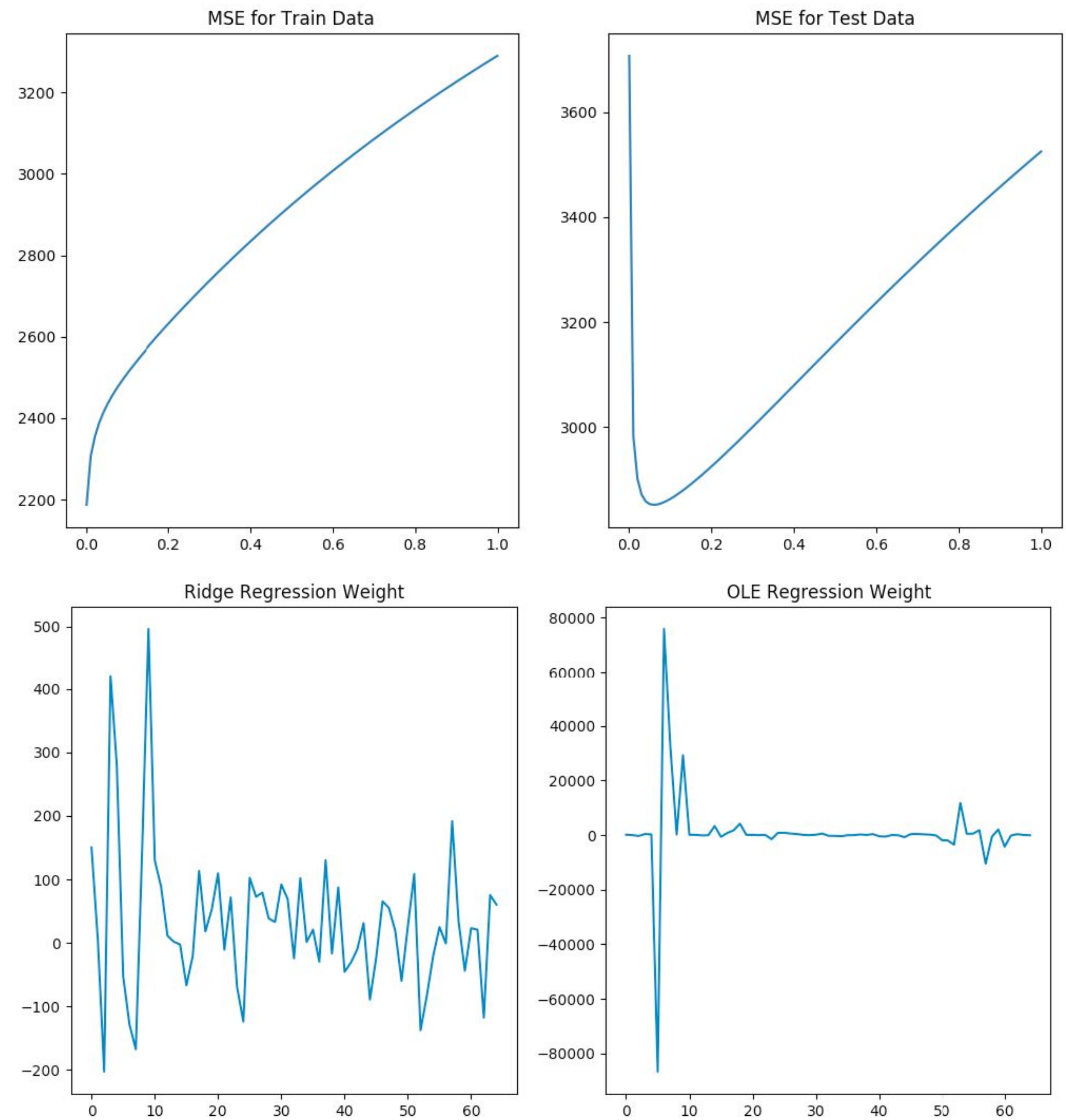
2.

MSE without intercept Test Data 106775.36155789
MSE without intercept Training Data 19099.44684457

MSE with intercept Test Data 3707.84018132
MSE with intercept Training Data 2187.16029493

As you can see the MSE with using an intercept is the better than MSE without an intercept.

