

CS677 HW 3
By: Ahnaf Tajwar

Question 1

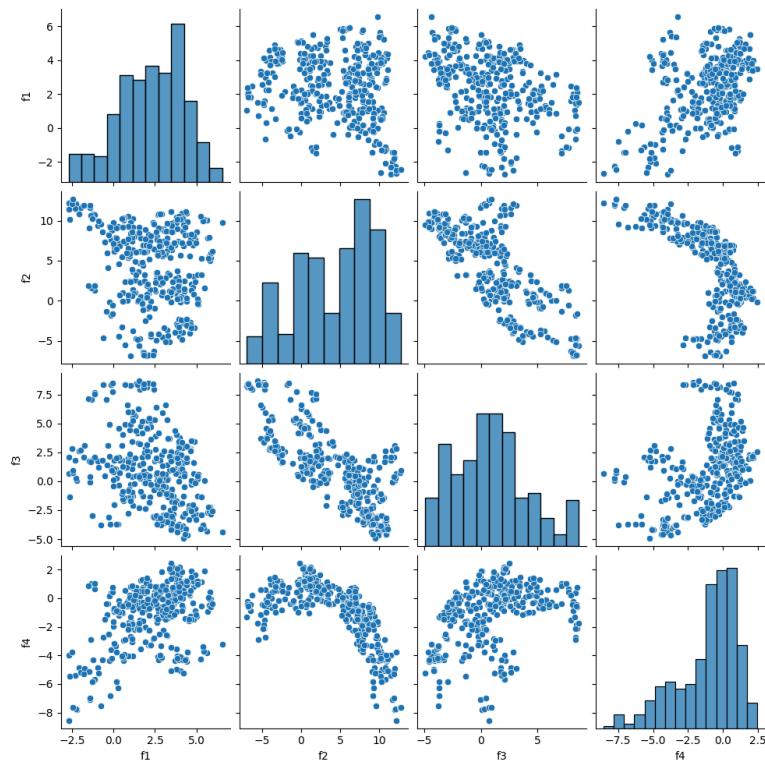
1. See code
- 2.

class	f1_mean	f1_std	f2_mean	f2_std	f3_mean	f3_std	f4_mean	f4_std
0	2.28	2.02	4.26	5.14	0.8	3.24	-1.15	2.13
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

3. For class 0 banknotes, it seems like there is a much higher mean value for the variance and skewness. Whereas, for the class 1 banknotes, there is a much higher mean value for the kurtosis. Entropy is similar for both classes.

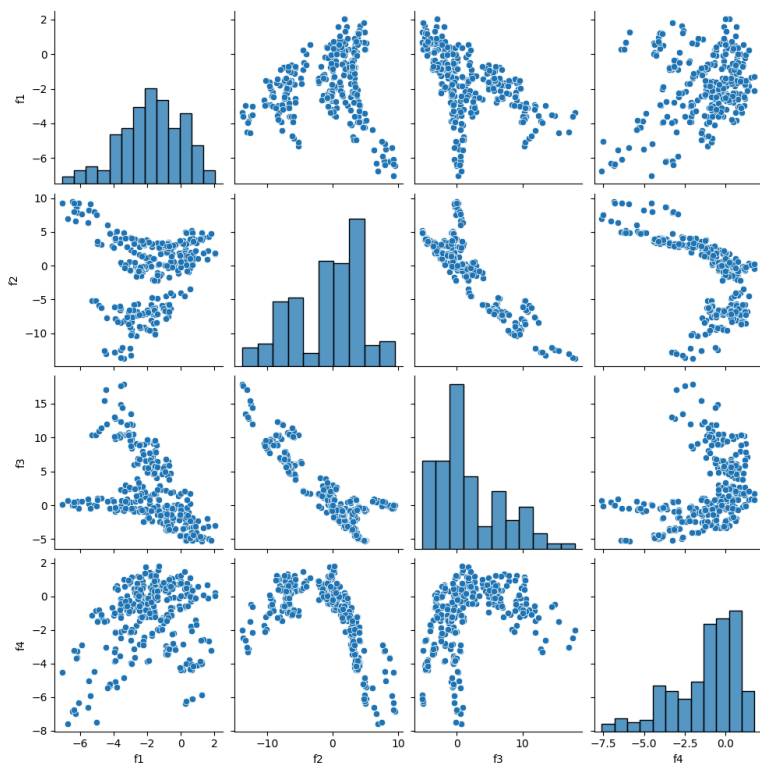
Question 2

1. See code
 - a. Good Bills



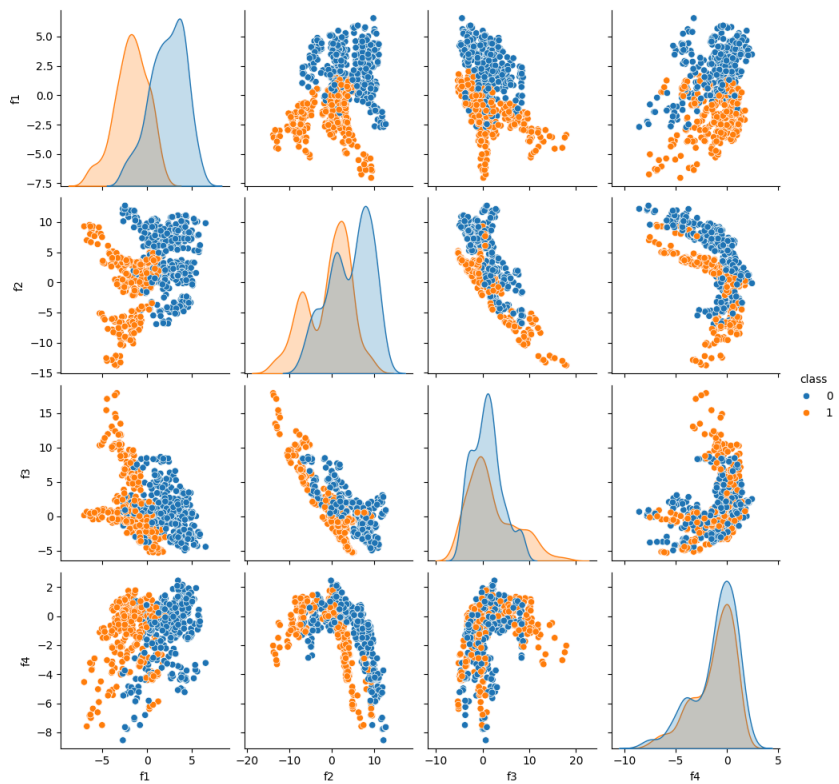
i.

