CSE574
Programming Assignment 2

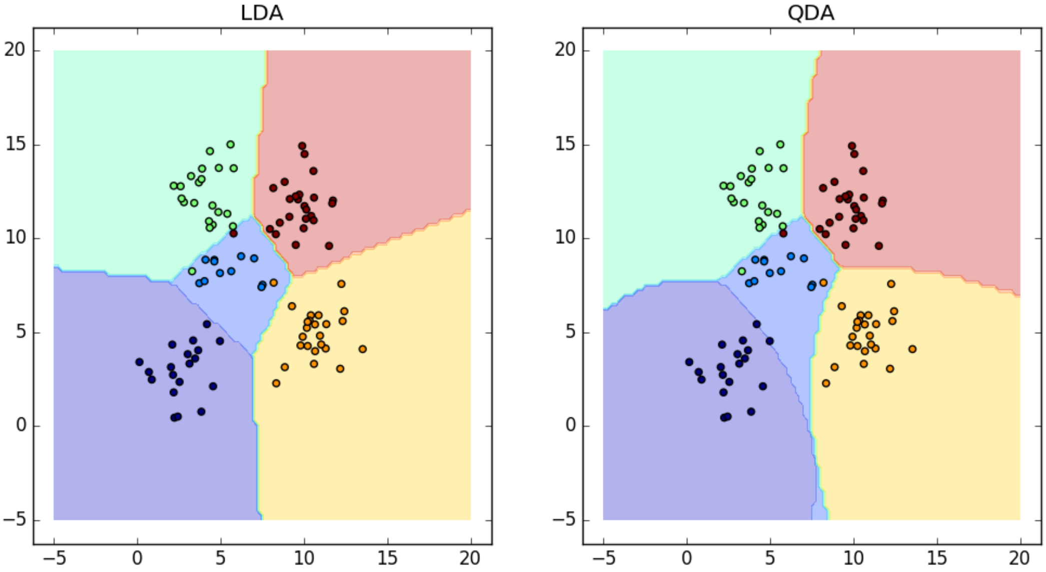
# Classification and Regression

574 Group 46

| | |
|---|---|
| Darshana Balakrishnan | 50207616 |
| Dhruva Kumar Srinivasa | 50207093 |
| Rajeev Vaswani | 50204855 |

**Problem 1:**

LDA Accuracy = 97%
QDA Accuracy = 96%




The difference in the two boundaries is because LDA boundaries are linear due to the non-existence of quadratic term.

**Problem 2:**

Train data:
MSE without intercept [[ 19099.44684457]]
MSE with intercept [[ 2187.16029493]]

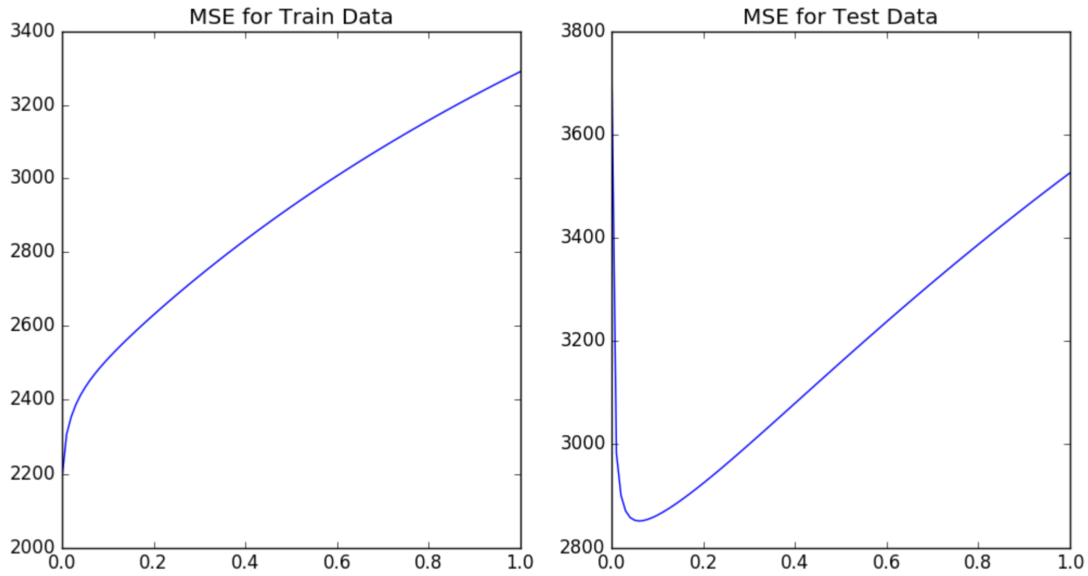
Test data:
MSE without intercept [[ 106775.36155512]]
MSE with intercept [[ 3707.84018132]]

