CSE 574

Intro to Machine Learning

Assignment 1

Submitted by: Tushar Deswal

Harsimranpreet Singh

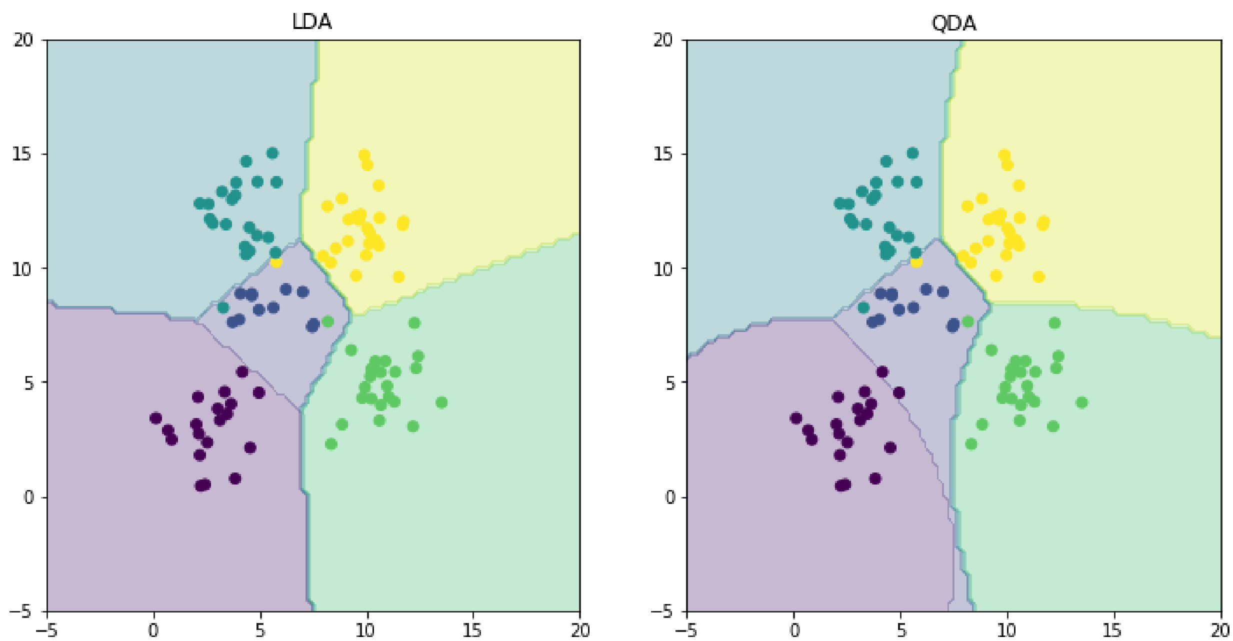
Question 1:

Accuracy of the models:-

LDA = 97

QDA= 94

Plot of the discriminating boundary:



The boundary of LDA is linear, as expected and that of QDA is non-linear.

LDA performs marginally better than QDA in this case as the training data here is less. If training data increases, QDA would do a better job maybe. The difference in the boundaries is because of the fact that QDA assumes different covariance matrix for the classes whereas LDA takes it as same for all the classes.

Question 2: Training error for both the cases:

testOLERegression(w,X,y) : MSE without intercept: 19099.446844570746

testOLERegression(w_i,X_i,y) : MSE with intercept: 2187.16029493

Test error for both the cases:

