Programming Assignment 1

Classification and Regression

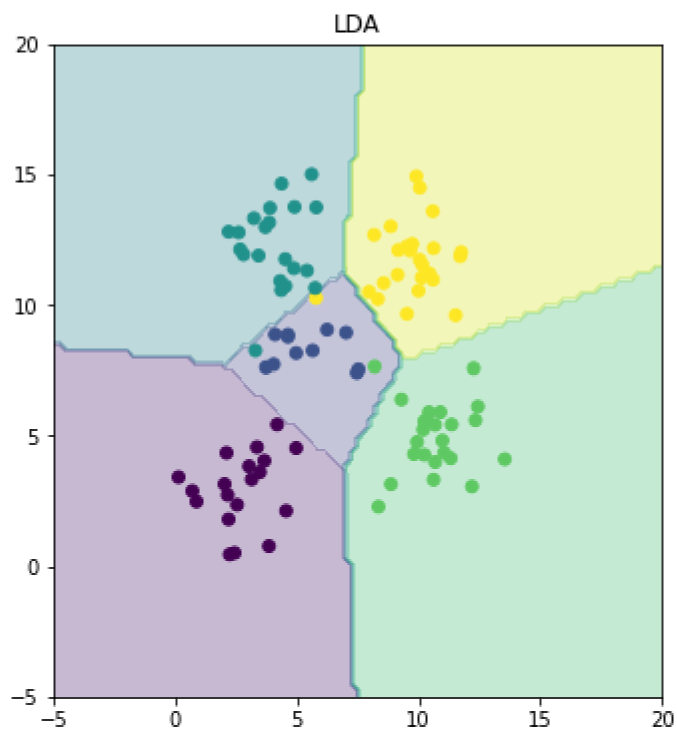
Team Members:

Pinaz Shaikh (50413528)
Janhavi Desale (50412898)
Krati Sharma (50416821)

Problem 1: Experiment with Gaussian Discriminators

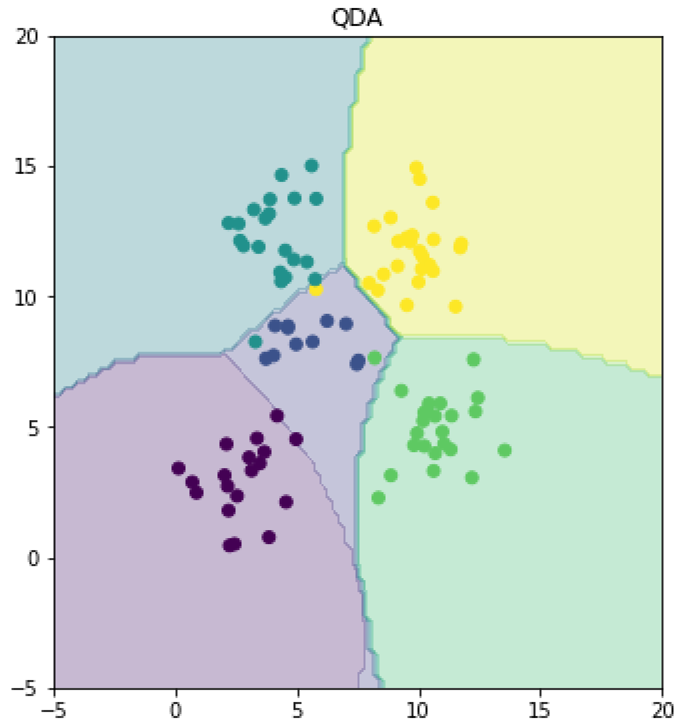
Linear Discriminant Analysis:

Accuracy: 97%



Quadratic Discriminant Analysis:

Accuracy: 96%



There is a difference between two boundaries as plotted above because QDA parameters are quadratic and also it estimates covariance matrix for each class whereas LDA uses single covariance matrix for all classes.

