Programming Assignment 2
Classification and Regression

CSE 574 - Group No. 49
Miraj Kheni (mirajsub)
Rushin Shah (rushinka)

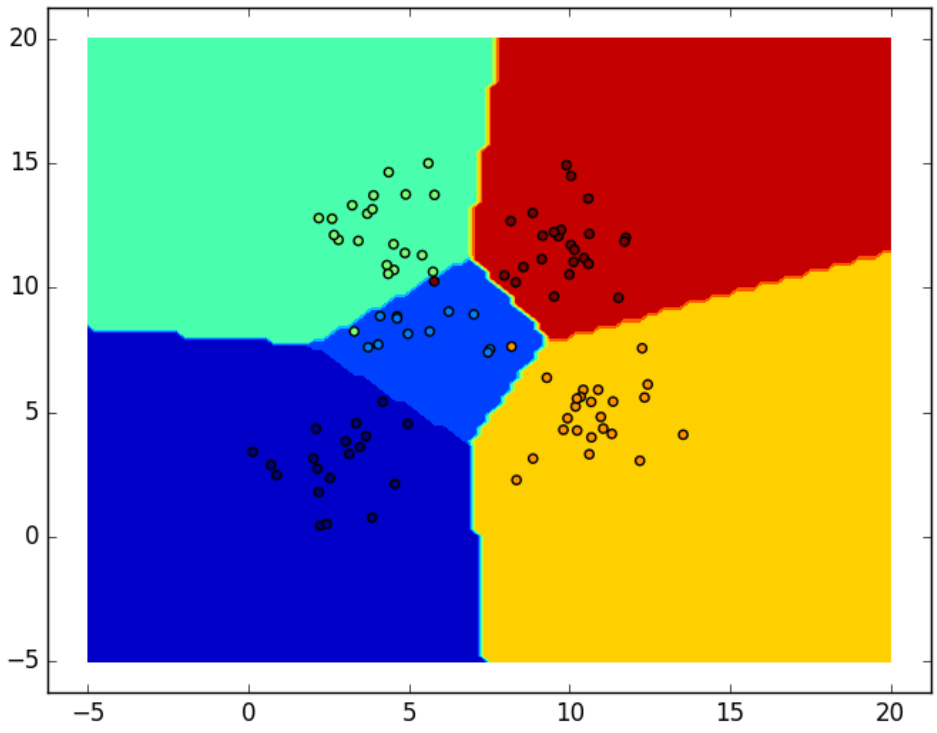
Problem 1 - Experiment with Gaussian discriminators

We achieved below accuracies for LDA and QDA:

