Classification and Regression (PA-2)

CSE 574 - Introduction to Machine Learning

Team 20

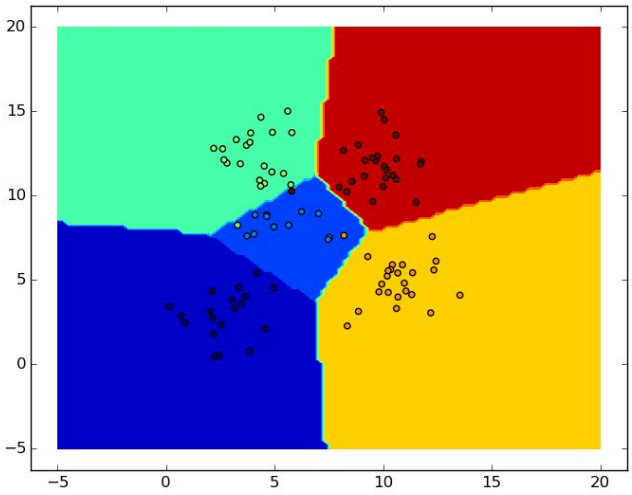
Harish Shankar,

Sukanya Moorthy,

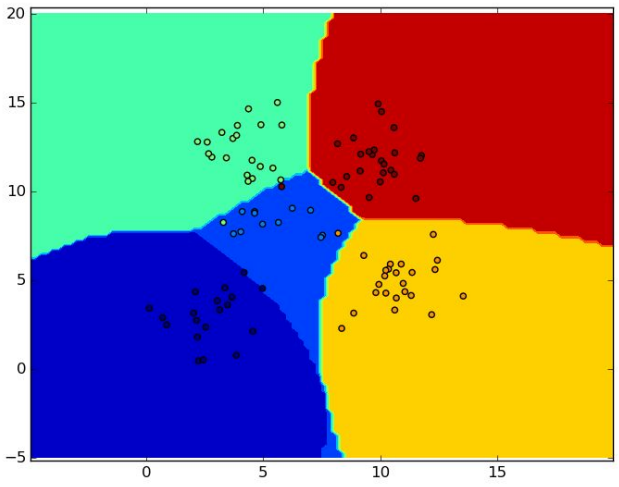
Vasavi Manasa Chindalur Lakshminarayana Gupta

Problem 1 - Experiment with Gaussian Discriminators

LDA - Accuracy Determined : 97%



QDA - Accuracy Determined : 96%



The graphs shown above represent the Discriminating Boundaries for the LDA and QDA. As can be seen, the difference in the 2 boundaries can be attributed to the use of covariance in each of them. Linear - covariance same for all classes, Quadratic - different covariance for each class.

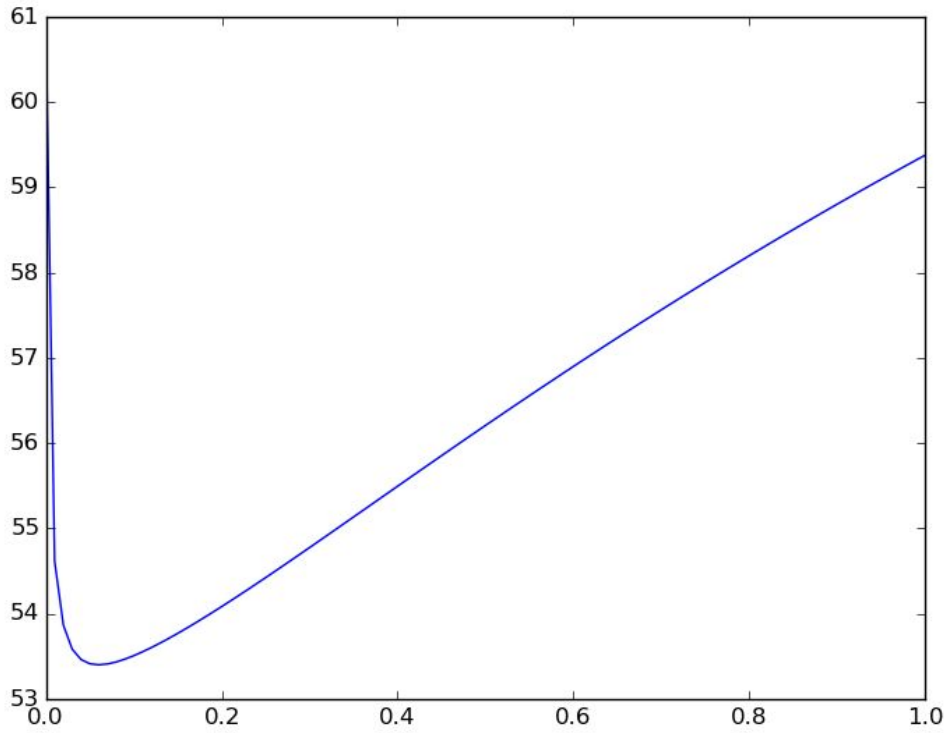
Problem 2 - Experiment with Linear Regression