Problem 2: Experiment with Linear Regression

Output

```
MSE for test data without intercept 106775.36156176544
MSE for test data with intercept 3707.8401817245654
MSE for train data without intercept 19099.446844570768
MSE for train data with intercept 2187.160294930389
```

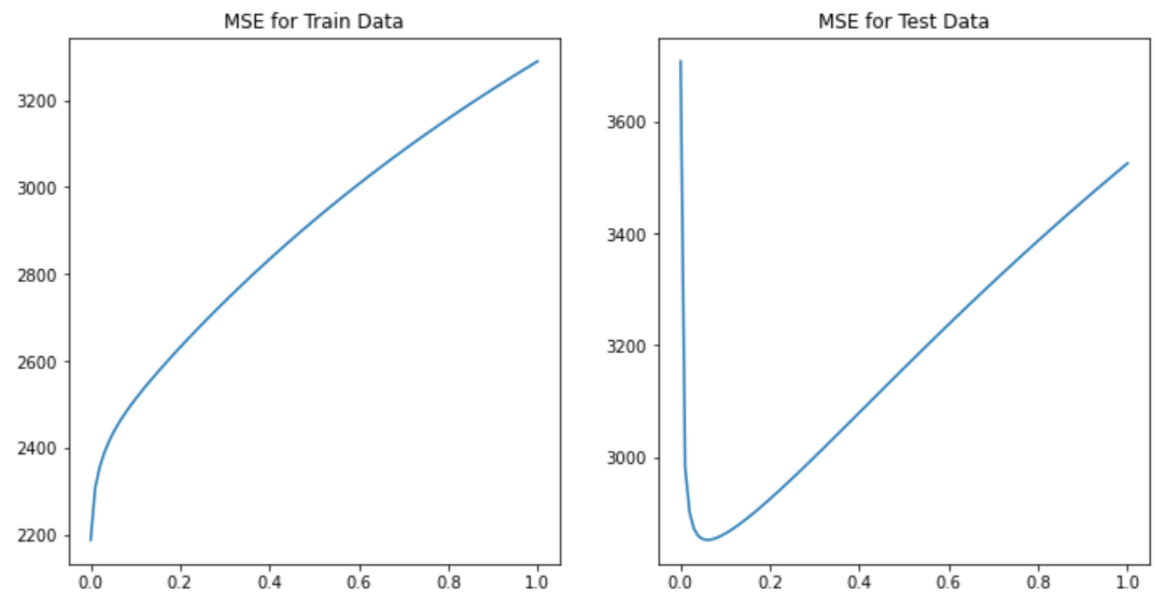
The better one is with intercept since the smaller RMSE, the better fit the model is. Regression line, when plotted in a graph, might pass through the origin if an intercept is not added. This causes the line to be forced to go through the origin. Hence adding an intercept rules out such anomalies.

Thus for Linear Regression we have :

Model	**Train Data**	**Test Data**
With Intercept	2187.160	3707.84
Without Intercept	19099.446	106775.361

Problem 3: Experiment with Ridge Regression

<u>Output</u>:




```
Ridge Regresion, mean of magnitude of weights: 4408.784304987702
The minimum MSE for Test Data occurs at lambda: 0.06 with an MSE of 2851.330213443848
```

<u>Explanation</u>:-

As linear regression, we can fit a model containing all p predictors using a technique that constrains or regularizes the coefficient estimates, and shrinks the coefficient estimates towards zero. It is not immediately obvious that such a constraint should improve the fit, but it turns out that shrinking the coefficient estimates can significantly reduce their variance.

Model	**Testing Data**	**Training Data**
OLE	3707.84	2187.160
Ridge Regression	2851.33	2187.160

When comparing the RMSE value for the two approaches, the Ridge regression is certainly better than the OLE regression for the error in the test data, while the error for the training data is similar for both models.

