# CSE 574 - Machine Learning

# Group – 77

Sathish Kumar Deivasigamani

Sankara Vadivel Dhandapani

Kaushik Raj Palanichamy

# Table of Contents

# Table of Figures

## Problem 1: Experiment with Gaussian discriminators

### Linear Discriminant Analysis (LDA)

In LDA we got 97% accuracy and the contour plot shows the classification boundaries. It's evident that LDA generates "simple boundaries" (approximately straight lines) and works on Occam's razor philosophy.

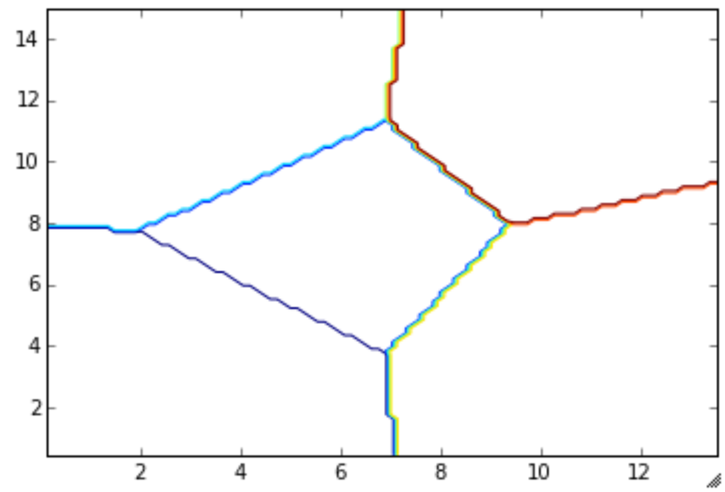


**Figure 1: Linear Discriminant Analysis**

### Quadratic Discriminant Analysis (QDA)

In QDA we got 94% accuracy. As shown in the below contour plot, the boundaries are not straight and are more complex than LDA. This resulted in overfitting.

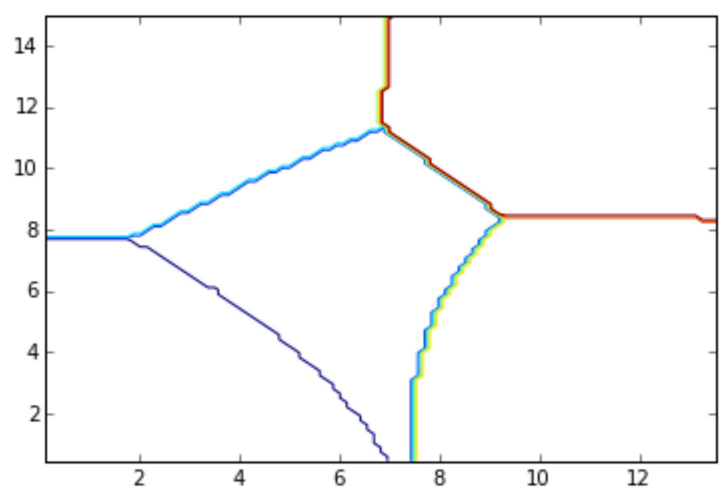


**Figure 2: Quadratic Discriminant Analysis**

## Comparison

The region enclosed in the circle in the below diagram is completely different from QDA. Intuitively, it looks like the prediction should be either left bottom or right bottom than the middle one as classified by the QDA due to overfitting of training data.

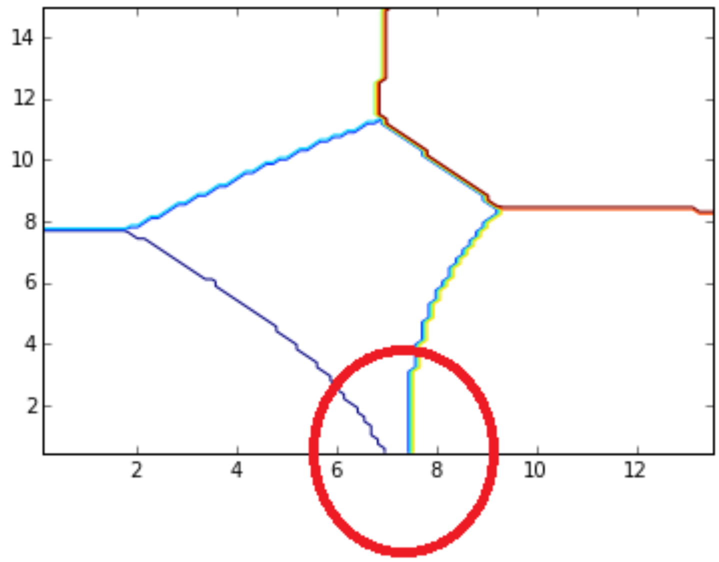


**Figure 3: LDA - QDA Comparison**

## Problem 2: Experiment with Linear Regression:

It is evident from the data that Root Mean Square Error is higher if we DO NOT use intercept. This is because we are forcing the classification line to pass through origin. This will not always yield best results, as the true label classification can be anywhere.

```
RMSE without intercept 23.10577434
RMSE with intercept     4.30571724
```