Test Data	Training Data
RMSE without intercept: 326.7649	RMSE without intercept: 138.2007
RMSE with intercept: 60.8920	RMSE with intercept: 46.7670

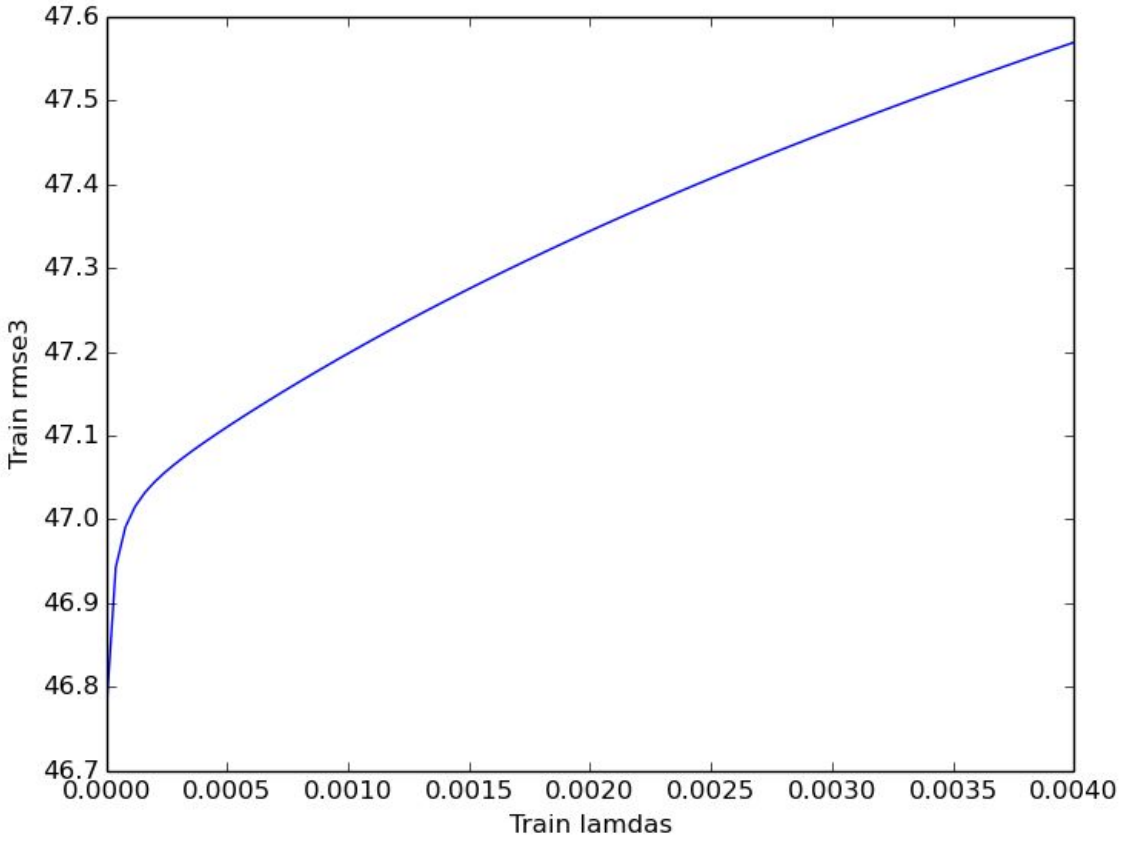
The RMSE value clearly is better/ lesser when used with intercept for both cases. Also, It is observed that RMSE for training data is relatively lower to that of the Test Data, and this could probably be due to the fact that size(Training Data)<size(Test Data).

