COMP 551: Assignment 2

Anand Kamat
ID# 260773313

Question1:

The dataset DS1 has been provided with the source code.

Question2:

The parameters calculated are provided in the folder.

LDA Model:

Accuracy	**0.956666666667**
Precision	**0.953642384106**
Recall	**0.96**
F- Score	**0.956810631229**

Quesion3:

KNN Model:

	K = 1	K = 2	K = 3	K = 4	K = 5
Accuracy	0.535	0.52333333	0.52333333	0.5175	0.545
Precision	0.53523489	0.54516129	0.52348993	0.52741514	0.54530201
Recall	0.53166666	0.28166666	0.52	0.33666666	0.54166666
F-Score	0.53344481	0.371428571	0.52173913	0.41098677	0.54347826

The KNN classifier fails to perform as well as the LDA model. The best performance achieved by the KNN classifier (in terms of F-Score) is recorded when K = 5.

Question 4

The dataset DS1 has been provided with the source code

Question 5:

The parameters calculated are provided in the folder.

LDA Model:

Accuracy	**0.438333333333**
Precision	**0.439739413681**
Recall	**0.45**
F- Score	**0.444810543657**

KNN Model:

	K = 1	K = 2	K = 3	K = 4	K = 5
Accuracy	0.64416666	0.60833333	0.67	0.6475	0.66166666
Precision	0.67474747	0.75390625	0.70318725	0.76261127	0.69877049
Recall	0.55666666	0.32166666	0.58833333	0.42833333	0.56833333
F-Score	0.61004	0.45093457	0.64065335	0.54855923	0.62683823

We now observe that the KNN classifier performs significantly better, however the success of the KNN classifier can be attributed to the low performance stats of the LDA model. The LDA model shows poor performance compared to the performance it showed in Question 2. The performance of the KNN classifier has slightly increased compared to its stats in Question 3.

Question 6:

The KNN shows similar performance for both DS1 and DS2 prediction however there is a slight increase in performance in the later case. This shows the KNN wasn't significantly affected by the dataset containing mixture of gaussians. In the case of the LDA model, we see it performing extremely well on DS1 with excellent accuracy and F-score. However, LDA fails miserably when it came to classifying data from a mixture of gaussians. This may be due to the fact that the mixture

of gaussians in the dataset required a non-linear classifier (like KNN) and confused a simple classifier such as the LDA.