LDA Accuracy = 97.0 %

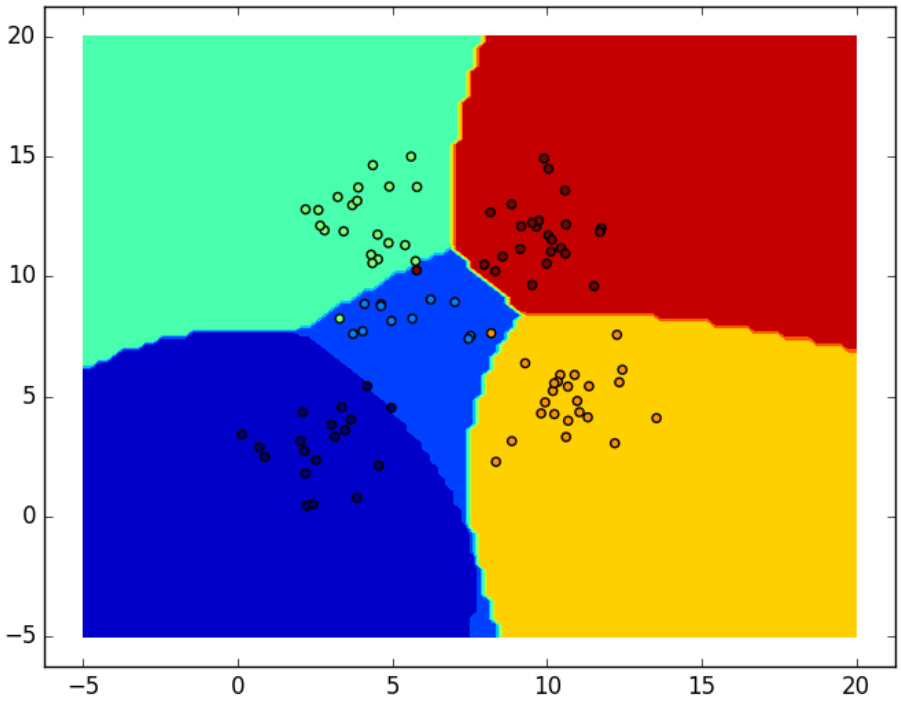
QDA Accuracy = 96.0 %

Discriminating boundary for LDA:



In above figure, you can see discriminant boundaries for LDA which differentiate 5 classes. Boundaries for LDA are linear as we assume that covariance matrix for different classes are same.

Discriminating boundary for QDA:



In above figure, you can see discriminant boundaries for LDA which differentiate 5 classes. Unlike LDA, in QDA there are no assumptions about covariance matrix. Hence, QDA learns quadratic decision boundaries as opposed to LDA which learns linear boundaries.

Problem – 2 Experiment with Linear Regression

RMSE without intercept for testing data: 326.7649943909467

RMSE with intercept for testing data: 60.892037096982534

RMSE without intercept for training data: 138.20074835025636

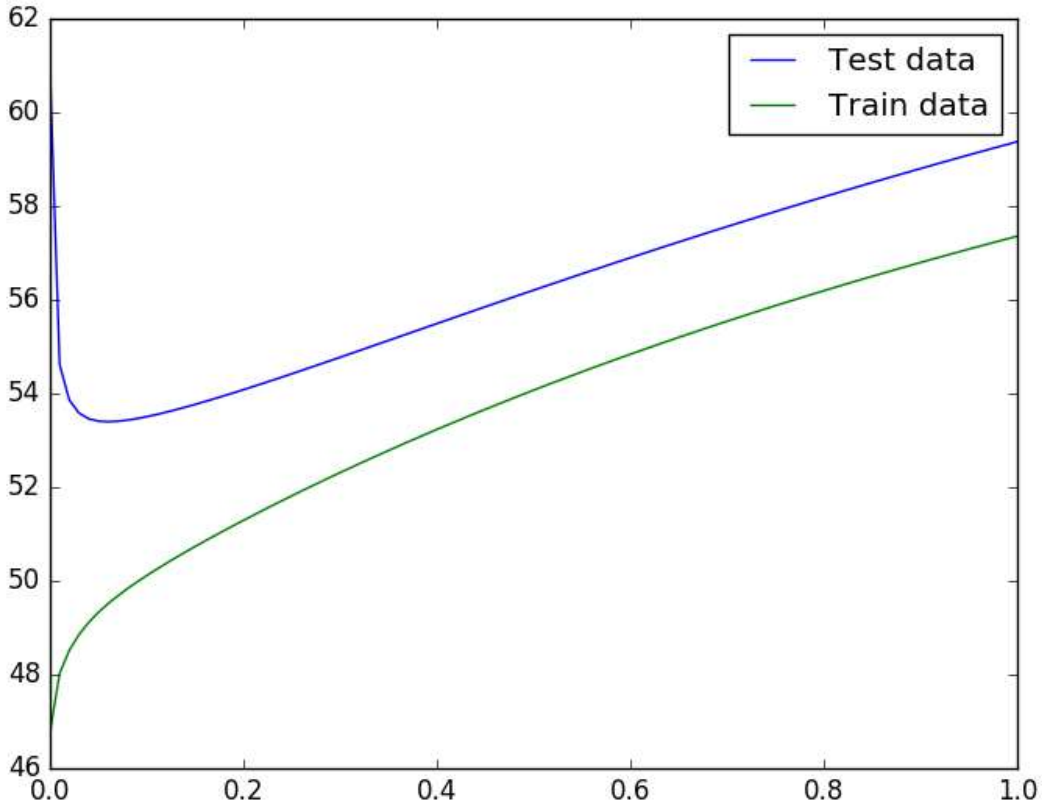
RMSE with intercept for training data: 46.7670855937206

Ordinary Least Squares method works better using intercept both for testing and training data.

