CSCA 5622 Supervised Learning Final Project

Bradley Leavitt

Banknote Authentication

Lohweg, V. (2012). Banknote Authentication [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C55P57.

https://github.com/brle1242/CSCA-5622-Supervised-Learning-Final-Project

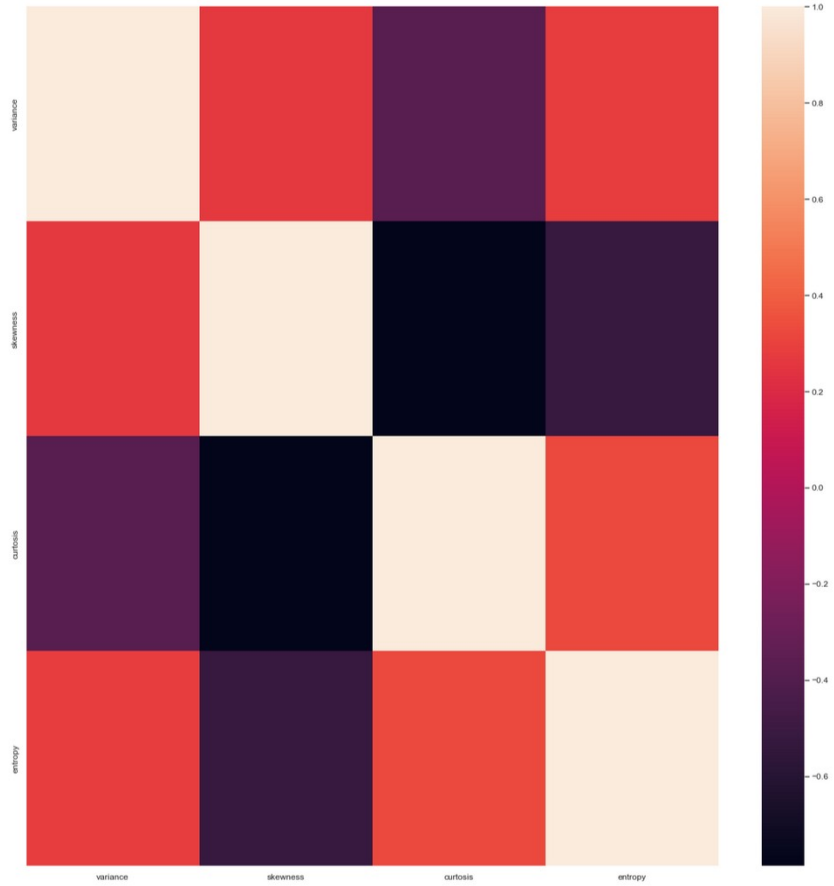
Data was extracted from images that were taken from genuine and forged banknote-like specimens.

Banknote Authentication

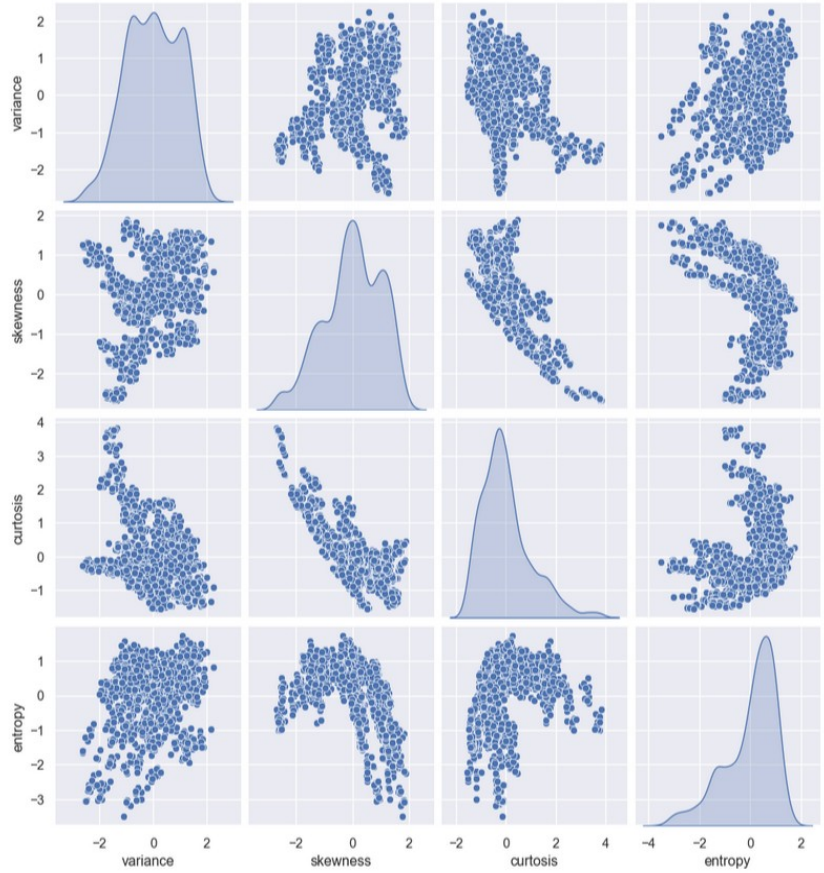
This data was precleaned, but I did explorativ analysis:

