CS677A3Fall2023_WillMohr_Banknotes

Q1

	variance		skewness		curtosis		entropy	
	mean	std	mean	std	mean	std	mean	std
0	2.28	2.02	4.26	5.14	0.80	3.24	–1.15	2.13
1	–1.87	1.88	–0.99	5.40	2.15	5.26	–1.25	2.07
all	0.43	2.84	1.92	5.87	1.40	4.31	–1.19	2.10

fake and real banknotes have similar mean and std for entropy. For kurtosis, the real notes have a lower mean and std, but not totally differentially so considering the relatively high std for both. Regarding skewness, the mean is about a standard deviation higher for the real notes, with real and fake notes having similar std. Variance is the most clear-cut, with real notes having a higher mean that is over two standard deviations above the fake notes. Again, both real and fake notes have similar std.

Q2

CM_Result	TN	TP	FN	FP	TPR	TNR	accuracy
Value	334.0	221.0	84.0	47.0	0.72	0.88	0.81

The accuracy is much better than p=.5 random coin flipping in terms of overall accuracy but also TPR and TNR. True Negative rate is still better than true positive rate.

Q3

	TP	FP	TN	FN	accuracy	TPR	TNR
Metric Value	305.0	0.0	381.0	0.0	1.0	1.0	1.0

K-nn is perfect and better than my simple classifier.

Both my simple classifier and the knn classifier predict a 'Good' or class 0 banknote

Q4

	TP	FP	TN	FN	accuracy	TPR	TNR
variance	279.0	26.0	361.0	20.0	0.932945	0.933110	0.932817
skewness	296.0	9.0	366.0	15.0	0.965015	0.951768	0.976000
curtosis	290.0	15.0	374.0	7.0	0.967930	0.976431	0.961440
entropy	302.0	3.0	380.0	1.0	0.994169	0.996700	0.992167
Full_Model	305.0	0.0	380.0	1.0	0.998542	0.996732	1.000000

Knn feature significance ordering: variance > skewness > curtosis > entropy. The full model was the best

Q5

	TP	FP	TN	FN	accuracy	TPR	TNR
LogReg Metrics	305.0	0.0	375.0	6.0	0.991254	0.980707	1.0

Better than simple, worse than best k-NN.

Logistic regression also predicts a "Good" bill based on the last 4 digits of my BUID.

Q6.

	TP	FP	TN	FN	accuracy	TPR	TNR
variance	253.0	52.0	300.0	81.0	0.806122	0.757485	0.852273
skewness	267.0	38.0	354.0	27.0	0.905248	0.908163	0.903061
curtosis	246.0	59.0	355.0	26.0	0.876093	0.904412	0.857488
entropy	305.0	0.0	375.0	6.0	0.991254	0.980707	1.000000
Full_Model	305.0	0.0	375.0	6.0	0.991254	0.980707	1.000000

Knn feature significance ordering: variance > skewness > curtosis > entropy
LogReg feature significance ordering: variance > curtosis > skewness > entropy. Skewness, curtsies are switched. Removing entropy had no effect