

學號:0411509 姓名:許家維

1. 選用 DATA

banknote authentication Data Set

- <http://archive.ics.uci.edu/ml/datasets/banknote+authentication>

他的參數主要是透過 Wavelet Transformed 得到鈔票圖像的

- A. Variance
- B. Skewness
- C. Kurtosis

並也把圖像亂度(entropy)也當成 DATA,

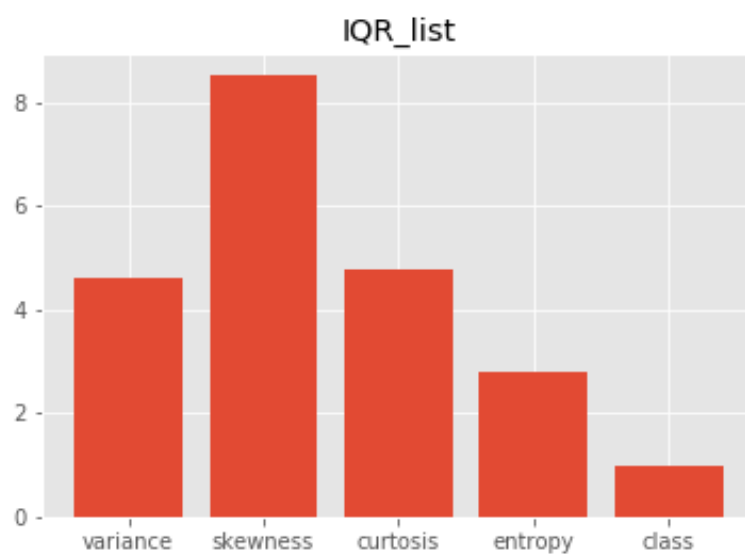
目標設定為能夠辨識出真假鈔票(class)

