Automatic Banknotes forgery detection with artificial intelligence - report

Author: Paweł Marczak

The task of binary classification of true and forged banknotes seems to be an ideal fit for AI methods application. Before testing an actual model to fit the data some pre-analysis must be done. Data used in experiments was downloaded from https://www.openml.org/search?type=data&status=active&id=1462&sort=runs .

It is represented by forgery - honest labels and 4 variables representing features extracted by wavelet transformation - advanced tool used in computer vision. It consists of 1372 samples, from which 610 are labeled as forgery and 762 as honest notes, so the dataset is balanced in respect to decision classes. That's usually good for training machine learning models, but it might raise doubts about the quality of tests. In real life, deployed model would be mostly used for classifying honest notes and forgeries are more uncommon. Due to the fact that 1372 samples is relatively not much for both training and testing, all samples were used. It's worth mentioning that increasing the volume of the data set undoubtedly would increase both accuracy of the model and reliability of tests.

	V1	V2	Class
count	1372.000000	1372.000000	1372.000000
mean	0.433735	1.922353	0.444606
std	2.842763	5.869047	0.497103
min	-7.042100	-13.773100	0.000000
25%	-1.773000	-1.708200	0.000000
50%	0.496180	2.319650	0.000000
75%	2.821475	6.814625	1.000000
max	6.824800	12.951600	1.000000

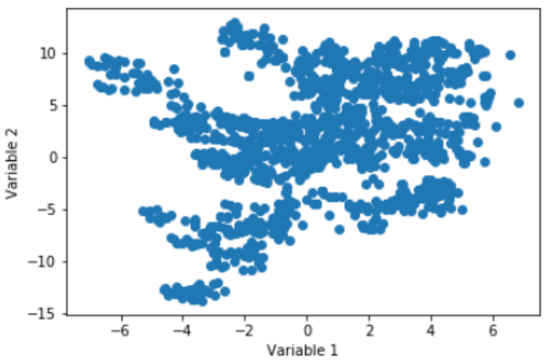
Data set statistics

	V1	V2	Class
count	610.000000	610.000000	610.0
mean	-1.868443	-0.993576	1.0
std	1.881183	5.404884	0.0
min	-7.042100	-13.773100	1.0
25%	-3.061450	-5.810025	1.0
50%	-1.806100	0.172775	1.0
75%	-0.541770	3.189275	1.0
max	2.391700	9.601400	1.0

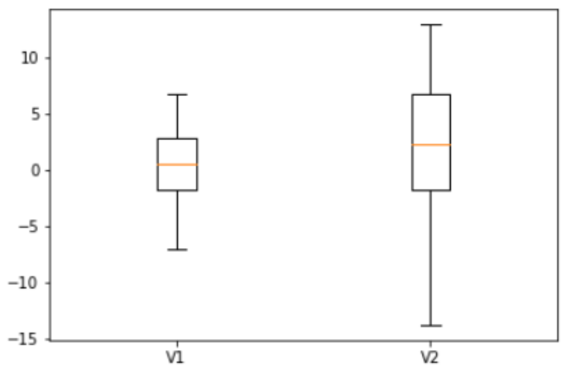
Statistics for forgery banknotes

	V1	V2	Class
count	762.000000	762.000000	762.0
mean	2.276686	4.256627	0.0
std	2.019348	5.138792	0.0
min	-4.285900	-6.932100	0.0
25%	0.883345	0.450063	0.0
50%	2.553100	5.668800	0.0
75%	3.884450	8.691975	0.0
max	6.824800	12.951600	0.0

