# Question 1:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class | Mean(f1) | Std(f1) | Mean(f2) | Std(f2) | Mean(f3) | Std(f3) | Mean(f4) | Std(f4) |
| 0 | 2.27 | 2.02 | 4.25 | 5.14 | 0.8 | 3.24 | -1.15 | 2.12 |
| 1 | -1.87 | 1.88 | -0.99 | 5.4 | 2.15 | 5.26 | -1.25 | 2.07 |
| all | 0.43 | 2.84 | 1.92 | 5.87 | 1.4 | 4.31 | -1.19 | 2.1 |

For mean value of f1 and f2, class 0 is greater than class 1.

The standard deviations for f2 and f3 are pretty large.

F4’s mean and standard deviations are both quite similar between class 1 and 0.

# Question 2:

The simple predictor is in the helper.py

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TP | FP | TN | FN | Accuracy | TPR | TNR |
| 28 | 0 | 308 | 349 | 0.4905 | 0.074 | 1.0 |

It gives me better accuracy identifying fake bills.

Overall, it’s worth than a coin flip (49.05%<50%).

# Question 3:

A blue line on a white background

Description automatically generated

The best Ks is 5 and 7, both at 1.0 accuracy. We are going to use k\*=5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TP | FP | TN | FN | Accuracy | TPR | TNR |
| 384 | 0 | 302 | 0 | 1.0 (682/686) | 1 | 1 |

My kNN is way better than the simple predictor. It’s worse in TNR but better in all other aspects.

Result from simple predict: green

Result from knn when k=5: green

# Question 4:

For dropping f 1 Accuracy is: 0.9533527696793003

For dropping f 2 Accuracy is: 0.9795918367346939

For dropping f 3 Accuracy is: 0.9737609329446064

For dropping f 4 Accuracy is: 0.9941690962099126

The accuracy from 3 is 1.0 (682 correct out of 686). With dropping features, the accuracy did not increase.

When removing f1, we have the greatest loss of accuracy.

When removing f4, we have the least loss of accuracy.

# Question 5:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TP | FP | TN | FN | Accuracy | TPR | TNR |
| 372 | 2 | 300 | 12 | 0.9796 | 0.96875 | 0.9934 |

It’s better than the simple classification method.

It does not do better than the kNN model (since that model has 1.0 accuracy).

Predicted based on BUID is green. It is the same as predicted by the kNN model.

# Question 6:

Accuracy for removing f 1 is: 0.8177842565597667

Accuracy for removing f 2 is: 0.9052478134110787

Accuracy for removing f 3 is: 0.8790087463556852

Accuracy for removing f 4 is: 0.9810495626822158

The accuracy increased after removing f4.

Removing f1 contributes to the most loss of accuracy.

Removing f4 (since it’s gaining accuracy, otherwise it’s f2) contributes to the lest loss of accuracy.

The relative significance is indeed the same as I got for kNN.