1. Calculate IQR(Q3-Q1) for numeric variables

我將以上幾個 data 項目的 Q3-Q1 並列在下表，可以發現

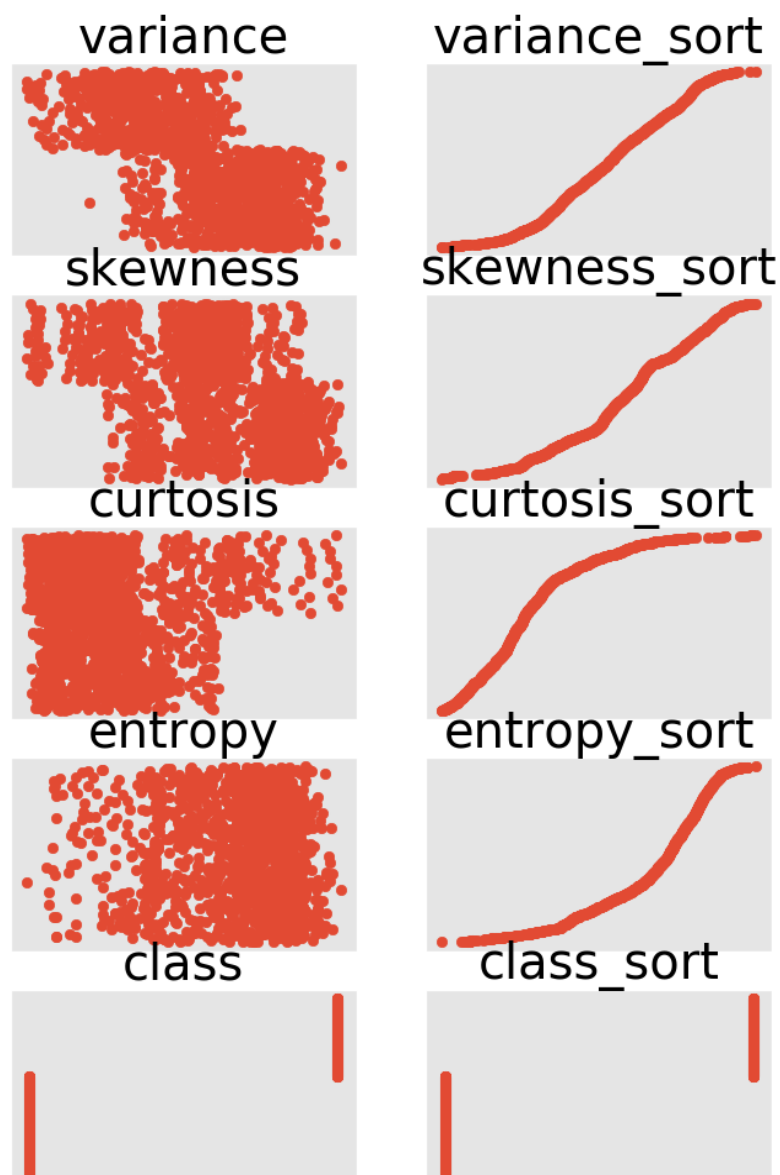
skewness 有著最大的四分位距，後面的 boxplot 也證明此項散步

較廣。



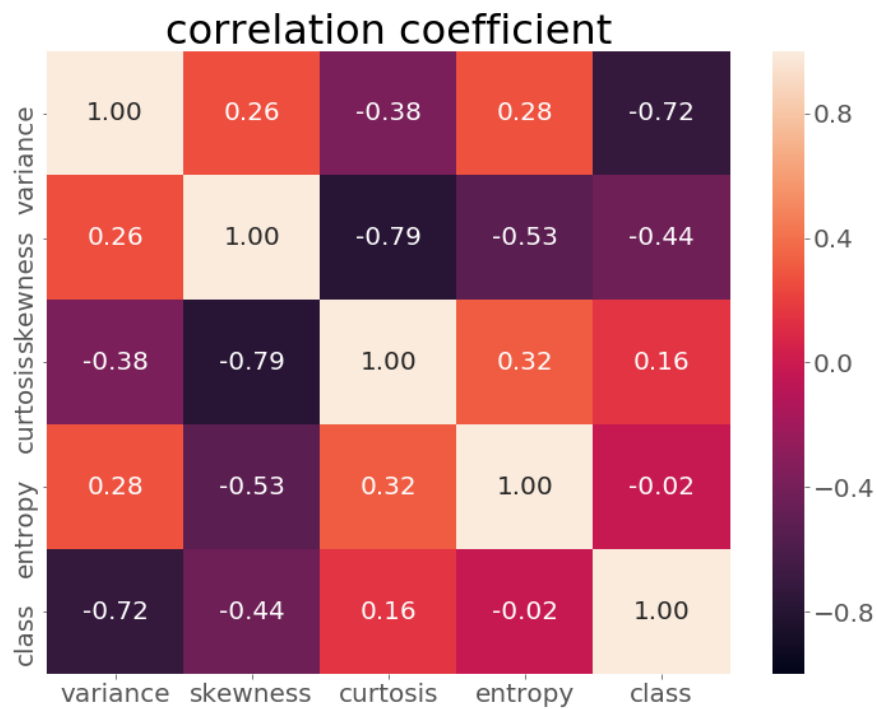
2. Sorting or ordering numeric variables

我為了觀察數據我將 x 軸當成數據，y 軸當成序號排列。左邊為未排序的數據，右邊為排序後的數據，可以看到有成功的以升序排列數據。



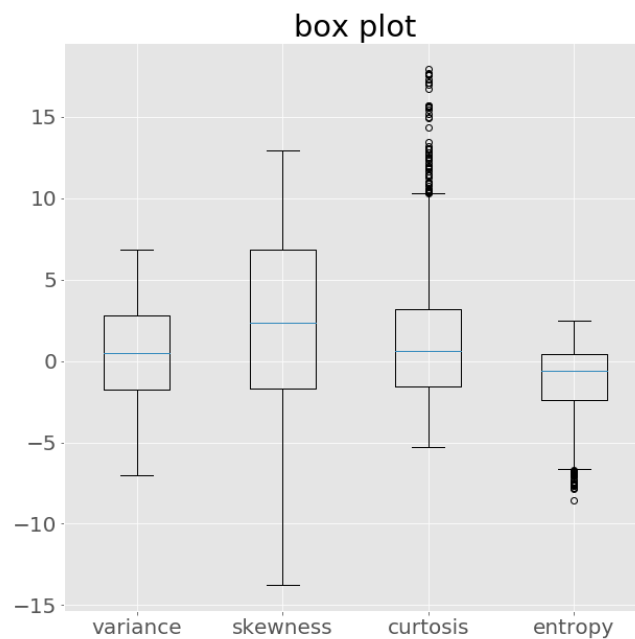
3. Show correlation matrix

我建立了各係數為 X,Y 軸的棋盤，並對應該係數的相關係數，其中暗色系為負相關，亮色系為正相關可以找到 variance 和 class 有高度負相關，這是蠻重要的線索。



