UNIVERSITY AT BUFFALO

CSE 574 INTRODUCTION TO MACHINE   
LEARNING, SPRING 2017

Programming Assignment 2

Classification and Regression

PROJECT REPORT

Submitted by

Anurag Devulapalli (5020 8153)

Sandeep Shenoy (5020 )

Vipin Kumar (5020 8397)

April 11, 2017

Problem 1: Experiment with Gaussian Discriminators

Requirement :

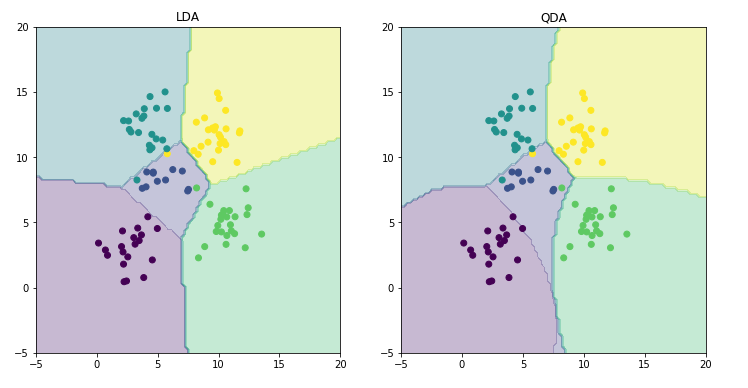
Train both methods using the sample training data (sample train). Report the accuracy of LDA and QDA on the provided test data set (sample test). Also, plot the discriminating boundary for linear and quadratic discriminators. The code to plot the boundaries is already provided in the base code. Explain why there is a difference in the two boundaries.

**a. Report the accuracy of LDA and QDA on the provided test data set (sample test)**

LDA Accuracy = 97

QDA Accuracy = 96

**b. Also, plot the discriminating boundary for linear and quadratic discriminators**

****

**c. Explain why there is a difference in the two boundaries**

For LDA we use a common covariance matrix irrespective of the class. In QDA we use a different covariance matrix for each class. This results in a straight boundary (linear) for LDA whereas the boundaries become slightly curved (quadratic) for QDA.

Problem 2: Experiment with Linear Regression

Requirement :

**a. Calculate and report the MSE for training and test data for two cases: first, without using an intercept (or bias) term, and second with using an intercept. Which one is better?**

|  |  |  |  |
| --- | --- | --- | --- |
| Data Type | Mean Squared Loss (MSE)  Without Intercept | Mean Squared Loss (MSE)  With Intercept | % Improvement |
| Train Data | 19099.4468446 | 2187.16029493 | 88.54% |
| Test Data | 106775.361558 | 3707.84018132 | 96.5% |

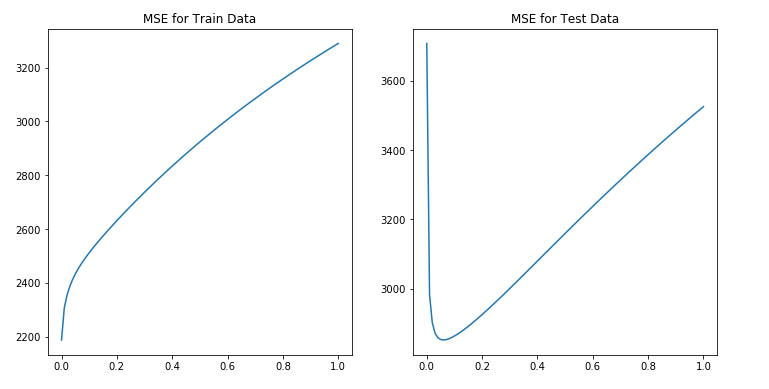
**b. Which one is better?**

In both the cases (training and test data) we observe that MSE with intercept is smaller, hence it is better.