As we can see from the above values, using an intercept gives better results in both cases i.e train data and test data. The intercept acts as a bias term for the classifier thereby always producing an output class ∈ k even when the sample does not have valid data to classify it i.e X = 0s.

**Problem 3:**

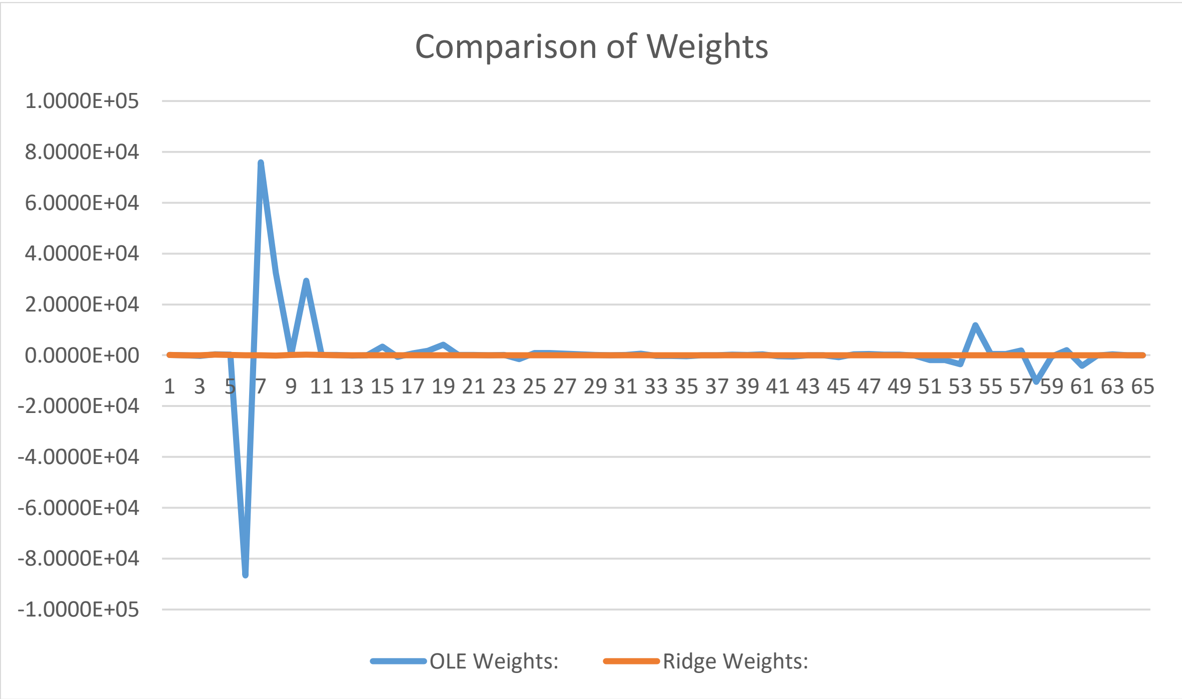
Train MSE:  2187.16029493
Test MSE:  2851.33021344

Plot of MSE for Train and Test data for different lambda values 0 to 1 in steps of 0.01




Weight comparison between Ridge and OLE Regression:

As observed in the below graph, the weights obtained from Ridge Regression are normalized whereas the weights obtained from OLE Regression are not. This is due to fact that the insignificant features are penalized in Ridge Regression.



