**Assignment 2 – COMP 551**

**Olivier Simard-Morissette**

**260563480**

**February 11th, 2018**

2.

*LDA DS1 Results*

Accuracy: 0.9625

Precision: 0.9646

Recall: 0.9630

F1 Measure: 0.9638

The class zero and class one mean vectors can be found in the zip file 260563480.zip labeled as DS1-Class-X-Mean-Vector and the covariance matrix may be found in DS1-covariance-matrix.

3.

*KNN DS1 Results*

k was tested from 1 to 20 just like in the lecture slides for class.

Do you do better or worse than LDA?

The performance of KNN across all evaluation measurements was worse than linear discriminant analysis. LDA performs well since the points for class 0 and class 1 share the same covariance matrix, which is an assumption of the LDA classifier.

Are there particular values of k which perform better?

If a value of k is even, it has a much lower recall than a value of k which is odd. This may be due to the KNN implementation, which in the event of a tie, will always choose the negative class 0. This default behaviour should be changed to randomly choose between the two classes in the event of a tie instead of always choosing class 0.

Report the best fit accuracy, precision, recall and f-measure achieved by this classifier.

The best fit was determined by looking at the Accuracy from K = 1 to 20. This was determined to be when k=13. The accuracy, precision, recall and F1 were:

k = 13

Accuracy: 0.5508

Precision: 0.5387

Recall: 0.6233

F1 Measure: 0.5779

5.

*LDA Results:*

Accuracy: 0.5375

Precision: 0.5352

Recall: 0.5343

F1 Measure: 0.5348

*KNN Results:*

The best fit was determined by looking at the Accuracy from K = 1 to 20. This was determined to be when k=15. The accuracy, precision, recall and F1 were:

k = 15

Accuracy: 0.5592

Precision: 0.5612

Recall: 0.5226

F1 Measure: 0.5412

*Analysis:*

The performance of the LDA is reduced dramatically. This can be explained by LDA's assumption that all sample classes have the same covariance matrices. In this case, the samples were generated using separate covariance matrices and we got a poor performance as a result.

The performance of the KNN did not improve from DS1 to DS2. The performance of KNN in this case was better than LDA when an odd number of K was used.