Problem – 3 Experiment with Ridge Regression

Errors – Ridge regression:

Below graph shows errors on train and test data for different values of Lambda. Our observation with different values of testing error and lambda shows that optimum value for lambda is: **0.06**.



When we compare OLE and ridge regression method in terms of error on train and test data, **Ridge regression is better** as RMSE value for test data is much lesser in case of Ridge regression.

Weights:

Looking at the weights observed from OLE approach in Problem 2 and Ridge regression, we can conclude that **Ridge regression is better** as weights are lesser in this approach.

(Mean weight for Ridge regression: 17.3219)

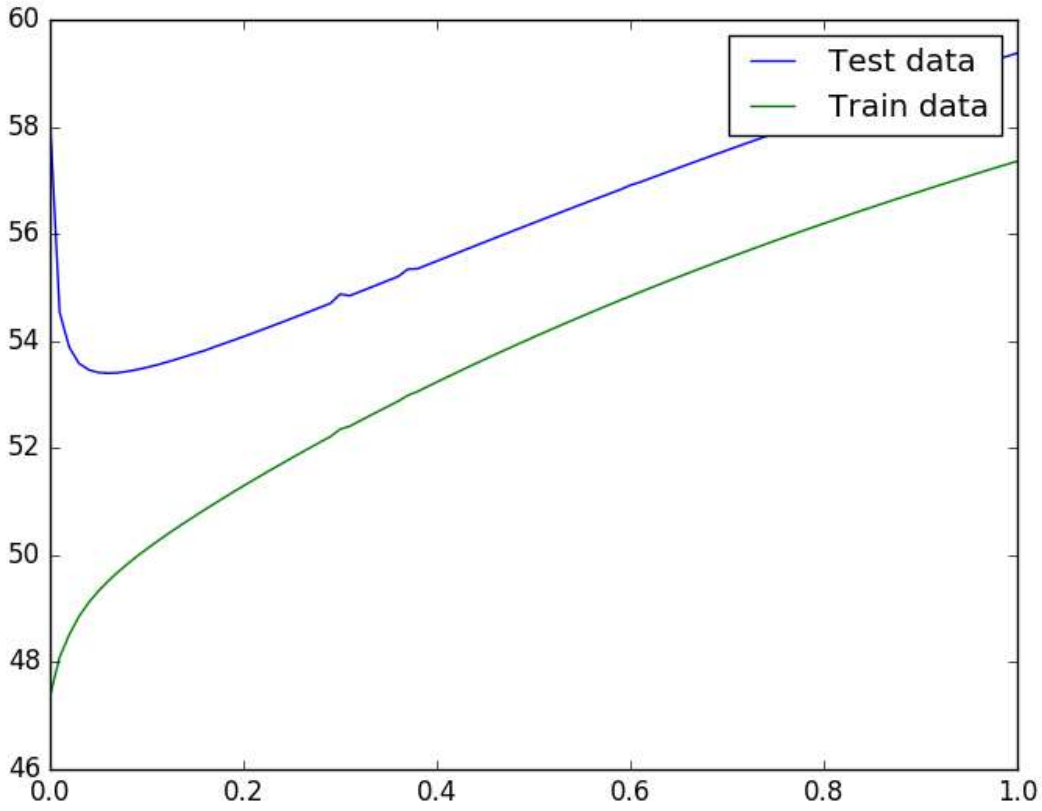
(Mean weight for OLE approach: 882.80762)