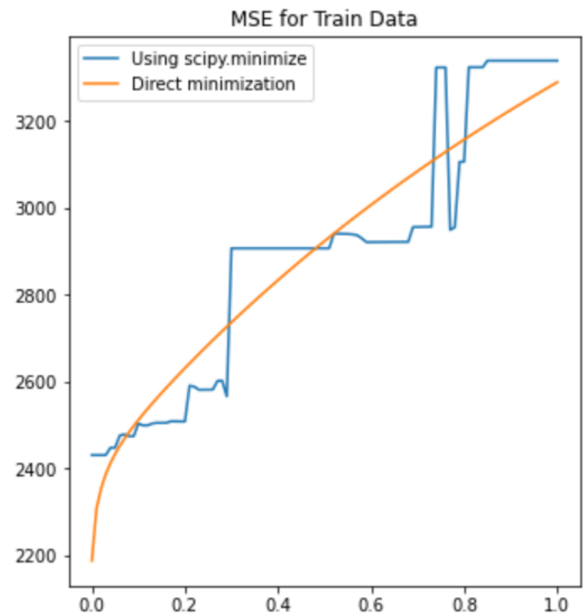
Calculating Lambda Parameter :
As with least squares, ridge regression seeks coefficient estimates that fit the data well, by minimizing RMSE . However, the shrinkage penalty is low when X1, . . . Xp are close to 0, and so it has the effect of shrinking the estimates of Xi towards 0. The tuning parameter λ serves to control the relative impact of these two terms on the regression coefficient estimates. When λ = 0, the penalty term has no effect, and ridge regression will produce the least squares estimates. However, as λ, the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will approach 0.

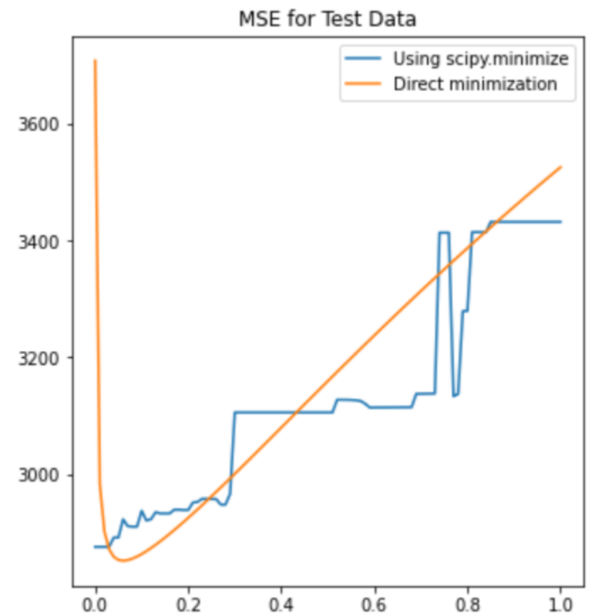
Lambda	MSE
0	3707.84018172
0.01	2982.44611971
0.02	2900.97358708
0.03	2870.94158888
0.04	2858.00040957
0.05	2852.66573517
0.06	2851.33021344
0.07	2852.34999406
0.08	2854.87973918
0.09	2858.44442115

For our problem statement, we have observed that λ = 0.06 is the best optimal value.

Problem 4: Using Gradient Descent for Ridge Regression Learning

Output:





Explanation:-

The lowest value of MSE by applying gradient descent is 2851.33.While comparing this value to minimum value by ridge regression that is in result 3 which is 2851.33 ,the values are the same.

The optimal value of lambda for both the solutions is 0.06.