OLE Regression:
Train data MSE with intercept 2187.16029493
Test data MSE with intercept 3707.84018132

Ridge Regression:
Train data MSE with intercept 2187.16029493
Test data MSE with intercept 2851.33021344

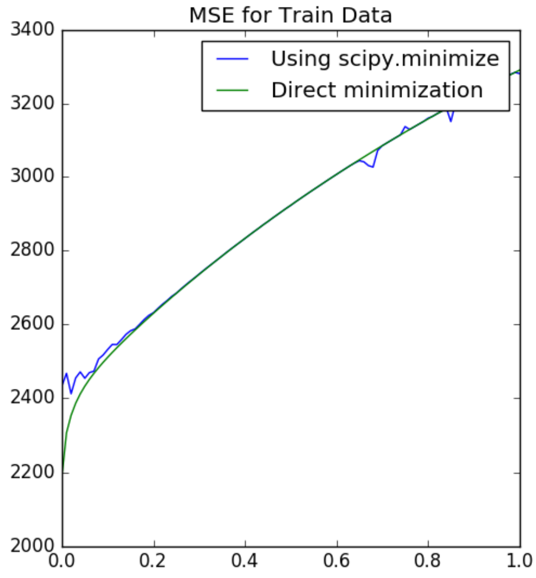
As we can see, Ridge Regression gives a better result for Test data but there is no change in the result for Train data.

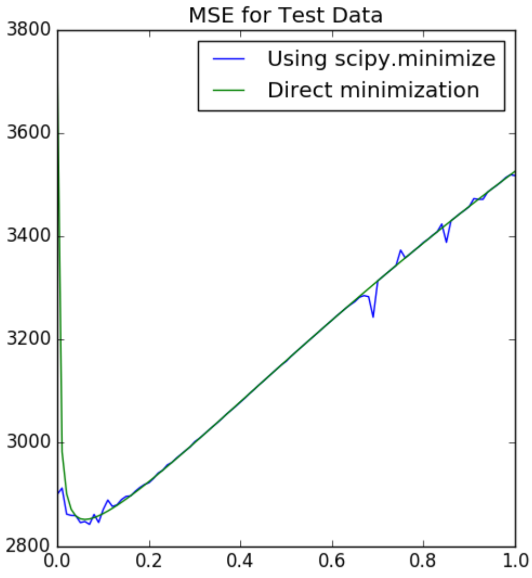
When Lambda = 0.06, the MSE for Test data is the lowest.
Therefore, Optimal Lambda = 0.06

**Problem 4:**

Using the minimize method to estimate the weights by minimizing the error. The below plot compares the MSE for Train and Test data obtained by using direct minimization from the previous problem and by using the scipy.minimize method for lambda values ranging from 0 to 1 in increments of 0.01.

