Jon Organ

CS677

Homework 7, Bank

Good bills:



Fake bills:



**Question 1.1:**

CSV file read and column added for color...

**Question 1.2:**

Saving Q1.2 Table...

A table with numbers and symbols

Description automatically generated

**Question 1.3:**

It seems that generally if a bank note has a positive value for F1 (variance), then it's green. If the F2 (skewness) value is greater than 2, it seems to be green. If the F3 (curtosis) value is greater than 1.4, its green. Finally if the F4 (entropy) value is greater than -1.19, then its green. The volatility doesn't seem to show any significant patterns

**Question 2.1:**

Pair plots generated...

**Question 2.2:**

Rules for comparison:

if F1 > 0 and F2 > 1.9 and F3 > 1.4 then its good, else fake

**Question 2.3:**

Simple classifier labels predicted...

**Question 2.4:**

Simple classifier labels evaluators computed...

**Question 2.5:**

Simple classifier table:

TP FP TN FN accuracy TPR TNR

22 0 343 321 0.53 0.06 1.0

**Question 2.6:**

My simple classifier is perfect at identifying fake bills, and very bad at identifying real bills. The overall accuracy is just barely better than flipping a coin (53%)

**Question 3.1:**

kNN accuracy computed...

**Question 3.2:**

Saving k Accuracy by Value graph...

A graph with a line

Description automatically generated

**Question 3.3:**

kNN classifier table:

TP FP TN FN accuracy TPR TNR

385 0 301 0 1.0 1.0 1.0

**Question 3.4:**

My kNN classifier was much more accurate in every way compared to my simple classifier

**Question 3.5:**

Simple classifier prediction for BUID: Green

kNN classifier prediction for BUID: Green

**Question 4.1:**

kNN minus 1 feature accuracy computed for each feature

**Question 4.2:**

kNN minus F1 accuracy: 96.21%

kNN minus F2 accuracy: 97.23%

kNN minus F3 accuracy: 96.36%

kNN minus F4 accuracy: 98.4%

None of these were more accurate than all 4 working together

**Question 4.3:**

Removing feature F1 (variance) caused the greatest accuracy loss

**Question 4.4:**

Removing feature F4 (entropy) caused the least accuracy loss

**Question 5.1:**

Logistic regression classifier accuracy: 0.99%

**Question 5.2:**

Logistic regression classifier table:

TP FP TN FN accuracy TPR TNR

377 1 300 8 0.99 0.98 1.0

**Question 5.3:**

My logistic regression classifier was much better than my simple classifier in its accuracy and TPR, they both had the same TNR however

**Question 5.4:**

My logistic regression classifier and my kNN classifier are comparable in terms of their performance. However, the kNN classifier was slightly more accurate as it had no false positives or negatives, whereas the logistic regression had 1 false positive, and 8 false negatives.

**Question 5.5:**

Logistic regression classifier prediction for BUID: ['Green']

This is the same label as predicted by kNN

**Question 6.1:**

Logistic regression classifier minus 1 feature accuracy computed for each feature

**Question 6.2:**

LRC minus F1 accuracy: 81.92%

LRC minus F2 accuracy: 89.36%

LRC minus F3 accuracy: 87.17%

LRC minus F4 accuracy: 98.54%

None of these were more accurate than all 4 working together

**Question 6.3:**

For LRC, removing the F1 feature (variance), caused the greatest accuracy loss

**Question 6.4:**

For LRC, removing the F4 feature (entropy), caused the least accuracy loss

**Question 6.5:**

Comparing kNN with LRC after removing each feature one by one shows LRC had a much greater loss of accuracy by removing features F1, F2, and F3 when compared with kNN. The relative significance of each feature is comparable otherwise. For both, removing F4 had little effect