CSE 574

INTRODUCTION TO MACHINE LEARNING

PROJECT ASSIGNMENT – II

CLASSIFICATION AND REGRESSION

Report done by

CSE 574 Group 14

Anuradha Ashavatha Rao (aashavat)

Nagadeesh Nagaraja (nagadees)

Nagesh Rao (nageshra)

Problem 1:

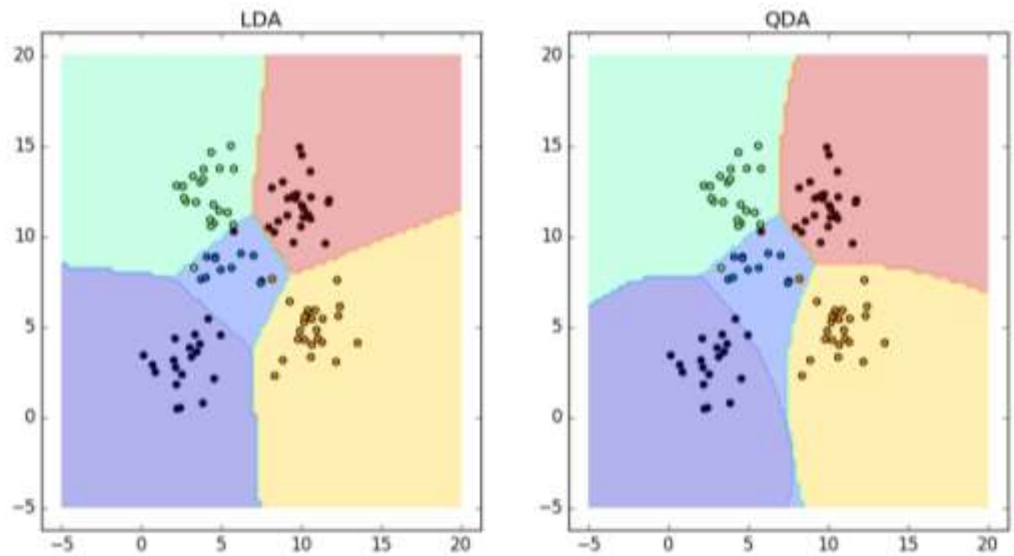
Introduction:

The QDA and LDA classifiers are close-form classifiers that are multiclass in nature. The main difference of these classifiers as compared to Naïve Bayes' classifier is that the covariance matrix is not required to be diagonal. Another advantage of these classifiers is that they do not require additional hyper-parameters to tune to get a result.

A) Accuracy:

i) LDA Accuracy = 97
ii) QDA Accuracy = 96

B) Plotting:



C) Why there is difference?

As we observe from the plots above we can see that the LDA classifier can learn only linear boundaries while the QDA classifier can learn quadratic boundaries. This allows the QDA to be more flexible while classifying data.

Problem 2:

A) Accuracy:

i) Training data

MSE without intercept [106775.36155929]

MSE with intercept [3707.84018148]

B) Which is better?

Linear Regression with intercept is better as it produces the lower error (Mean Square error).

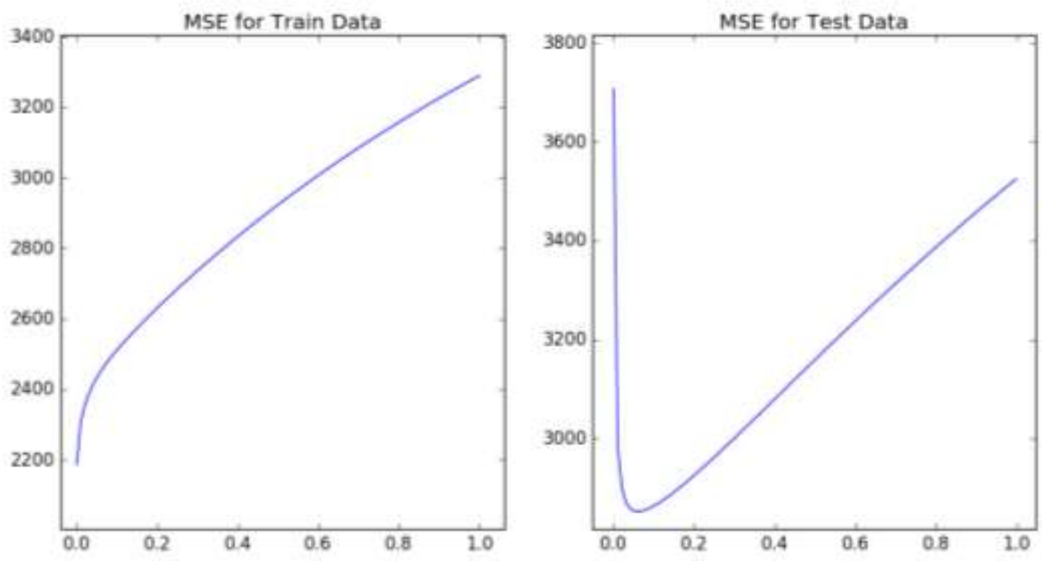
Problem 3:

A) Error (MSE)

For train data??

For test data??