- correlation matrix
- pair plot

Correlation Matrix



Pair Plot

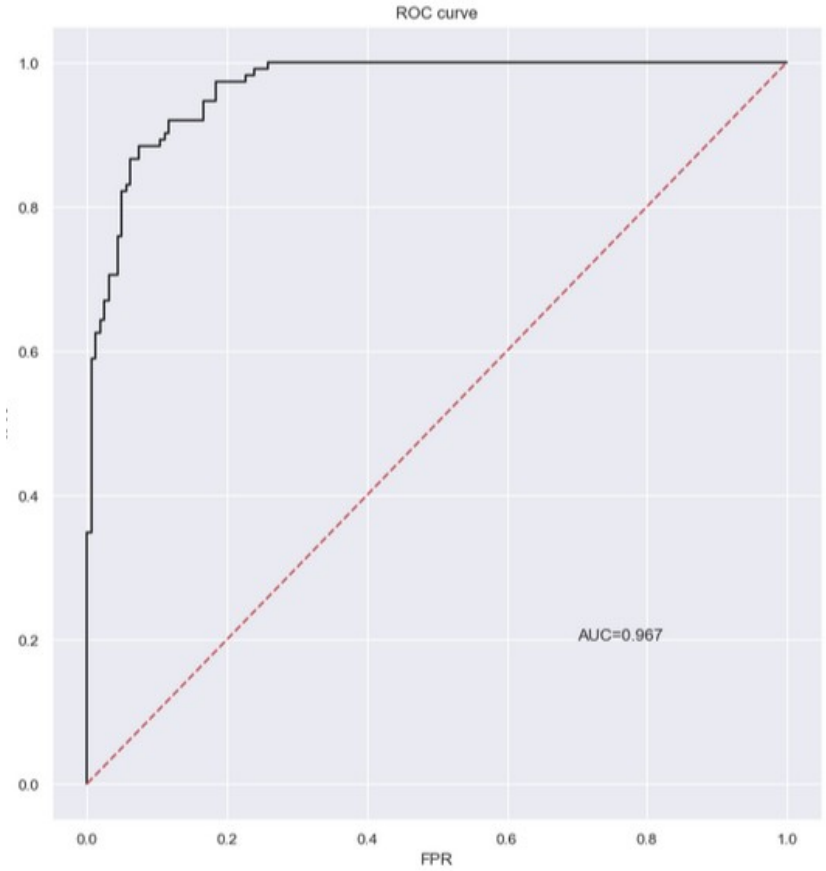

variance
skewness
curtosis
entropy
variance
skewness
curtosis
entropy

Banknote Authentication

First Attempt: Logistic Regression, without Skewness

- ~89.1% accurate

ROC Curve (Without Skewness)

