mnavane@g.clemson.edu

To Evaluate the Authenticity of Currency Notes

Introduction

Although producing currency notes is strictly monitored by the government, the cases of counterfeit money production and circulation pops up in the news worldwide every now and then. Counterfeit money is currency that is produced without the legal sanction of the state or government and in deliberate violation of that country's laws[1].

In this project, an attempt has been made to detect such currencies by using the Wavelet Transformed images of the currency notes (both genuine and fake) - obtained from the one of the datasets of UCI Machine Learning Repositories. I've trained the model on training data and tested the model with a validation set for predicting the authenticity of the notes and observed that the predictors – Variance, Skewness and Kurtosis of Wavelet Transformed Image of the notes play significant role in determining the genuineness of the currency notes. An accuracy of 97.81% was achieved using this model.

Dataset

The dataset used for the analysis of bank notes authentication is obtained from Volker Lohweg and Helene DÃrksen[2]. Data were extracted from images that were taken from genuine and forged banknote-like specimens. Wavelet Transform tool were used to extract features from images.

Basic Information about the dataset:

No of observations: **1372 records**

Predictors – 'Variance', 'Skewness' and 'Curtosis' of Wavelet Transformed Image and 'Entropy'

Output - 'Class' with values 0s (Counterfeit) and 1s (Genuine)

Training to Test Data Ratio - 80:20

Snippet of the Dataset:

Variance	Skewness	Curtosis	Entropy	Class
3.6216	8.6661	-2.8073	-0.44699	0
4.5459	8.1674	-2.4586	-1.4621	0
3.866	-2.6383	1.9242	0.10645	0
3.4566	9.5228	-4.0112	-3.5944	0
0.32924	-4.4552	4.5718	-0.9888	0
4.3684	9.6718	-3.9606	-3.1625	0
3.5912	3.0129	0.72888	0.56421	0
2.0922	-6.81	8.4636	-0.60216	0

mnavane@g.clemson.edu

Exploratory Data Analysis

Before proceeding on with model selection and prediction, EDA was performed to get an idea of the main characteristics of the predictors in the datasets. Following are my observations:

- Histograms of all four predictors were plotted in R to visualize the data distribution
    a. Observed that histograms of Curtosis and Entropy are right and left skewed respectively
    b. Histogram of Variance appears to be normally distributed whereas as for Skewness, it appears to be multi-modal

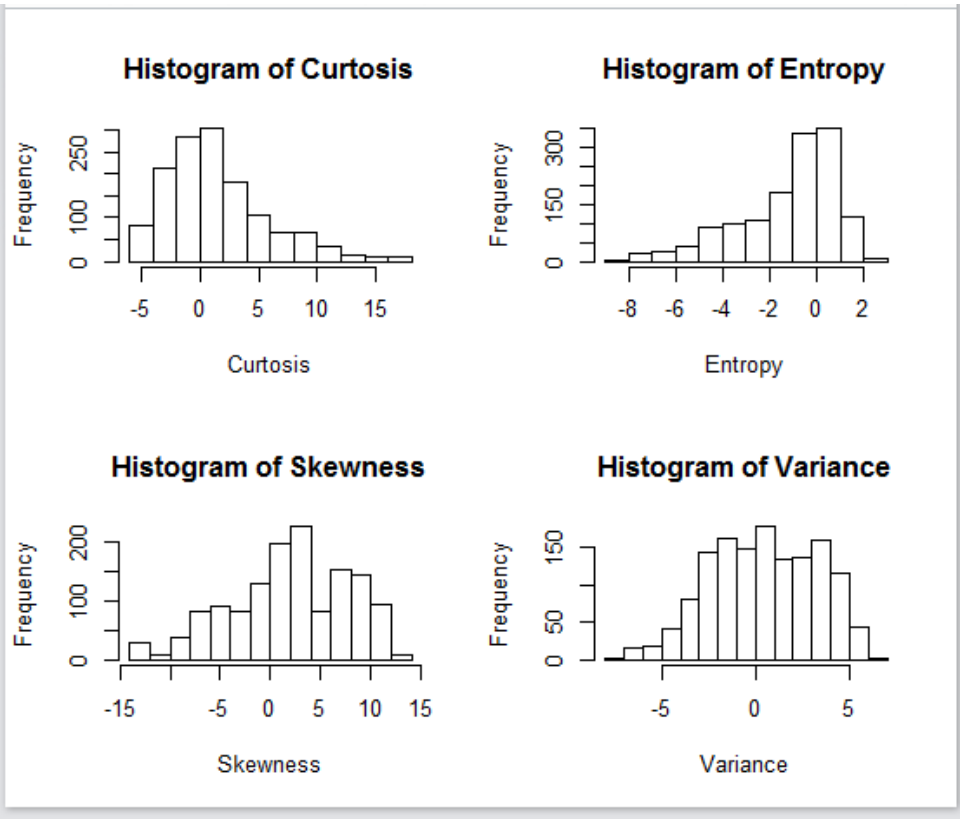

Figure 1 – Histograms of four Predictors – Curtosis, Skewness, Entropy and Variance