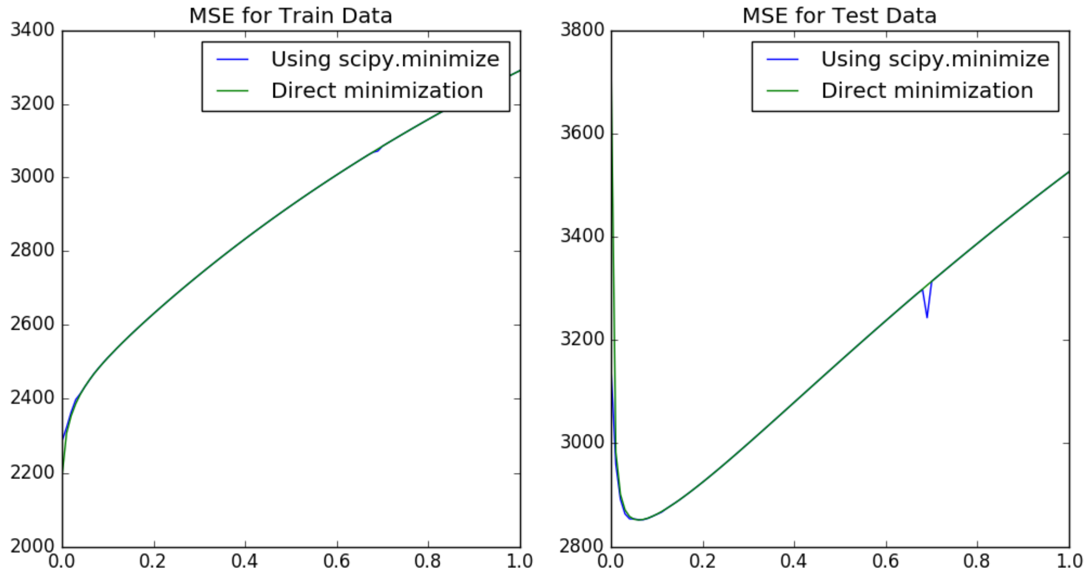
20 iterations



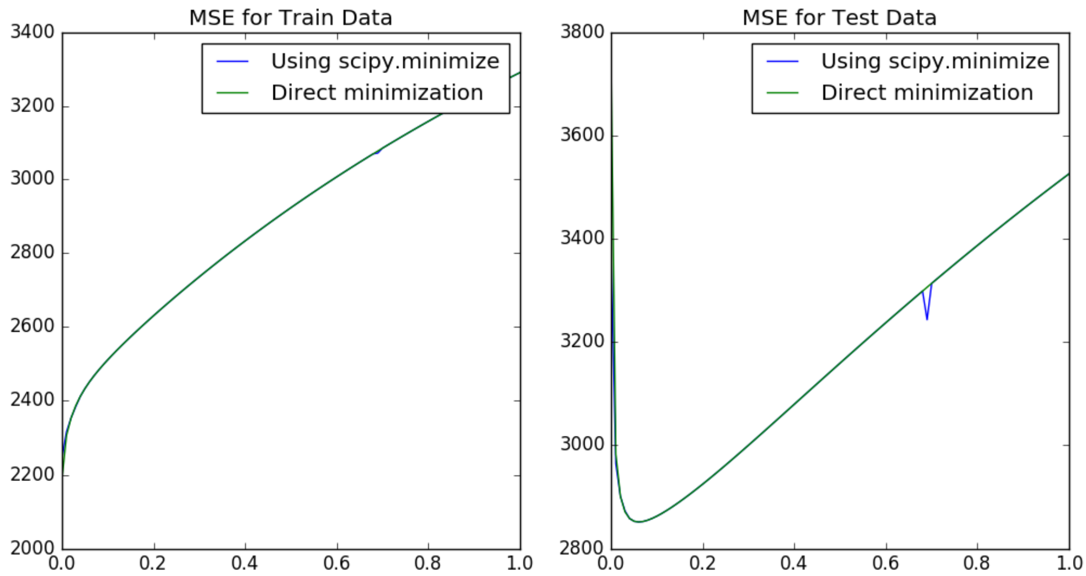




The minimize method can be used to obtain results similar to direct minimization by increasing the number of iterations. The below plots show that both curves are nearly identical for higher iterations.

50 iterations




100 iterations



**Problem 5:**

Train data

| | lambda = 0 | lambda = 0.06 |
|---|---|---|
| p = 0 | 5650.71053890 | 5650.71190703 |
| p = 1 | 3930.91540732 | 3951.83912356 |
| p = 2 | 3911.83967120 | 3950.68731238 |
| p = 3 | 3911.18866493 | 3950.68253152 |
| p = 4 | 3885.47306811 | 3950.68233680 |
| p = 5 | 3885.40715740 | 3950.68233518 |
| p = 6 | 3866.88344945 | 3950.68233514 |