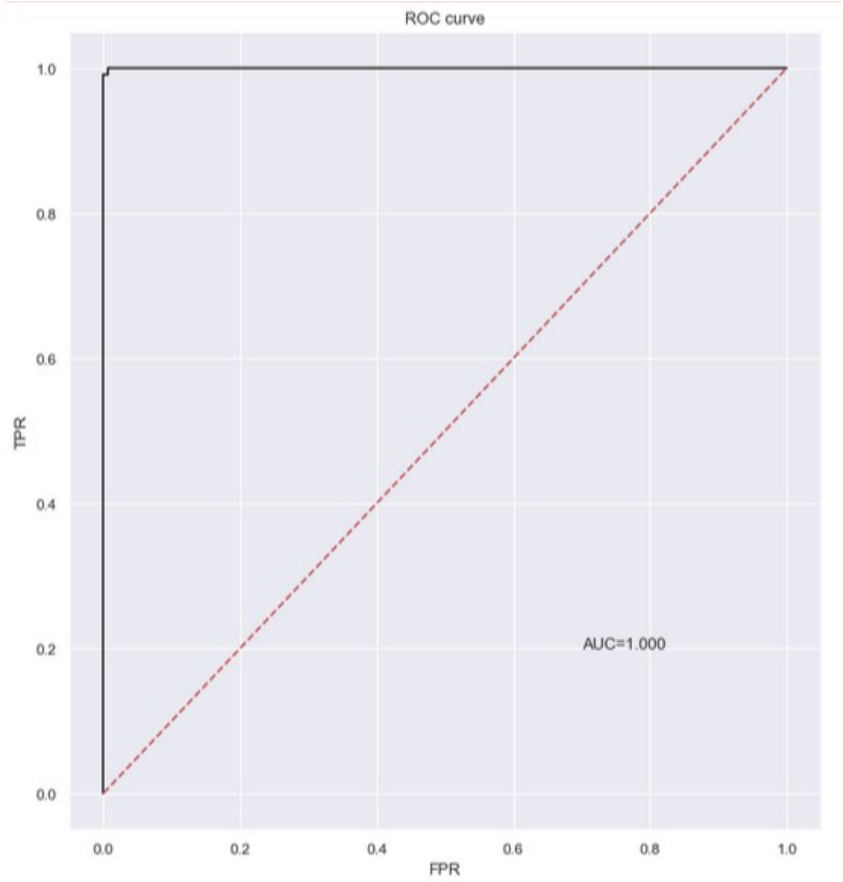


Banknote Authentication

Second Attempt: Logistic Regression, with Skewness

- ~98.54% accurate

ROC Curve (With Skewness)