- To understand the correlation among the predictors and the target, **correlation matrix** was generated (Figure 2):
    - As we can see from the matrix below, the Pearson coefficients for the predictors do not suggest any multi-linearity and hence we can take all the predictors into consideration
    - We also notice that 'Variance', 'Skewness' and 'Entropy' are negatively correlated to our target 'Class', whereas 'Curtosis' is positively correlated
    - An overview shows that 'Variance' and 'Skewness' could be strong predictors, however further analysis will give us more insight on the actual behavior

	Variance	Skewness	Curtosis	Entropy	Class
Variance	1.0000000	0.2640255	-0.3808500	0.27681670	-0.72484314
Skewness	0.2640255	1.0000000	-0.7868952	-0.52632084	-0.44468776
Curtosis	-0.3808500	-0.7868952	1.0000000	0.31884089	0.15588324
Entropy	0.2768167	-0.5263208	0.3188409	1.00000000	-0.02342368
Class	-0.7248431	-0.4446878	0.1558832	-0.02342368	1.00000000

Figure 2 – Pearson's Correlation Matrix for Predictors and Target

- Summary() was performed on the dataset understand the dataset furthermore

```
> summary(mydata)
    Variance          Skewness          Curtosis          Entropy
 Min.   :-7.0421   Min.   :-13.773   Min.   :-5.2861   Min.   :-8.5482
 1st Qu.:-1.7730   1st Qu.: -1.708   1st Qu.:-1.5750   1st Qu.:-2.4135
 Median : 0.4962   Median :  2.320   Median : 0.6166   Median :-0.5867
 Mean   : 0.4337   Mean   :  1.922   Mean   : 1.3976   Mean   :-1.1917
 3rd Qu.: 2.8215   3rd Qu.:  6.815   3rd Qu.: 3.1793   3rd Qu.: 0.3948
 Max.   : 6.8248   Max.   : 12.952   Max.   :17.9274   Max.   : 2.4495
```

Model Selection

Since the result of scrutinizing a bank note can either be 'Genuine' or 'Counterfeit' and nothing in between. Hence, training my dataset for linear regression is ruled out.

Therefore, I proceeded with '***Logistic Regression***' model where outcome is categorical or discrete (0 for Counterfeit notes and 1 for Genuine notes)

I performed **Best Subset Selection** for feature selection, using regsubsets() and observed that 'Variance' has the highest significance followed by 'Skewness', 'Curtosis' and 'Entropy'. Since all 4 variables have predicting power, all of them are used in training the model.

```
> bestsub = regsubsets(Class ~   Skewness + Variance + Curtosis + Entropy, data = mydata, nvmax = 10)
> summary(bestsub)
Subset selection object
Call: regsubsets.formula(Class ~ Skewness + Variance + Curtosis + Entropy,
    data = mydata, nvmax = 10)
4 Variables  (and intercept)
          Forced in Forced out
Skewness      FALSE      FALSE
Variance      FALSE      FALSE
Curtosis      FALSE      FALSE
Entropy       FALSE      FALSE
1 subsets of each size up to 4
Selection Algorithm: exhaustive
         Skewness Variance Curtosis Entropy
1  ( 1 ) " "      "*"      " "      " "
2  ( 1 ) "*"      "*"      " "      " "
3  ( 1 ) "*"      "*"      "*"      " "
4  ( 1 ) "*"      "*"      "*"      "*"
>
```

Once the model was trained, Error Rate on trained model was calculated - 0.02367942.

 mnavane@g.clemson.edu

Validation

To evaluate the performance of our classifier on the ***test data set,*** I followed **validation set approach** and observed an **error rate of 0.02189781** with an **accuracy** of **97.81 %**

Conclusion

In this project, with the application of supervised classification (Logistic Regression with multiple predictors), the challenge to solve the problem of counterfeit currency notes is addressed. With an accuracy of 97.8%, the model stands good and can be concluded that the features – Variance, Skewness, Curtosis of the Wavelet Transformed Image of the currency notes (and with a measure of Entropy) play very important role in differentiating fake notes from genuine ones.

References

[1] https://en.wikipedia.org/wiki/Counterfeit_money

[2] https://archive.ics.uci.edu/ml/datasets/banknote+authentication #Source