

# Machine Learning Assignment 2

Ryan Cooper

April 20, 2017

2. With the way the data is arranged in the database, the LDA classifier reported an error rate of 2.5% for training and 0% for testing. This zero percent was very surprising to me, so I allowed the program to take an argument on the command line to randomize the order of the data before it is split, and this yielded results that were never greater than 7%, but average 3%, in both categories.

3. The QDA classifier reported an error rate of 1.66% for training and 0% for testing. Again, after shuffling, these values were observed to never exceed 7%, but averaged 3%, in both categories.

4. When the program is executed, it will display which categories are linearly separable for training and testing. Shuffling can often change which categories in the testing set are linearly separable, but because that data set is so small it doesn't mean very much. Iris-setosa is always reported as linearly separable, therefore it is the one that is truly linearly separable. Linear separability was determined by which categories had no errors when predicting which categories a data point would be in.

5. The program will display which variable is not useful for classification. It determined that the sepal width did not help classify the data. Usefulness was determined by the difference between the original data set error and the error when training without a feature in the data set.

6. When diagonal covariance ( $\Sigma$ ) matrices are used for LDA and QDA, LDA reports an error rate of 5% for training, and 0% for testing, while QDA reports 4.16% for training and 0% for testing. Again shuffling changes the testing error to a more acceptable value.

Github repo: <https://github.com/Pillager225/DiscriminativeAnalysis.git>