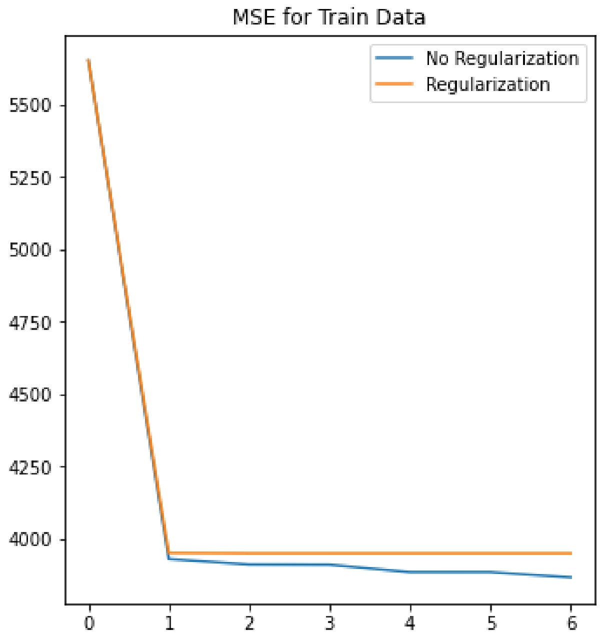
Problem 5: Non-linear Regression

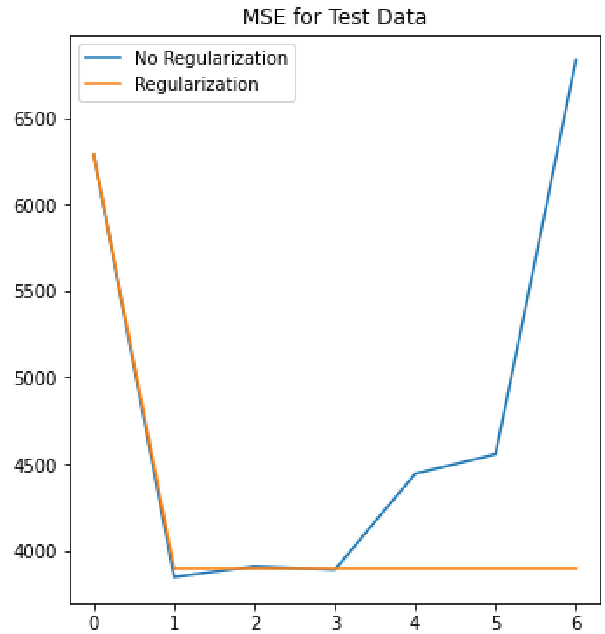
Using the λ = 0 and the optimal value of λ found in Problem 3, train ridge regression weights using the non-linear mapping of the data. Vary p from 0 to 6. Note that p = 0 means using a horizontal line as the regression line, p = 1 is the same as linear ridge regression. Compute the errors on train and test data. Compare the results for both values of λ. What is the optimal value of p in terms of test error in each setting? Plot the curve for the optimal value of p for both values of λ and compare.

Output:

```
Test MSE - Lambda 0
[[6286.40479168 6286.40479168]
 [3845.03473017 3845.03473017]
 [3907.12809911 3907.12809911]
 [3887.97553824 3887.97553824]
 [4443.32789181 4443.32789181]
 [4554.83037743 4554.83037743]
 [6833.45914872 6833.45914872]]

Test MSE - Lambda 0.06
[[6286.40479168 6286.88196694]
 [3845.03473017 3895.85646447]
 [3907.12809911 3895.58405594]
 [3887.97553824 3895.58271592]
 [4443.32789181 3895.58266828]
 [4554.83037743 3895.5826687 ]
 [6833.45914872 3895.58266872]]
```





We can observe from graph and MSE values, that when *lambda* is 0.06 p reaches its optimal value 1 and keeps on increasing slowly till 3 and then grows rapidly after 3 whereas when *lambda* is 0 p is optimal at 4 and remains almost the same till 6.

Problem 6: Interpreting Results

Using the results obtained for previous 4 problems, make final recommendations for anyone using regression for predicting diabetes level using the input features.

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

The results obtained are :-

OLE Regression
MSE for test data without intercept: 106775.36156176544
MSE for test data with intercept: 3707.8401817245654
MSE for train data without intercept: 19099.446844570768
MSE for train data with intercept: 2187.160294930389

Ridge Regression
Optimal value of lambda = 0.06

Training Error : 2187.160
Testing Error : 2851.33

Ridge Regression with Gradient Descent
Optimal λ = 0.06

Testing Error : 2851.33

Non Linear regression

P - Value	MSE
0	3845.03
1	3895.58

Conclusion :

On observing the results which are derived for all the different practices, we can say that Ridge regression or Ridge regression with Gradient Descent is the best way to get the least error (MSE) on the trained data.