## Problem 3: Experiment with Ridge Regression

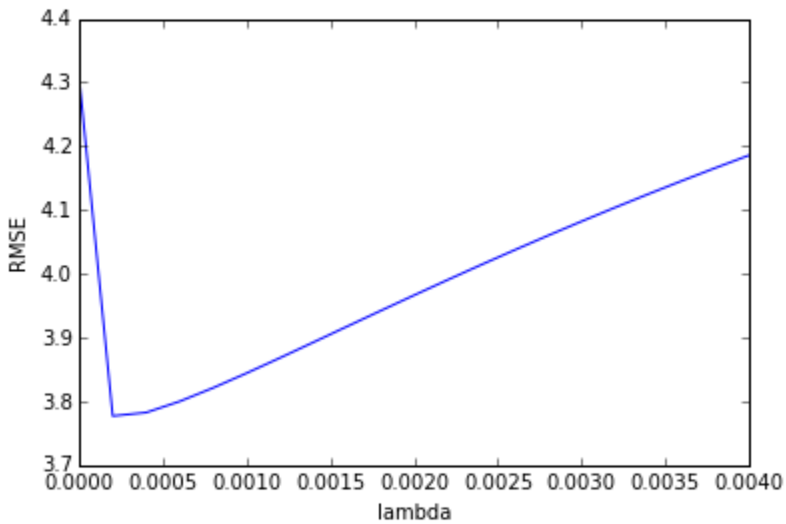


**Figure 4: Experiment with Ridge Regression**

```
Optimal value of lambda = 0.0002
```

If we don't use regularization, it's possible that the system may over fit the training data and would not perform well in the test data. If we use higher lambda for regularization it may result in under fitting i.e. not classifying data properly as it ignores slight variations in the input attributes. Choosing the right lambda is always a trade-off and there is no single answer. In this given test data, choosing lambda = 0.0002 results in the least RMSE. It may not be the case for a different test data.

## Problem 4: Using Gradient Descent for Ridge Regression Learning

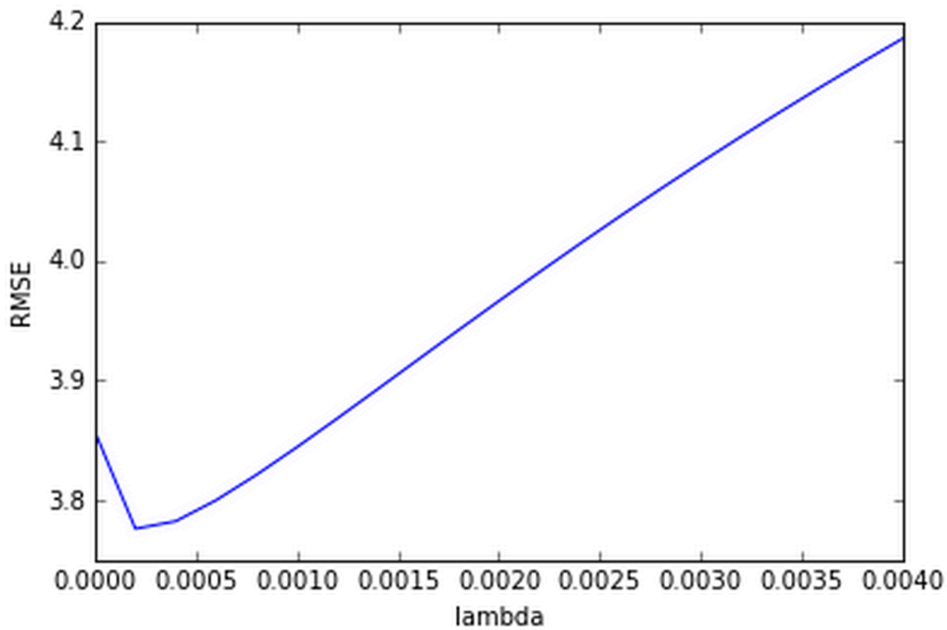


**Figure 5: Using Gradient Descent for Ridge Regression Learning**

```
Optimal value of lambda = 0.0002
```

The optimal value of lambda found using gradient descent algorithm is 0.0002. It is same as the optimal regularization parameter found using ridge regression learning.

