Introduction to Machine Learning
Programming Assignment 1

Aditya Sahay
Katyayni Shankar Kaushik
Shubham Sharma
Group No. 4

March 9, 2019

1 Experiment with Gaussian Discriminators

We implemented Linear Discriminant Analysis and Quadratic Discriminant Analysis and trained both the methods using the sample training data. We tested the methods using the test data provided, plotted the result in Figure 1

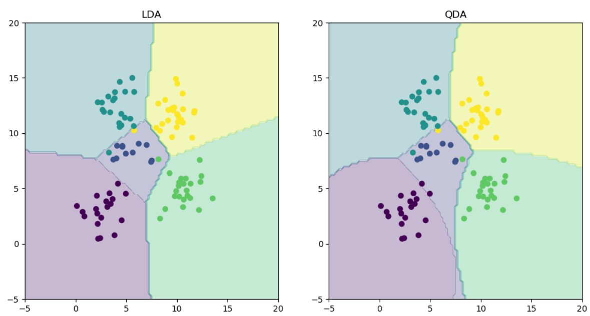

Figure 1: Plot of LDA and QDA on the test data

The accuracy of LDA is 97% whereas that of QDA is 94%. The reason for this is that QDA has non-linear boundaries and is therefore prone to overfitting and can lead to more mis-classifications when the data actually follow linear boundaries.

2 Experiment with Linear Regression

We performed Linear Regression on the given dataset, trained and tested the model with and without intercept term. Following are the Mean Squared Errors (MSE) for the various cases:

Train MSE without intercept: 19099.45
Train MSE with intercept: 2187.16
Test MSE without intercept: 106775.36
Test MSE with intercept: 3707.84

3 Experiment with Ridge Regression

We performed Ridge Regression on the given data and using the functions implemented in Problem2, calculated train and test errors for regularization parameter λ, ranging from 0 to 1 in steps of 0.01.

On comparing the magnitude of weights learnt using OLE and Ridge with minimum MSE, we observed that the weights for some of the features in Ridge were much less than that of OLE. This is because the regularization parameter penalizes those features that are less important, thus leading to reduction in their weights, hence improving the accuracy.

The plots for train and test MSEs can be seen in Figure 2. The minimum MSE for train data is 2187.16 and that of test data is 2851.33.

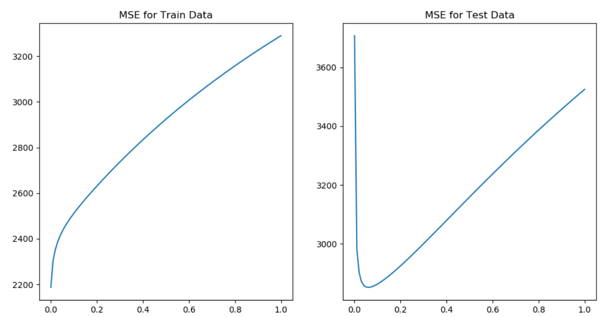

Figure 2: MSEs for Train and Test data over a range of λ in Ridge Regression

4 Using Gradient Descent for Ridge Regression Learning

We used Gradient descent to achieve Ridge Regression in order to avoid computing matrix inverse. As seen in Figure 3, the gradient descent performs almost equally well for the range of values of the regularization parameter λ.

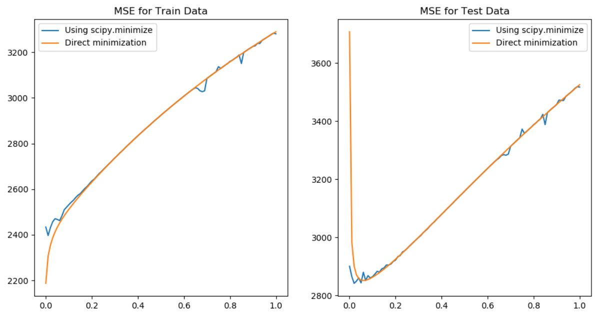

Figure 3: Train and Test MSEs for Gradient Descent

5 Non-linear Regression

For this problem, we used the third variable and formed a non-linear basis function. We used this function to train the model with regularization parameters 0 and 0.06 (the one that gave the lowest test MSE in Problem 3). We can see the corresponding plot for the Train and Test MSEs in Figure 4.

Although, the No-Regularization MSEs are quite similar for both the values of λ, the Regularized MSEs vary. The latter shows a downward trend in Train MSE and in case of Test MSE, we get the lowest error for $p = 1$, i.e. Linear Ridge Regression.

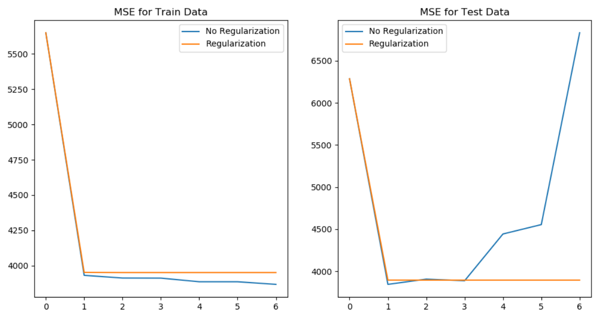

Figure 4: Train and Test MSEs for Non-Linear Regression

6 Interpreting Results

The comparison of MSEs for various methods can be seen in Figure 5.

Linear				Ridge	
Mean Square Error (Train)		Mean Square Error (Test)		MSE Train	MSE Test
With Intercept	Without Intercept	With Intercept	Without Intercept		
2187.160295	19099.44684	3707.840182	106775.3616	2187.160295	2851.330213
Gradient Descent		Non-Linear			
MSE Train	MSE Test	Mean Square Error (Train)		Mean Square Error (Test)	
		Lambda = 0	lambda = 1	Lambda = .06	lambda = .06
2396.708691	2841.57743	3866.883449	3950.682335	3845.03473	3895.582668

Figure 5: Comparison of MSE values for various methods

Thus, upon comparing the different MSE values of various methods, we observed that Gradient Descent showed the best result i.e. Train MSE = 2396.70 and Test MSE = 2841.57. This shows that variance in results of MSE values of Gradient Descent for Train and Test is the lowest in comparison to all the other methods.

Hence, we see that the Test MSE is the metric used for choosing the best setting.