MSE without intercept 106775.36155789059

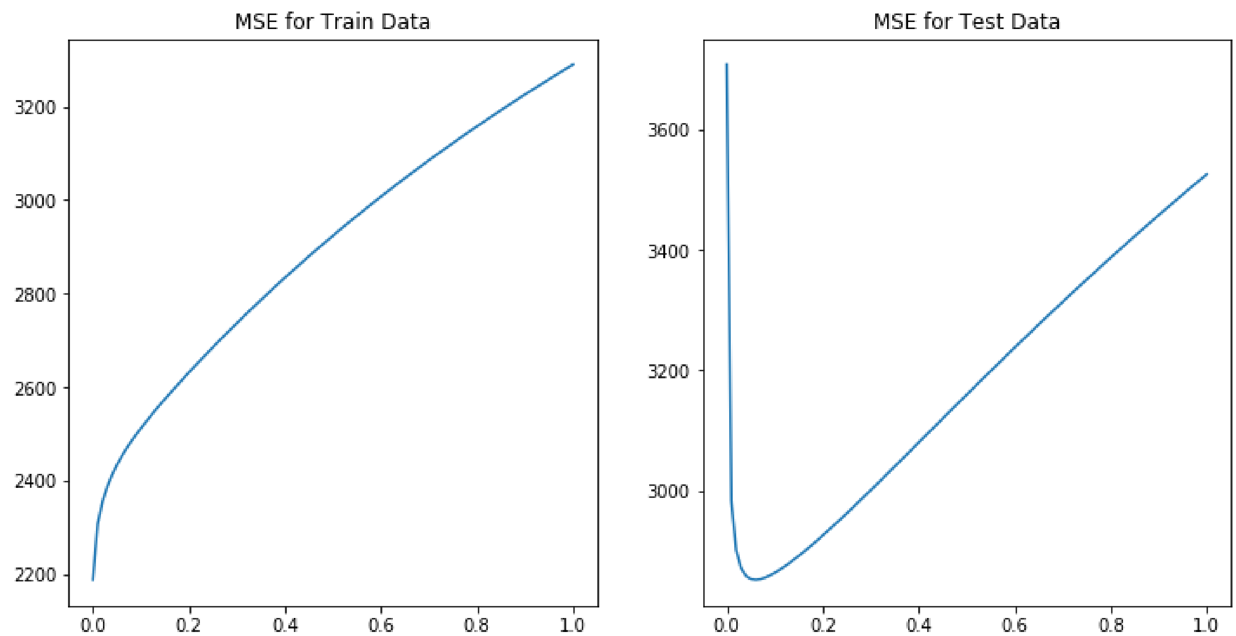
MSE with intercept 3707.8401813150163

Interpretation: MSE with intercept is better in both the cases- train and test.

Train MSE values(with or without intercept) are better for the training sample.

Question 3:

Plot of MSE: Train & Test



Least value of test MSE is 2851.330213443848 at lambda=0.06.

Optimal value of lambda is 0.06.

For train MSE, the increase in lambda increases the MSE because of which the model would not do good(especially for values of lambda over 0.1 in this case).

The test MSE hits minimum at the optimal value of lambda(which is 0.06), after that MSE increases with increase in lambda.

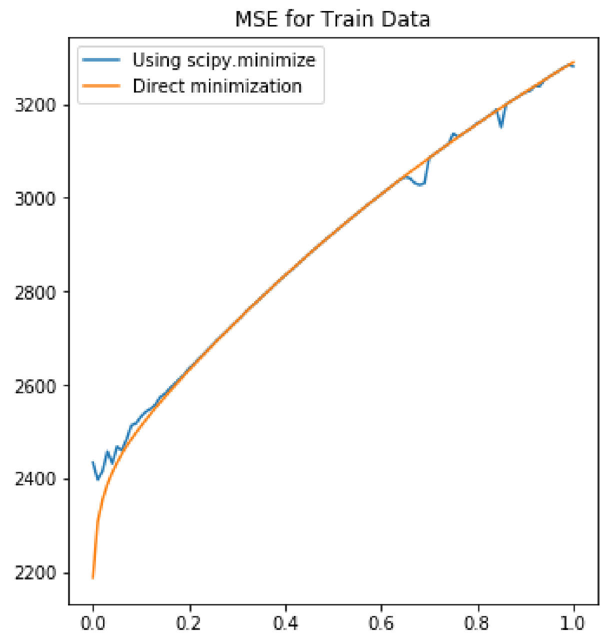
<u>Weights learnt by OLE Regression vs Ridge Weights</u>:

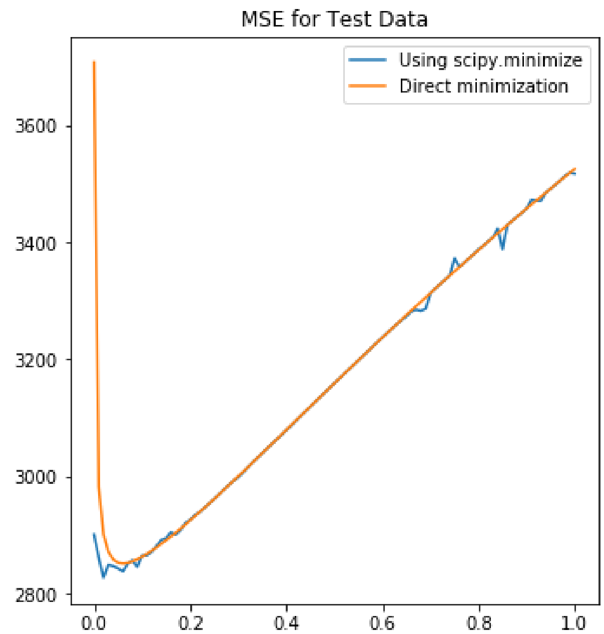
Because of the penalty imposed in Ridge(lambda), the weights in Ridge are very low(for most of the cases) as compared to that in OLE.

```
[[-4.12173302e+02]
 [-3.45940349e+02]
 [ 5.78814085e+02]
 [ 5.89243805e+01]
 [-1.35891612e+06]
 [ 1.19462263e+06]
 [ 5.07036457e+05]
 [-1.34586972e+03]
 [ 4.47713283e+05]
 [ 4.77903799e+02]
 [-1.40658376e+02]
 [-9.19340329e+02]
 [-3.95968904e+02]
 [-7.25692640e+04]
 [-8.95093665e+04]
 [-3.23782609e+03]
 [ 1.40730022e+03]
 [ 3.91795186e+04]
 [ 2.65084262e+02]
 [ 5.12843509e+02]
 [ 2.01158086e+02]
 [ 6.99140521e+01]
 [-4.24307012e+03]
 [ 3.44644877e+03]
 [ 2.22399630e+03]
 [-1.77734526e+02]
```

```
array([[ -40.67699768],
       [ -42.25642637],
       [ 170.42723513],
       [ 131.89883618],
       [ -43.92802254],
       [ -42.93775201],
       [-115.96066513],
       [  52.4991263 ],
       [ 155.47663294],
       [ 127.90564113],
       [  45.41490248],
       [ -45.81437224],
       [ -20.38609793],
       [ -29.8017071 ],
       [ -36.51161981],
       [ -31.96769283],
       [  33.78073578],
       [ -70.97487785],
       [  67.82944814],
       [ 107.85064564],
       [  53.42760794],
       [  60.33513755],
       [ -12.70340642],
       [ -66.90393966],
       [ 106.03251587],
       [ -68.21438398],
       [  38.3554746 ],
       [  12.68876711],
```