Problem 3 - Experiment with Ridge Regression

Test Data



Train Data

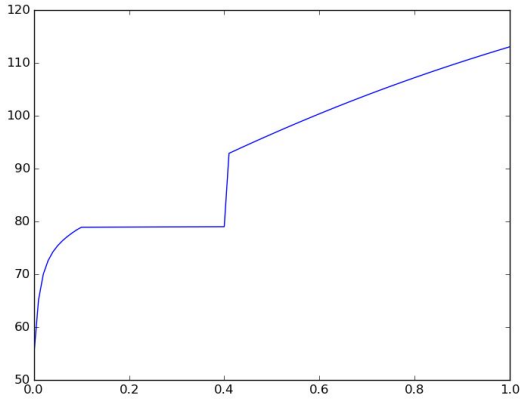


We observe the optimal value of lambda as one where the values converge, and clearly there is a point in the graph where the curve reaches a minima and starts to rise again.

This point, is optimal value of lambda = 0.06

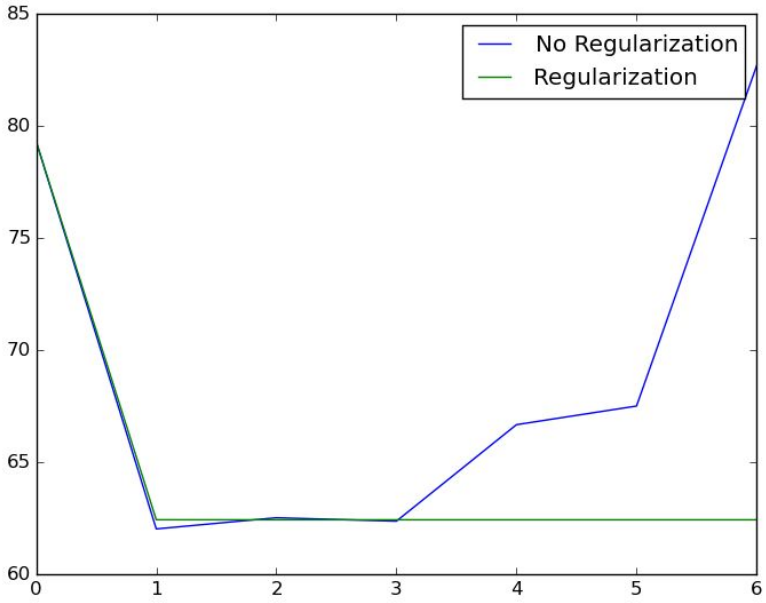
RMSE (Training Data) < RMSE (Test Data) because the number of observations are more in training data

Problem 4 - Gradient Descent for Ridge Regression Learning



Comparing problem 3 and 4, it can be said that the results for both are similar but in the previous case, the performance w.r.t increasing lambda values was better.

Problem 5 - Non Linear Regression



As can be observed from the graph, the Regularized method steadies itself from value p = 1 whilst the one with No Regularization seems to be less efficient.

Problem 6 - Interpreting the Results

The RMSE values being calculated across different is an estimate of the error associated with each method. That is, higher values of RMSE indicate a drift away from the ideal values that is desired, whilst lower values indicate lower errors. The above mentioned 4 methods only indicate that RMSE is dependent on the nature, size of data and that each model is preferred over the other depending on the situation given. However, comparing the results, it can be said that for this given scenario, Ridge Regression or Gradient Descent Ridge Regression can be chosen.