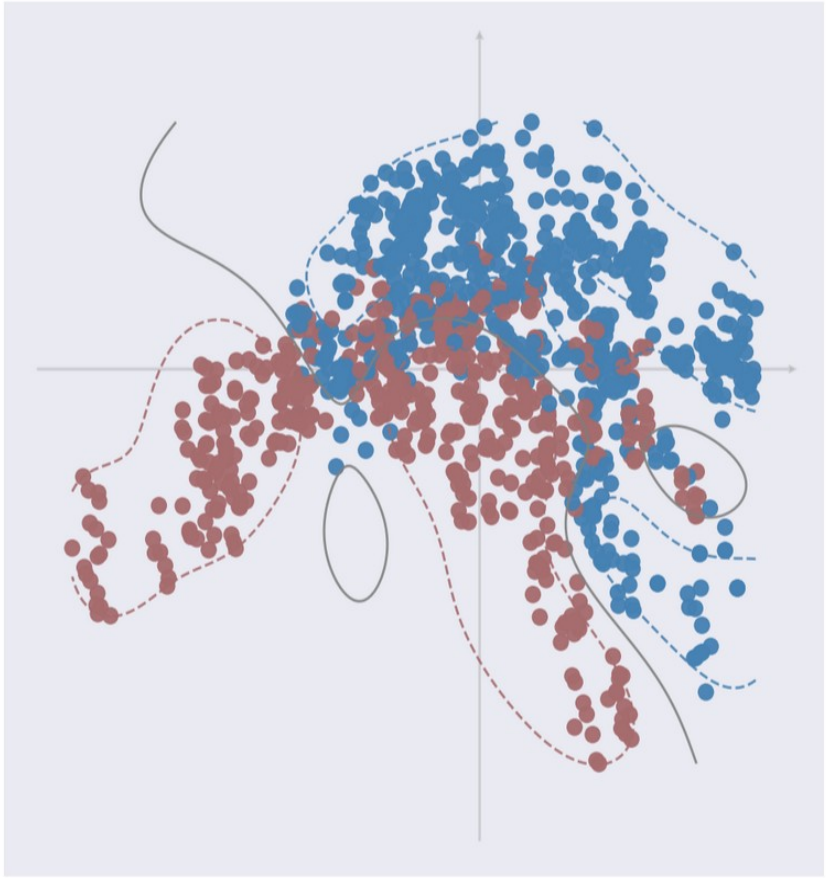


Banknote Authentication

Can we do better?

- Naive SVM accuracy 99.63%

SVM C=1 Gamma=1

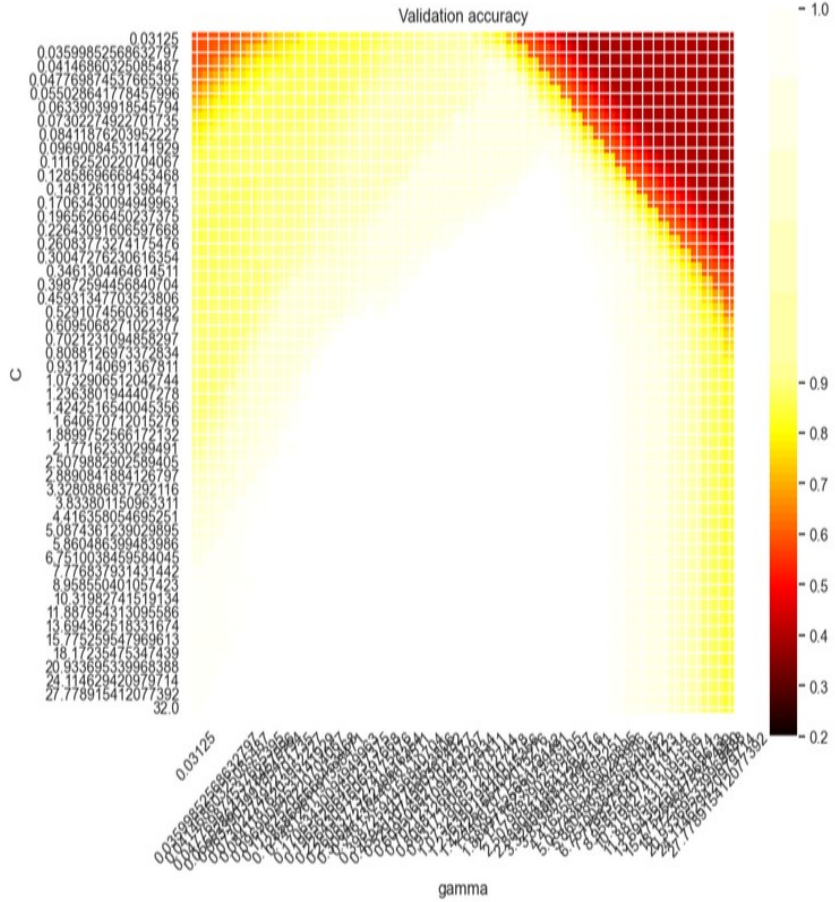


Banknote Authentication

Can we do better?

- perform a grid search on SVM parameters
- generate a heat map

SVM Grid Search Heatmap



Banknote Authentication

Grid Search Results

- SVM with 100% accuracy on training and test data

SVM

Banknote Authentication

Should we keep looking?

- naive KNN scores 99.63% accurate on test data
- grid search on KNN reveals a model with 100% accuracy on training data and 99.63% accuracy on test data

Banknote Authentication

Reality of bank note authentication

- Complex business process
- False positives and negatives both cost time and money

Banknote Authentication

What model to use?

- All models performed well
- Keep it simple!
- Grid search KNN is simplest model with high accuracy