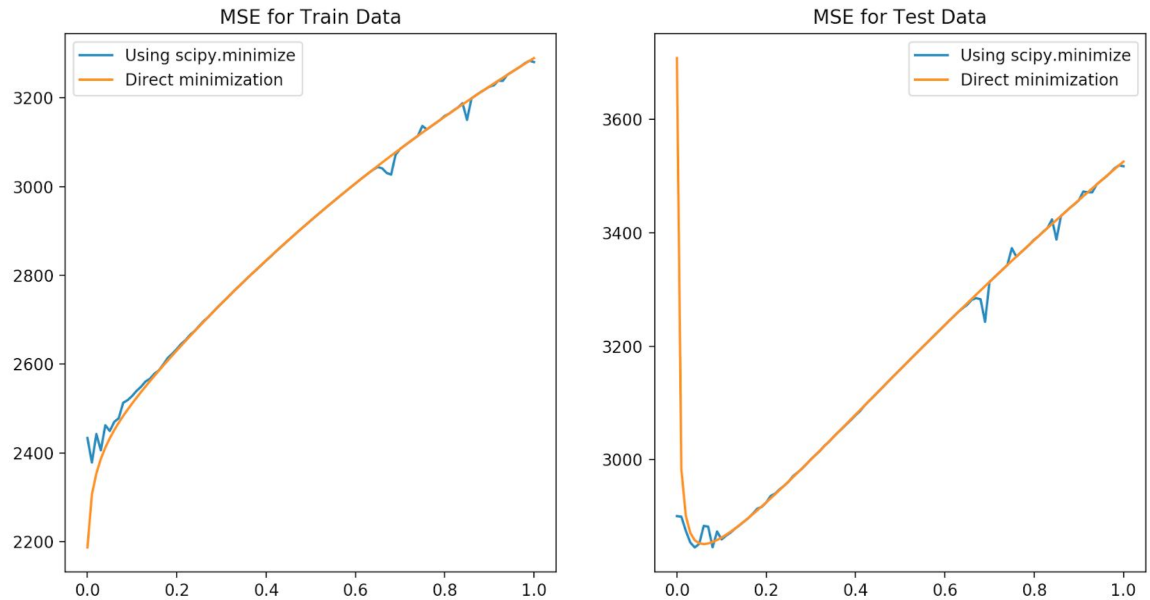
3.



The relative weights of OLE Regression seems to be less sporadic than the Ridge Regression weights. This pattern seems to take place with both the training data and the test data.

Using the MSES for the Test Data we computed λ to be 0.06. The optimal value for λ is 0.06 because it is the lowest MSE value observed for the test data set.

4.

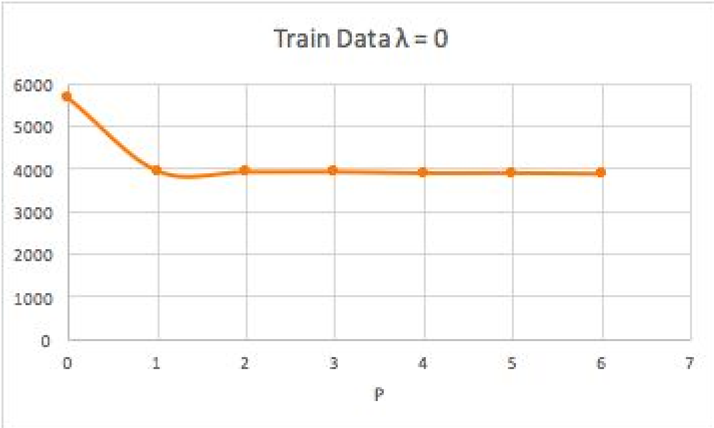


As we can see from the plot above, for the “Train data”, the error when we use the direct scipy.minimize is more than the error when we use “Direct Minimization”; But as the data set is getting larger, the error when we use the scipy.minimize is less than the error that generated by using the direct minimization;
For the “Test data”, for small data set, using scipy.minimize, the errors is somehow more or less than the error generated by using the direct minimization. For the larger data set, the errors is less than (Or “More than” when the lambda is close to the 0.8.) the error generated by using the direct minimization.

