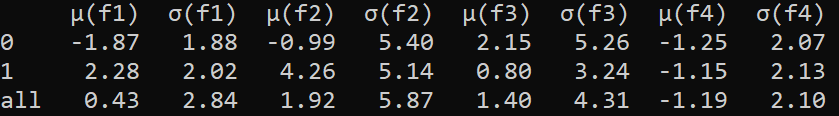
Jake Stephens

Assignment #3

July 27, 2021

**Question #1:**

1. Done



1. I notice in the above table that for the false bank notes (label 1) for f1 and f2, the means are positive, but for real bank notes (label 0) the means for f1 and f2 are negative. After noticing this, I feel like I could make a pretty good decision on whether a bank note is false or not according to the values of f1 and f2 alone.

**Question #2:**

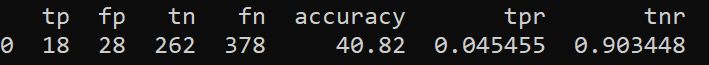
1. Done – See fake\_bills.pdf and good\_bills.pdf in the output folder.
2. Here are my comparison rules:

Rule 1: f1 > -4

Rule 2: f2 > -10

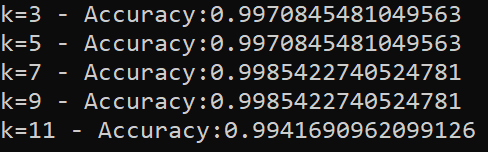
Rule 3: -4 > f3 > -6

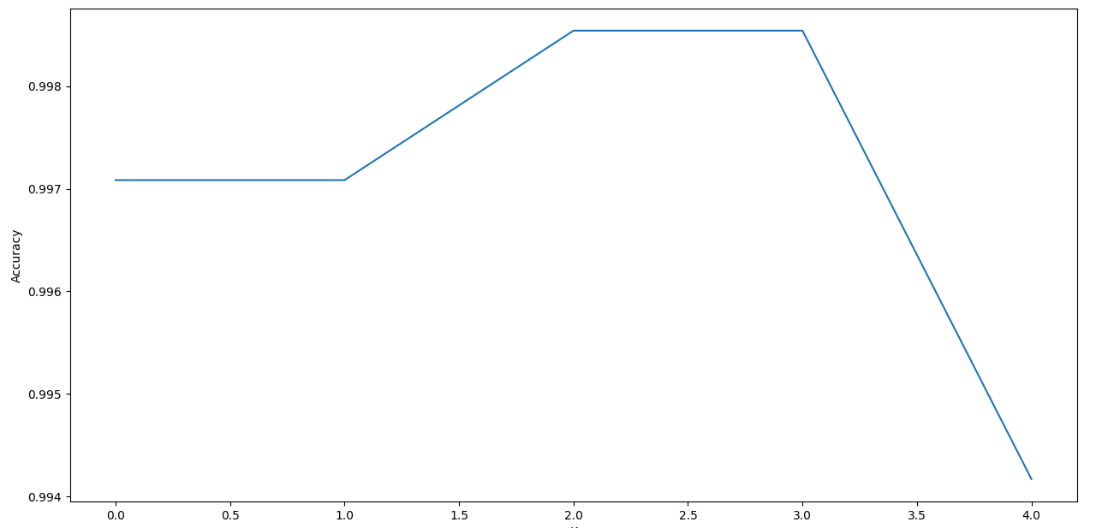
1. Done
2. Done



1. My simple classifier does not perform well, being that the accuracy is 40.82%. You would be better off flipping a coin to decide whether a bank note was fake or not. This was surprising to me because I felt like I had some good rules to classify bank notes, these results show me how hard it can be to come up with your own classification rules.

**Question #3:**





I found the optimal value of k to be 7 or 9. When k=7 the accuracy is the same as when k=9.



1. Yes the knn classifier is much better than my simple classifier I wrote myself for question 2. I believe knn is so good at predicting the label because the data is clumped together, and it is easy to tell if you look at the data distributed, 2 blobs, 1 blob being the real bank notes and another blob for false bank notes that is distinctively separated from the other blob.

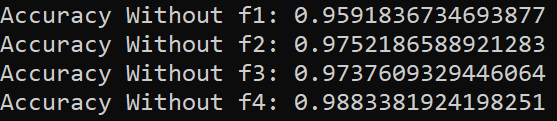
The predicted value for my BU ID Bill using my simple classifier is: 1

My Bill is Fake...

The predicted value for my BU ID Bill using Knn where k=7 is: [0]

My Bill is Real!!!!

**Question #4:**



1. No, removing any of the features did not improve the accuracy of the model. However, I will note that there is a significant difference in the accuracy between removing f1 and removing f4.
2. Removing f1 had the most loss of accuracy.
3. Removing f4 had the least loss of accuracy.

**Question #5:**

1. Done

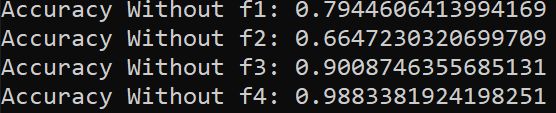


1. Yes, linear regression is much better at predicting the label than my simple classifier.
2. No, the knn classifier is the better model. Linear regression works pretty good but not better than knn for any of the fields (tp, tn, accuracy, tpr, tnr)
3. The predicted value for my BU ID Bill is: [0]

My Bill is Real!!!!

This is the same label predicted by knn, which makes sense as both algorithms (knn and logistic regression) are very accurate models for this dataset.

**Question #6:**



1. No, none of the accuracy’s increased after removing any of the fields. However, removing the accuracy after removing the f4 field is very close to the accuracy when all fields are considered.
2. Removing f2 had the most loss of accuracy.
3. Removing f4 had the least loss of accuracy.
4. Both algorithms (knn and logistic regression) when removing the field f4 from the dataset, have the least loss of accuracy, which is interesting.