Problem 4 - Using Gradient Descent for Ridge Regression Learning

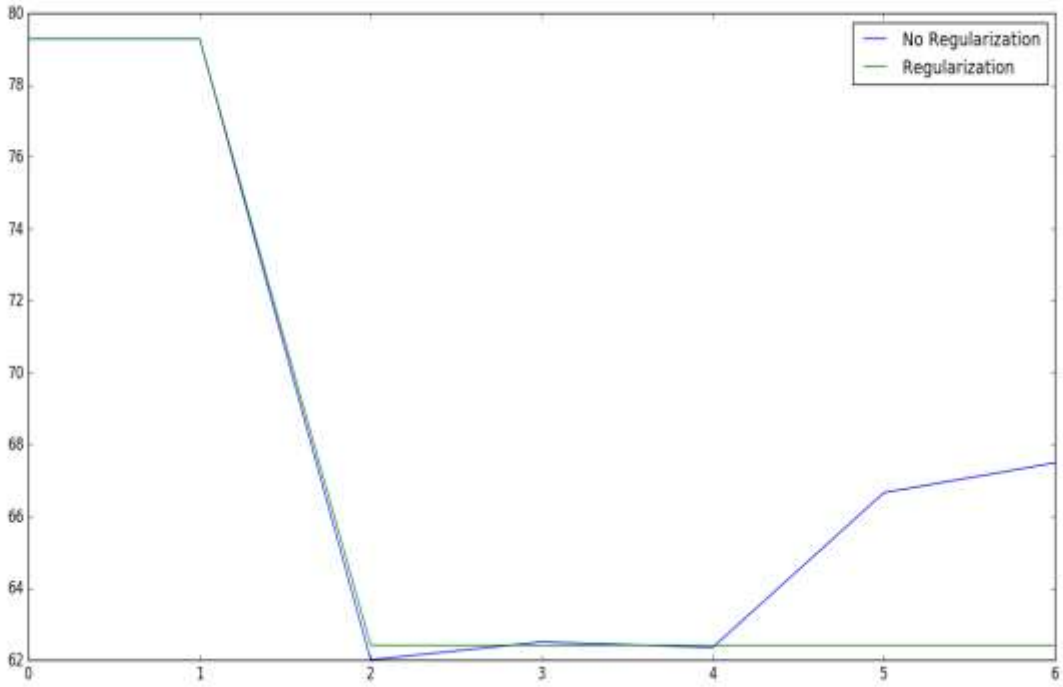
Below graph shows errors on train and test data for different values of Lambda. When we compare the results with ones obtained in problem 3, we come to know that there is no much difference in them.

Conclusion:

Errors obtains in both problems are almost similar.



Problem 5 - Non-linear Regression



Test Data

P	Test Error (Lambda = 0)	Test Error (Lambda optimum = 0.06)
0	79.2868513165	79.2898604296
1	79.2868513165	79.2898604296
2	62.0083440367	62.416796333
3	62.5070243981	62.4146141215
4	62.3536329193	62.4146033867
5	66.6582919959	62.4146030051
6	67.4894834581	62.4146030085

Conclusion:

Optimum **p = 2** when **lambda = 0**

Optimum **p = 5** when **lambda (optimum) = 0.06.**

Problem – 6 Interpreting results

Linear regression:

From problem 2

RMSE without intercept on test data: 326.7649943909467
RMSE with intercept on test data: 60.892037096982534

RMSE without intercept on training data: 138.20074835025636
RMSE with intercept on training data: 46.7670855937206

Ridge regression:

From problem 3

Lambda optimum value = 0.06 and below are error values for Test and train data:

RMSE test data: 53.3978484
RMSE train data: 49.51291236

Weights (mean):
OLE approach weights 882.807625044
Ridge regression weights 17.3219272625

Ridge regression with Gradient descent:

This gives almost similar results compared to Ridge regression.

Conclusion:

Comparing all above results for RMSE values, it can be seen that OLE approach gives lesser RMSE value for Training data and Ridge regression approach gives lesser RMSE value for Test data. But, overall, Ridge regression performs better giving lesser RMSE values for both test and train data.

Ridge regression addresses some of the problems of Ordinary Least Squares by imposing a penalty on the size of coefficients. The ridge coefficients minimize a penalized residual sum of squares.

Therefore, Ridge regression with Lambda value of 0.06 should be used as the best setting for given data.