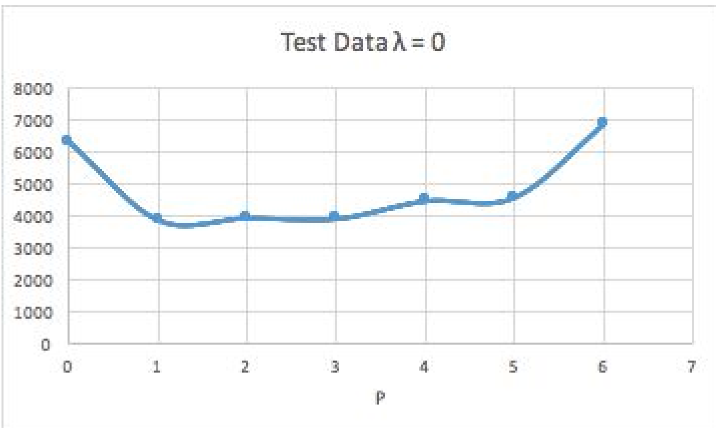
5.

Using $\lambda = 0$ and the optimal value of λ which we found to be 0.06 in Problem 3. We varied p from 0 to 6. The graphs below show the results of these settings on both the test data and the training data. The plot is of the mean square error vs p.

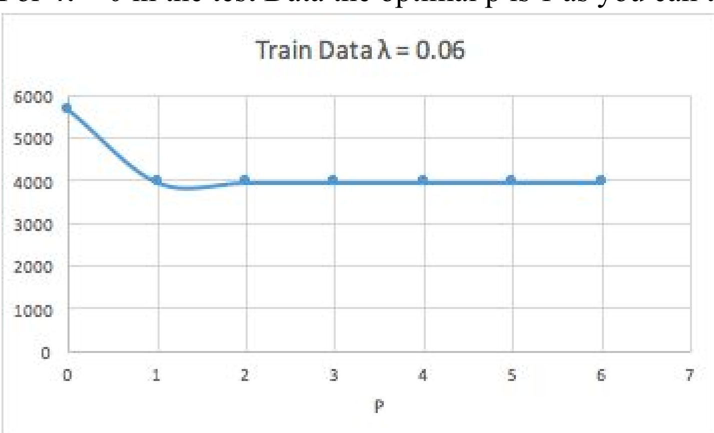
The optimal value of P of test error for $\lambda = 0$ is 1 as you can clearly see in the MSE for Test Data Graph the MSE is the lowest when p = 1. Similarly when $\lambda = 0.06$ the optimal value of p is 6.



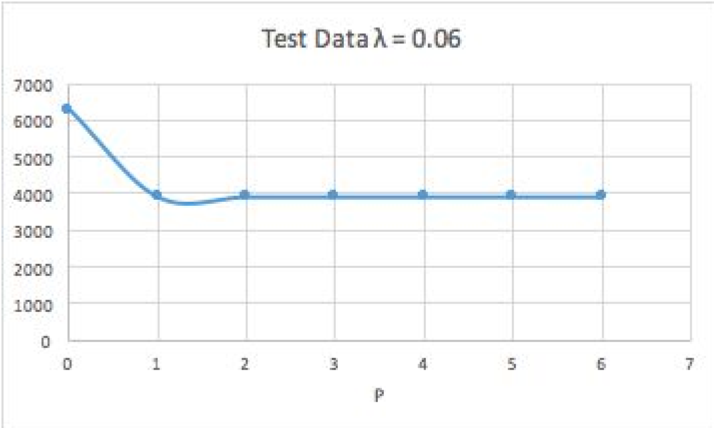
For λ = 0 in the training Data the optimal p is 6 as you can tell by the graph.



For λ = 0 in the test Data the optimal p is 1 as you can tell by the graph.



For λ = 0.06 in the training Data the optimal p is 1 as you can tell by the graph..



For $\lambda = 0.06$ in the test Data the optimal p is 1 as you can tell by the graph.

Overall all of the 1 seemed to be the optimal p in most of the settings that we witnessed. This is true for both when $\lambda = 0.06$ and $\lambda = 0$.

6.Compare the various approaches in terms of training and testing error. What metric should be used to choose the best setting?

As we can see, gradient descent without Ridge regression is better than ridge regression with gradient descent when the data set is small. On big data set, ridge regression with gradient descent is better.