4. Draw box plot for category

我使用前 4 個數據繪製 boxplot，可以發現 skewness 有很大的 range，並且額外發現 variance 和 kurtosis 雖然分布相似但是 kurtosis 包含了離群值。這些離群值都是鈔票為正確的 class

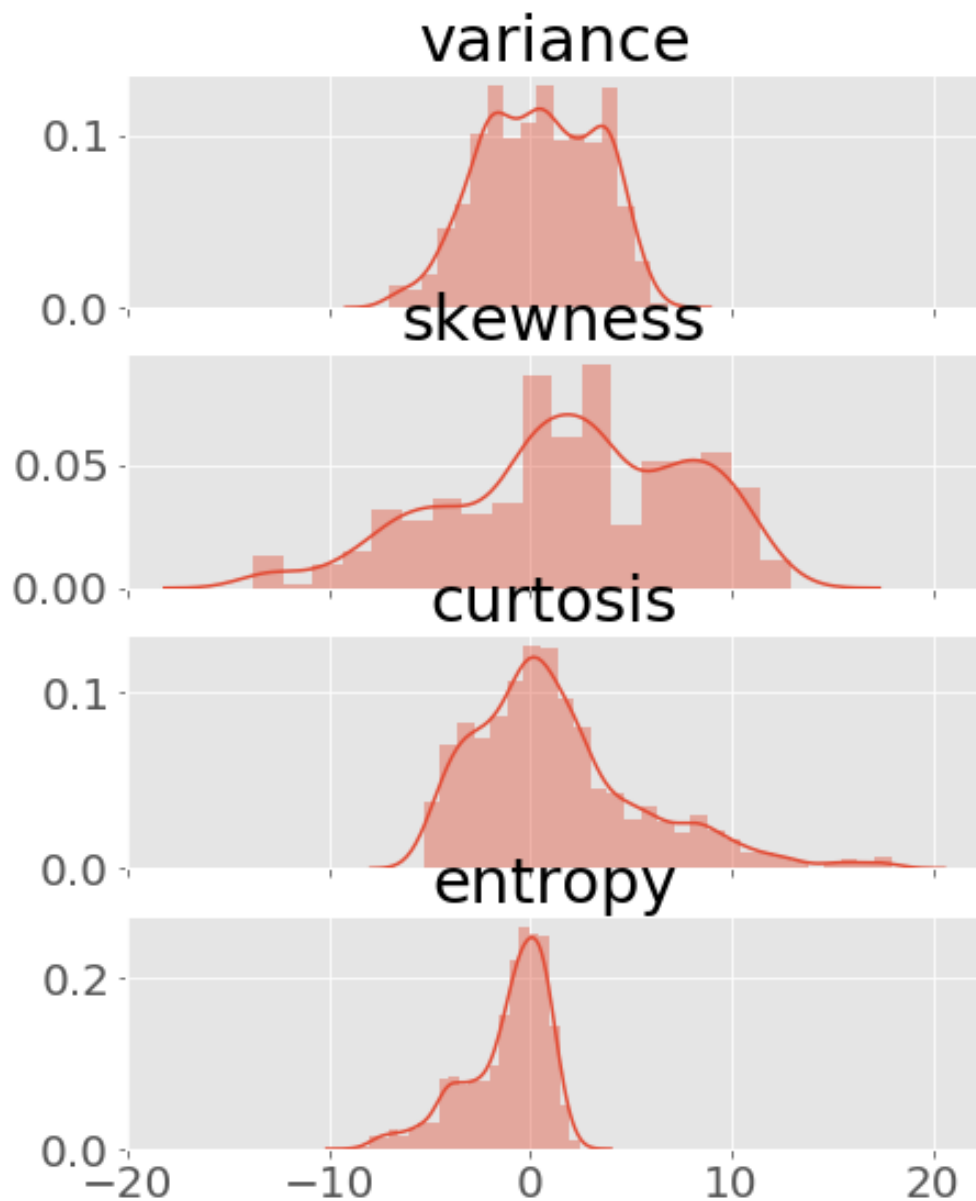


5. Draw a histogram with density curve

這是各係數的 histogram 並且描繪出密度曲線，可以看到

skewness 左偏，然後 kurtosis 右偏，但我檢查了左偏和右偏的區

域並未有特殊的 class 分布，所以我並未找到直接關係



6. Give your point of view what you found in this dataset

A. 其中 variance 和 class 有高度負相關

B. kurtosis 離群值皆為 class 1(真鈔)