

K-Means Clustering Project

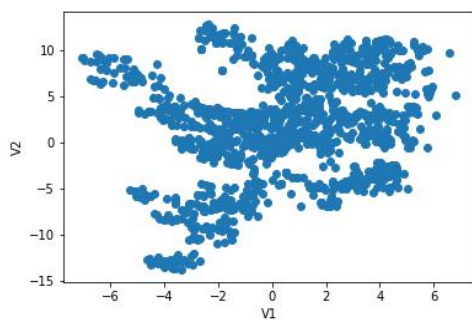
Purpose

The purpose of this project is to build a K-Means clustering model to detect the banknote is genuine or not, it's very helpful for Client OOO to check the authenticity of banknotes.

Data set

The data set was downloaded from OpenML. The visualization result is shown in the figure below.

- V1 is variance of this data set.
- V2 is skewness of this data set.

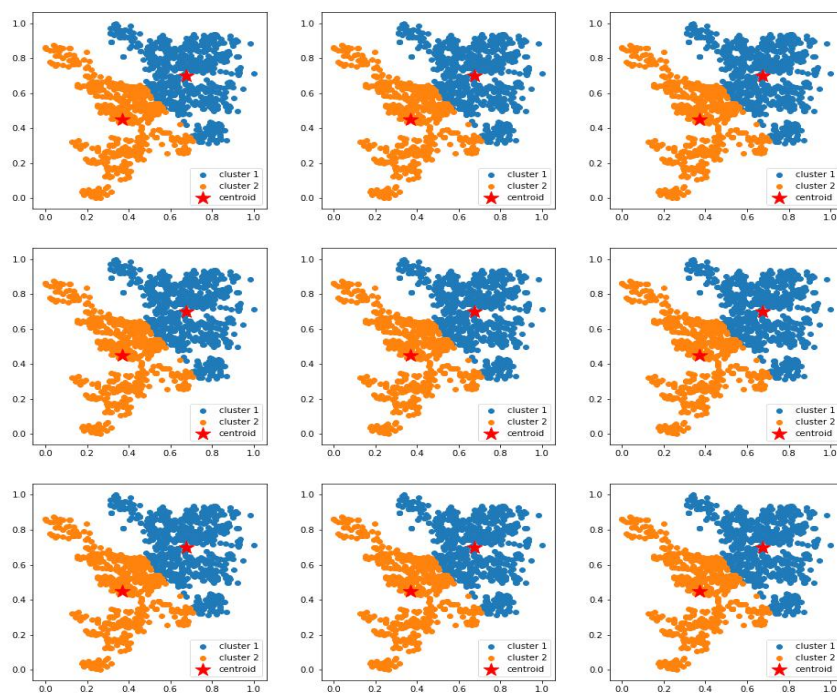


Methods

- Used **Numpy** to calculate the mean value and the standard deviation value of "Variance" and "Skewness". Find the minimum and maximum value, then we can use the formula to calculate **normalized versions** of the data set.

	V1	V2
count	1372.000000	1372.000000
mean	0.433735	1.922353
std	2.842763	5.869047
min	-7.042100	-13.773100
25%	-1.773000	-1.708200
50%	0.496180	2.319650
75%	2.821475	6.814625
max	6.824800	12.951600

- Used **K-Means model** to predict the classification status of the data.
- Try more times until the model is stable.

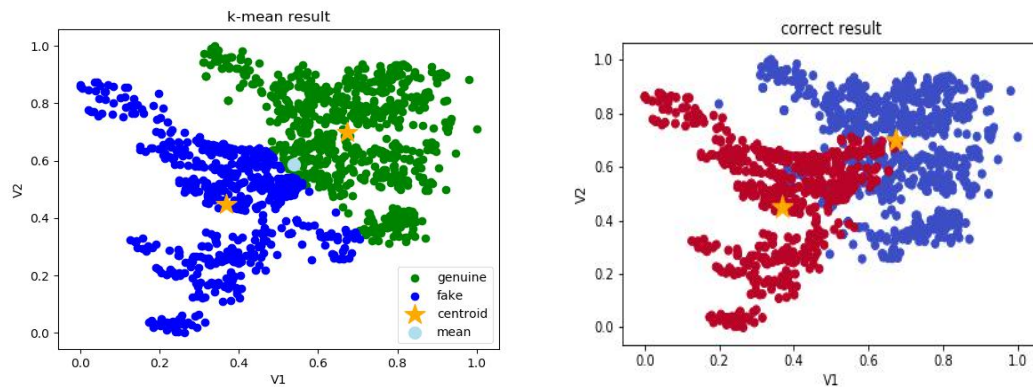


Summary of the results

The number of classification results for genuine banknotes was 765(Green), and the number of classification results for forged banknotes was 607(Blue).

The visualization result is shown in the figure "K-mean result".

I used the loop function to compare the K-mean predict result(fig.K-mean result) with the correct result data(fig. correct result). My result's calculation is as high as **96%**.



Recommend

I am very honored to be able to design this model for Client OOO to make a distinction with genuine and forged banknotes.

Since this data set is difficult to analyze with eyeballs, this is why you need use this model to train the data.

According to the results, it showed us a discrimination rate of up to 96%. They can almost find a lot of genuine banknotes. So I highly recommend this model to Client OOO.