B) Plotting:



C) Comparison between OLE and Ridge:

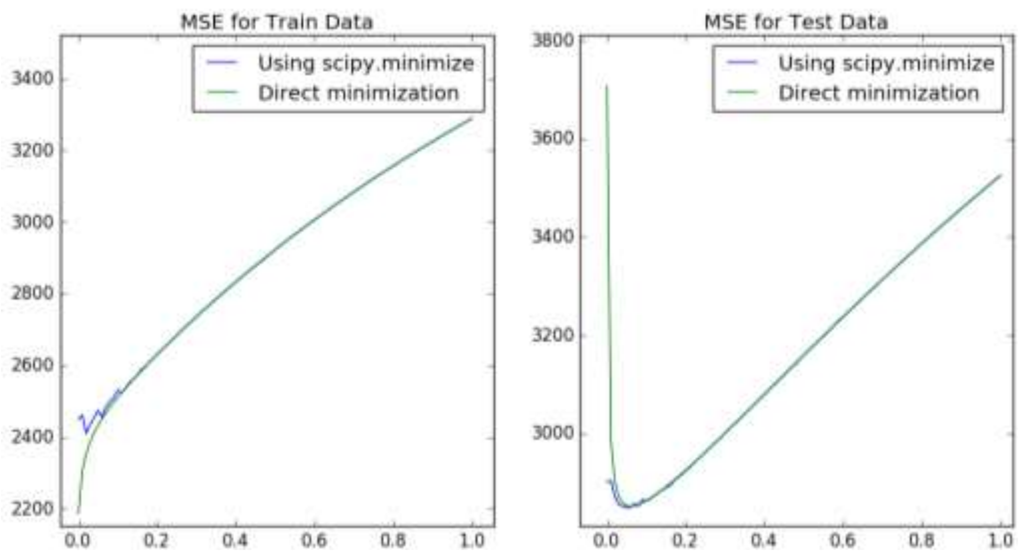
I) Compare two approaches for Error on both train and test data?

II) Weight learnt using OLE and ridge regression?

D) Optimal value of lambda is 0.06 – is the value at which we get lowest error.

Problem 4:

A) Plotting:



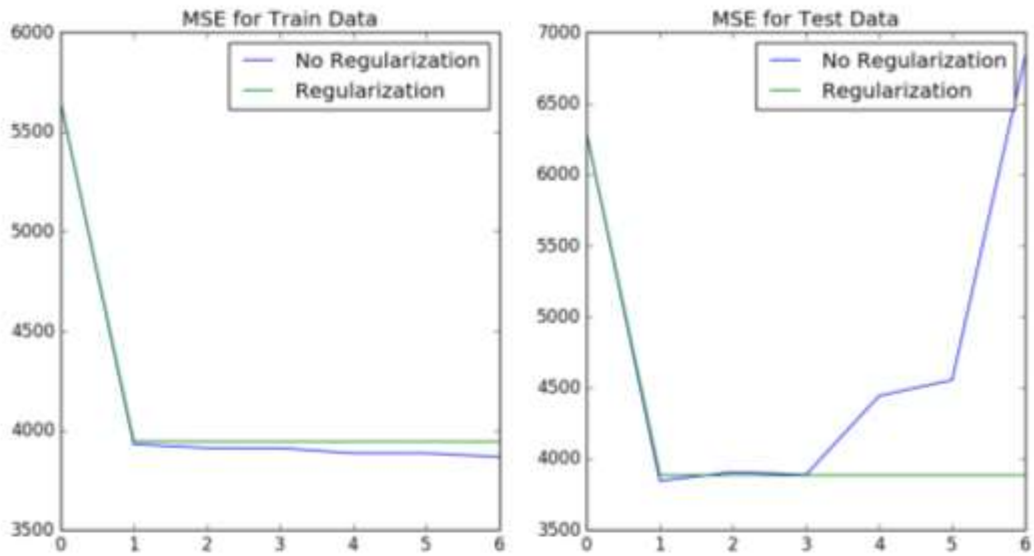
B) Comparison:

We observe that the Gradient descent converges to same values of errors, (and hence weights) as using the matrix formula in the problem 3 (with regularization). We notice some discrepancies between the two when there is no regularization (lambda is ~0). When we use no regularization, the difference comes due to the formulas used, while gradient decent, we remove the whole term (0.5*lambda*weights), however, while using the matrix formula, we add the identity matrix, which causes the different error.

Problem 5:

A) Errors in train and test data:

B) Plotting:



C) Comparison:

We notice that without regularization (lambda =0), the regression performs very well for train data, However, is very unstable, and the error increases as we increase the order of input (p) for the test data! (This is not desired outcome).

However, with regularization (lambda >0), the regression is stable, and it is not effected by the order of the input (for both train and test data). Furthermore, when we take the optimal lambda, the error will be lowest. Hence, regularization is better.

D) Optimal value of p:

Problem 6:

A) Brief comparison for the problem 2 to 5.

method		reason
linear Regression	with intercept	gives better MSE than without intercept. Hence, better
	without intercept	gives worse MSE than without intercept. Hence, not to be used
Ridge Regression	formula based	same result, but needs high computation power if using large features.
	Gradient based	Same result, uses iterations to compute the results(Might be slower if less features are present). Can replace formula method, if large features are present.
	effect of lambda	helps in not over learning the parameters and stabalize the error when higher order inputs are present(high correlation). Hence its better to use regularization.
	order of input	with regularization, order of input has approx 0 effect on the results (and error). Hence doesn't affect for Ridge regression. However, without regularization, the high correlation will effect the output adversely.

Linear regression is used for features with no correlation (practically very little correlation, how little? is decided by: the computational advantage Vs error trade-off). Ridge Regression is used for data with correlation.

When we compare the results (using the calculated error, MSE) of both linear and Ridge regression. We find that the MSE calculated using linear regression with intercept is 3707.84018148. Whereas, MSE calculated using the ridge regression (lambda = 0.06) with intercept is 2851.33021344. Because we are using this regression or health care application, the accuracy of the results is of at most importance. Additionally, if input features are changed in future the correlation among the features are bound to change, and linear regression will have very poor performance if correlation increases. Hence, we choose the best possible option we have at hand: ridge regression.

Furthermore, while using the ridge regression, we recommend choosing formula based approach to find the weights, as numbers of features are not too high for modern day computers.