COMP551 Assignment 2

Peter Park

260571481

*note all the data in this report and more are included in the jupyter notebook file. Please check the jupyter notebook file for detail.

1 QUESTION 1

Simulated data set is in the jupyter notebook file.

2 QUESTION 2

Below is the result for one of the simulated results. The first tuple shows prediction value for positive, (True positive, False Positive), and the second tuple shows prediction value for negative, (True negative, False negative).

```
(574, 26)
(29, 571)
The accuracy of this classifier is 95.41666666666667%
The precision of this classifier is 95.66666666666667%
The recall of this classifier is 95.19071310116087%
The F-measure of this classifier is 0.9542809642560266
```

3 QUESTION 3

$k = 1$, usually performs better than other values of k, below is the result on the performance.

```
[(288, 312), (461, 139), (324, 276), (435, 165), (324, 276), (405, 195), (306, 294), (394, 206), (306, 294), (383
7), (313, 287), (385, 215)]
[(310, 290), (450, 150), (313, 287), (406, 194), (308, 292), (396, 204), (320, 280), (381, 219), (305, 295), (391
9), (326, 274), (374, 226)]
value of k that performs the best is  2
The accuracy of this classifier is 50.916666666666664%
The precision of this classifier is 76.83333333333333%
The recall of this classifier is 50.603732162458844%
The F-measure of this classifier is 0.6101919258769027
```

4 QUESTION 4

Simulated data set is in the jupyter notebook file.

5 QUESTION 5

We get the following result for performance

```
(212, 388)
(217, 383)
The accuracy of this classifier is 49.583333333333336%
The precision of this classifier is 35.333333333333336%
The recall of this classifier is 49.4172494172494915%
The F-measure of this classifier is 0.8338192419825073
```

```
value of k that performs the best is  2
The accuracy of this classifier is 52.33333333333333%
The precision of this classifier is 78.16666666666666%
The recall of this classifier is 51.53846153846153%
The F-measure of this classifier is 0.6211920529801324
```

6 QUESTION 6

We can see that there is negligible difference in performance for $k-$NN classifier where as there is huge difference in performance for LDA model.For example, The data shows that LDA model for question 5 has much worse performance than question2.