## Problem 5: Non-linear Regression

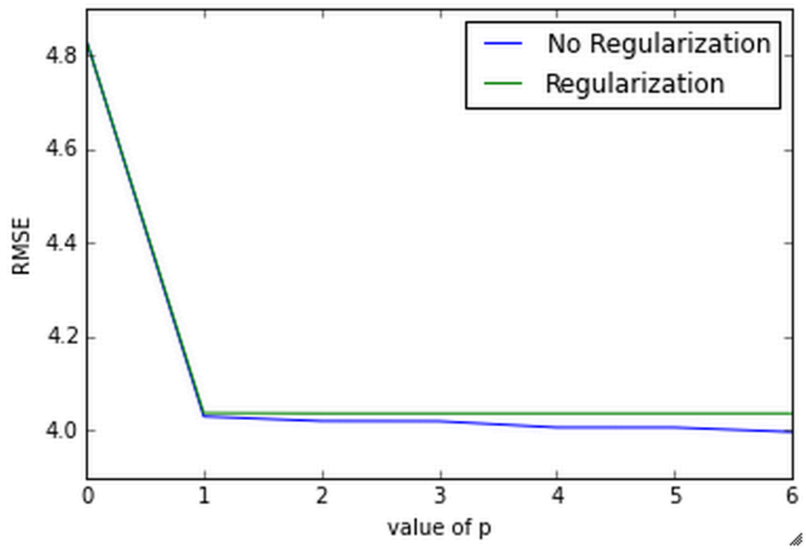


**Figure 6: Non-linear Regression (for Training Data)**

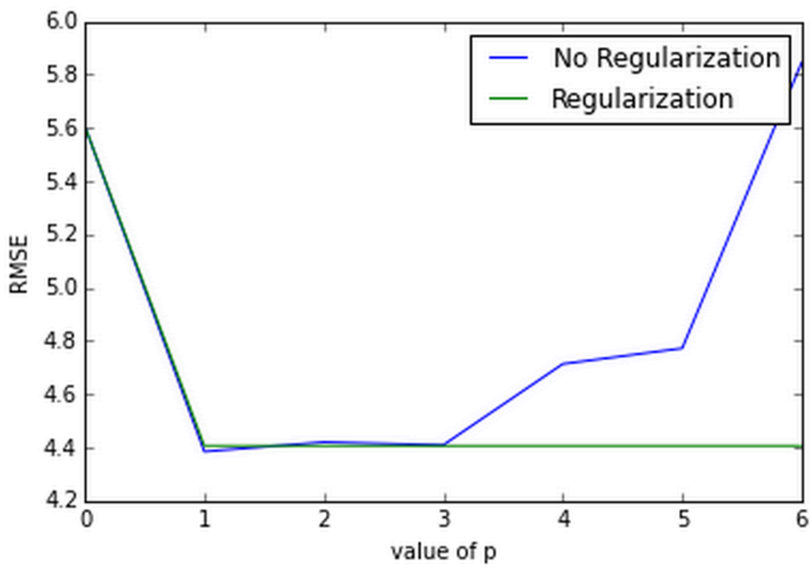


**Figure 7: Non-linear Regression (for Test Data)**

Nonlinear regression is an excellent example to demonstrate the over fitting problem using higher order polynomials. From the above figure, it is evident that when the degree of the polynomial increases, RMSE increases if lambda is kept zero. This is because, when the degree of the polynomial increases, the classification boundaries become so complex and defies Occam's razor. It performs well in training data but in test data it fails due to over-fitting.

## Problem 6: Interpreting Results

Repeatedly, the above results suggest that we should follow Occam Razor's philosophy, "simple solution is the best solution" while choosing the classification algorithm. By that, we should choose LDA or any other linear classifier since it gives the best results and does not suffer much from overfitting. We should also choose lambda experimentally as it depends completely upon the test data and its difference from the training data.

Linear classifiers are fast when compared to classifiers of higher degree polynomials. One special condition to choose higher order nonlinear classifiers is when training data and test data are almost same and overfitting is not an issue. Even in that case, regularization parameter should be approximately chosen to minimize the error.