Problem 3: Experiment with Ridge Regression

Requirement :

a. Calculate and report the MSE for training and test data using ridge regression parameters using the the testOLERegression function that you implemented in Problem

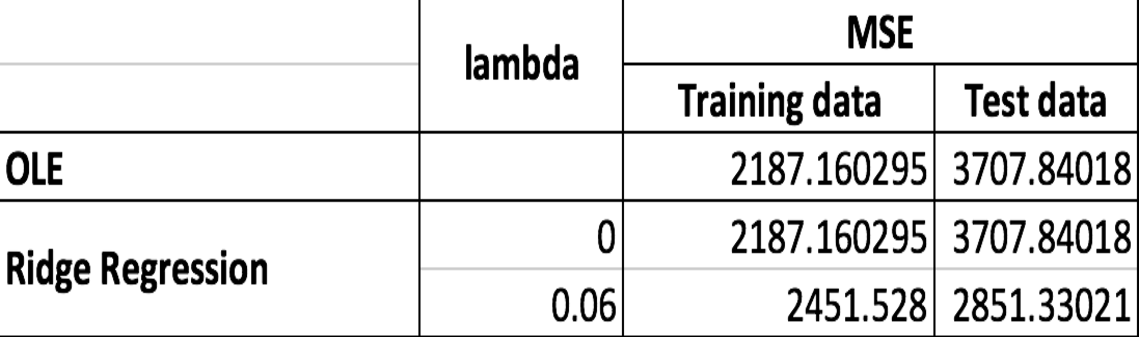


b. Compare the relative magnitudes of weights learnt using OLE (Problem 2) and weights learnt using ridge regression.





c. Compare the two approaches in terms of errors on train and test data.



d. What is the optimal value for λ and why?

Optimal value for λ is 0.06 in case of test data and 0 in case of training data since at this point the MSE is the least.

Problem 4: Using Gradient Descent for Ridge Regression Learning

Requirement :

a. Plot the errors on train and test data obtained by using the gradient descent based learning by varying the regularization parameter λ. Compare with the results obtained in Problem 3.

For iteration = 20



For iteration = 100



For iteration = 150

****

Observation :

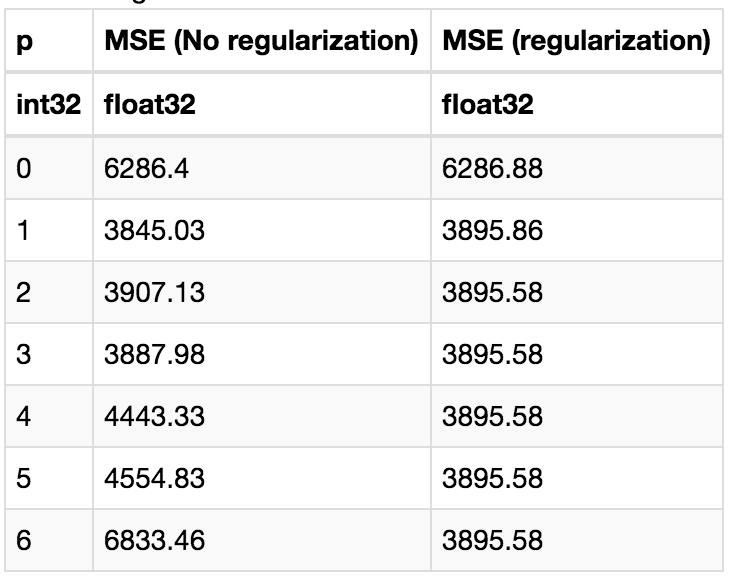
By using scipy.minimize at lower iterations such as 20, the errors obtained deviate a little from the results obtained by direct minimization. As we increase the number of iterations, the scipy.minimize learns more and gives results very similar to the ones obtained by direct minimization.

Problem 5: Non-linear Regression

Description:

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.

**a. Compare the results for both values of λ**

****

b. What is the optimal value of p in terms of test error in each setting?

Optimal p - No regularization 1

Optimal p - With regularization 4

**c. Plot the curve for the optimal value of p for both values of λ and compare.**

****