Question 4:





Comparing with results from Problem 3:

Minimum value of test MSE(with gradient descent): np.amin(mses4) : 2826.9525654027425 at lambda = 0.02

Minimum value of test MSE(without gradient descent): np.amin(mses3) : 2851.330213443848 at lambda = 0.06

Minimum value of train MSE(without gradient descent): np.amin(mses3_train):2187.1602949303892 at lambda = 0

Minimum value of train MSE(with gradient descent): np.amin(mses4_train): 2396.442108941406 at lambda=0.01

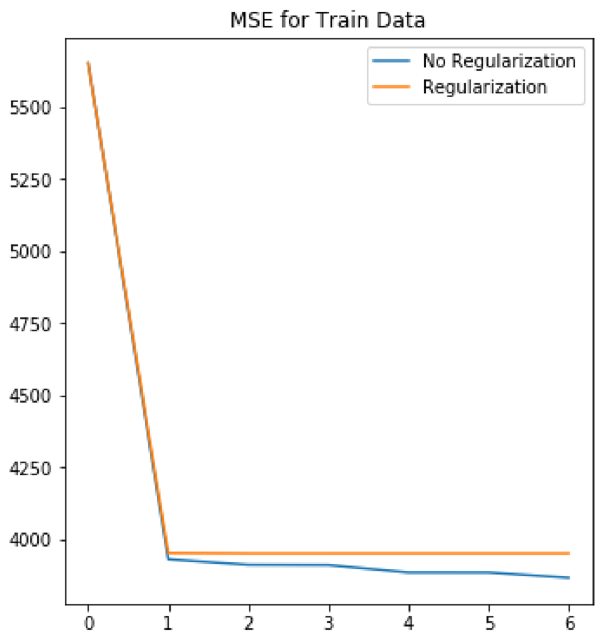
Interpretation: Using gradient descent, least value of MSE test is achieved for lambda 0.02.

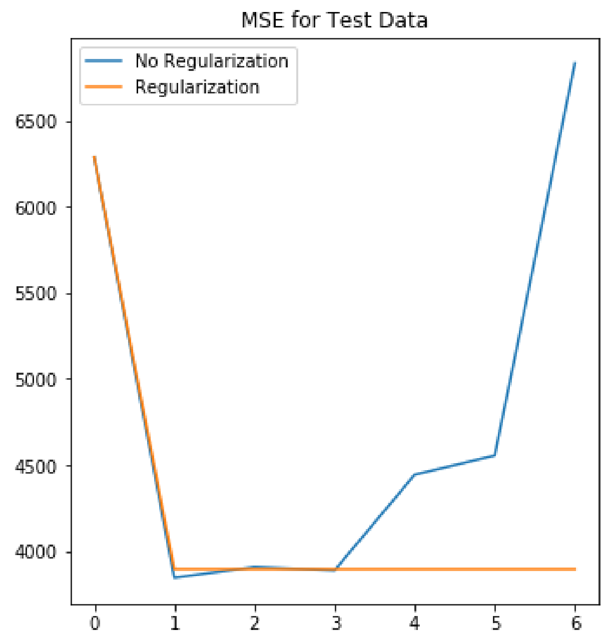
Question 5:

MSE for train, test plot with and without regularization.

We can see that if we don’t use regularization, the error keeps on decreasing for training sample. This would lead to over fitting as the model would only depend(learn) on the train data and as a result of it the test MSE would go high which can be observed from the MSE, test data plot here.

If we use regularization, the train and test MSE decreases and after a certain point remains constant because of the penalty imposed which is a good sign especially for test. Using regularization, we have avoided over fitting and minimized test error.





Question 6:

Out of all the methods, Gradient descent with regularization is the best method here. It avoids over fitting and obtains least value of test MSE at lambda=0.02.