b. Fake Bills



i.

c. All Bills



i.

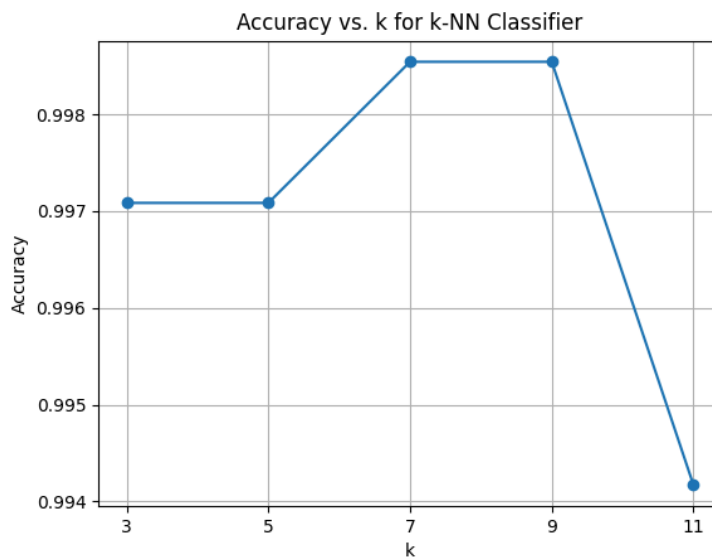
2. Visual examination and comparisons
 - a. From the pairplot with All Bills, it seems like feature 4 does not have much distinct difference between good or fake bills. For feature 1, a higher number usually indicates a good bill. For features 2 and 3, a higher number may sometimes indicate a good bill.
 - b. If $f1 > 1$ and $f2 > 2$ and $f3 < 2$, then bill is **good**
3. See code
4. See code
- 5.

TP	FP	TN	FN	accuracy	TPR	TNR
165	220	289	12	0.6618	0.9322	0.5678

6. My simple classifier gives me a higher accuracy in giving “fake” bills since the TNR is higher. My overall accuracy is higher than a coin flip since it is more than 0.5.

Question 3

1. See code
 - a. Accuracy for $k=3$: 0.9971
 - b. Accuracy for $k=5$: 0.9971
 - c. Accuracy for $k=7$: 0.9985
 - d. Accuracy for $k=9$: 0.9985
 - e. Accuracy for $k=11$: 0.9942



2.
 - a. The optimal value k^* of k is 7 and 9. I will use $k^* = 7$ for the following questions.
- 3.

TP	FP	TN	FN	accuracy	TPR	TNR
384	1	301	0	0.9985	1.0	0.9967

4. The kNN classifier is better than the simple classifier in all measures. It has a higher TP, TN, accuracy, TPR, and TNR.
5. $F1 = 8$, $F2 = 2$, $F3 = 6$, $F4 = 0$
 - a. Simple Classifier Prediction
 - i. 1 or “fake” bill
 - b. kNN Prediction
 - i. 0 or “good” bill

Question 4

1. Accuracy after dropping each f
 - a. F1 drop
 - i. Accuracy = 0.9548
 - b. F2 drop
 - i. Accuracy = 0.9708
 - c. F3 drop
 - i. Accuracy = 0.9679
 - d. F4 drop
 - i. Accuracy = 0.9927
2. The accuracy did not increase in any of the cases where one feature is dropped.
3. Dropping f1 contributed the most to loss of accuracy.
4. Dropping f4 came close to the accuracy of including all of them. This means that f4 contributes the least to loss of accuracy.

Question 5

1. Accuracy using logistic regression is 0.9898
- 2.

TP	FP	TN	FN	accuracy	TPR	TNR
381	4	298	3	0.9898	0.9922	0.9868

3. The logistic regression is better than the simple classifier in all measures. It has a higher TP, TN, accuracy, TPR, and TNR.
4. The logistic regression is not better than the kNN classifier in any measures. It has a lower accuracy, TPR, and TNR than kNN.

5. The class label predicted for BUID by logistic regression is 0. This is the same as the kNN classifier prediction.

Question 6

1. Accuracy after dropping each f
 - a. F1 drop
 - i. Accuracy = 0.8003
 - b. F2 drop
 - i. Accuracy = 0.8965
 - c. F3 drop
 - i. Accuracy = 0.8819
 - d. F4 drop
 - i. Accuracy = 0.9898
2. The accuracy did not increase in any of the cases where one feature is dropped.
3. Dropping f1 contributed the most to loss of accuracy.
4. Dropping 41 contributed the least to loss of accuracy. The accuracy is the same as not dropping any features.
5. The relative significance of dropping each feature is similar to that of kNN. The corresponding effects of each feature is the same. However, for logistic regression, dropping the features f1, f2, or f3 result in a larger loss of accuracy than kNN.