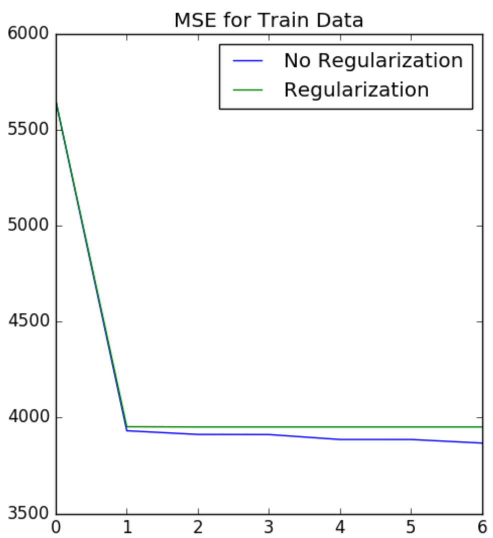
From the data shown above,
Lowest MSE with lambda = 0 is 3866.88344945 at p = 6
Lowest MSE with lambda = 0.06 is 3950.68233514 at p = 6
The tabulated values for MSE do not vary much with increasing values of p.

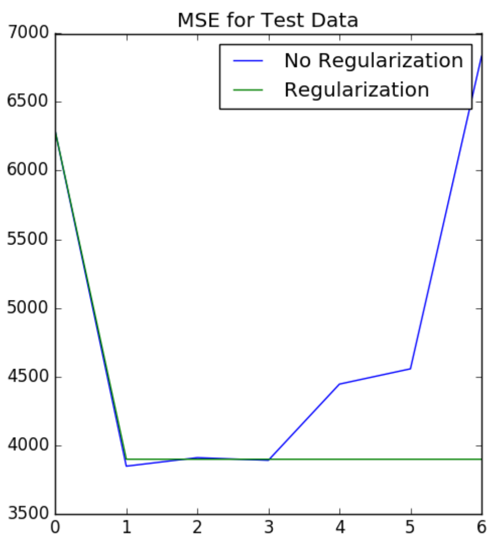
Test data

| | lambda = 0 | lambda = 0.06 |
|---|---|---|
| p = 0 | 6286.40479168 | 6286.88196694 |
| p = 1 | 3845.03473017 | 3895.85646447 |
| p = 2 | 3907.12809911 | 3895.58405594 |
| p = 3 | 3887.97553824 | 3895.58271592 |
| p = 4 | 4443.32789181 | 3895.58266828 |
| p = 5 | 4554.83037743 | 3895.58266870 |
| p = 6 | 6833.45914872 | 3895.58266872 |

From the data shown above,
Lowest MSE with lambda = 0 is 3845.03473017 at p = 1
Lowest MSE with lambda = 0.06 is 3895.58266828 at p = 4

The tabulated results show that for the optimal value of lambda the MSE values do not vary with increasing values of p. The optimal lambda (0.06) is a better regularization factor than 0.



**Problem 6:**

The MSE values obtained from Problem 2 – 5 for Train and Test are:
Train
OLE: 2187.16029493
Ridge: 2187.16029493
Ridge with gradient descent: 2246.10854366
Non-linear: 3950.68233514

Test
OLE: 3707.84018132
Ridge: 2851.33021344
Ridge with gradient descent: 2851.3344498
Non-linear: 3895.58266828

From the obtained results, Ridge Regression would be the best option for anyone using regression for predicting diabetes level using the provided features.
Since Ridge Regression incorporates regularization, overfitting can be avoided. Ridge Regression also helps in reducing the impact of correlated features in the input data.

Among the Ridge Regression implementations in this project, using gradient descent to minimize the square loss would be the more efficient approach. The analytical approach requires the computation of $(X^T X)^{-1}$ which would be computationally expensive for data with large number of features (of the order $d^3$ where d is the number of features). Using gradient descent will provide results very close to the analytical approach thereby making it a better option.

Therefore, from our implementation, Ridge Regression using gradient descent would be the best technique to predict diabetes level using the input features.