Statistics for real banknotes

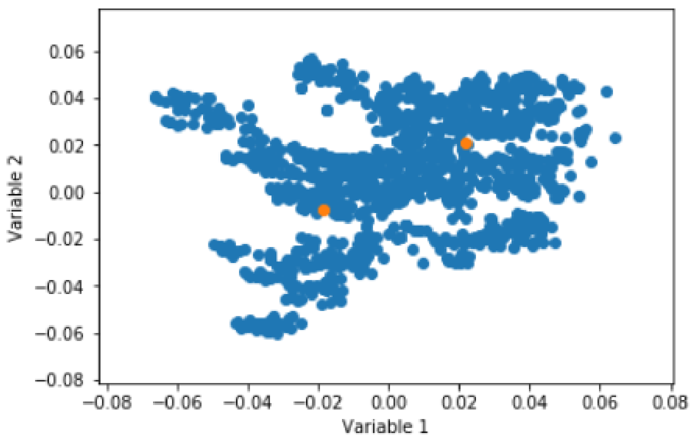


Dataset features overview (scatter plot)



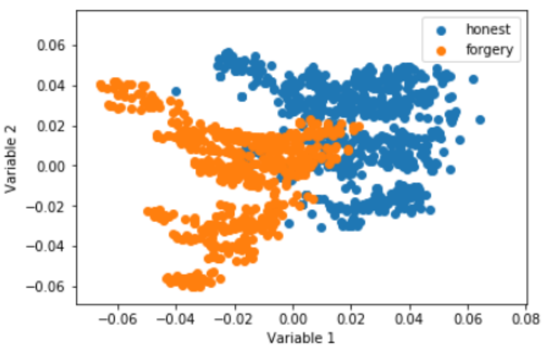
Features boxplots

The dataset was limited to only two most informative variables (V1, V2) to make visualizations more transparent and helpful in making inferences and to reduce possible harmful effects of the rest of variables on the model (too much information might not make the model smarter). Variables differed in value range, which was the reason to normalize them before introducing them into the model (to reduce bias that could be caused by higher values of one variable). Scatter plot of features did show some clearly distinct clusters of points, which motivated using K-means clustering classification model dedicated to cases like this.

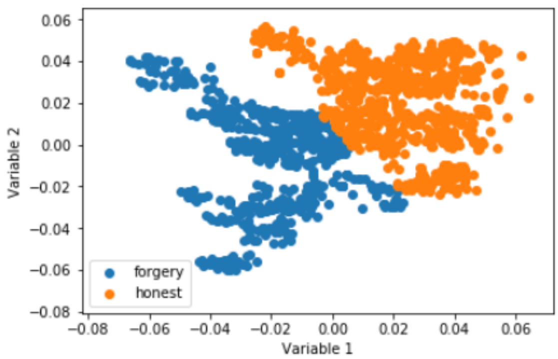


Clusters centers

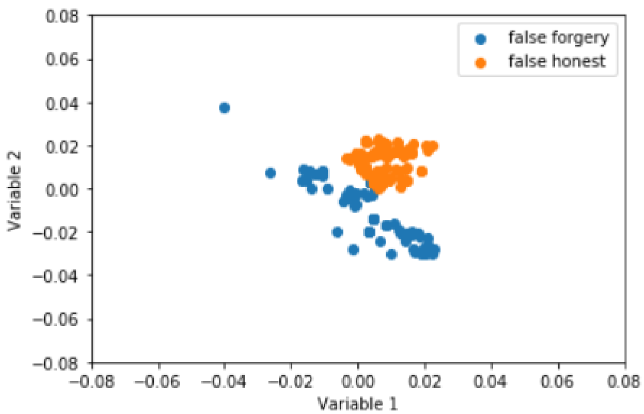
The pictures showing both true and predicted distribution of classified points prove that the chosen method of 2-means clustering is reasonable and fitted to chosen features (found clusters overlap well with ground truth picture). However, analysis of mistakes showed that near the border of classes there are a few misclassified points, indicating that another, nonlinear, supervised model could be a better match for the problem or supporting the model by additional variables could be tested. Both approaches would require detailed analysis and additional resources.



True class distribution



Predicted class distribution



Model mistakes analysis

The overall accuracy of the model was tested on the entire data set and it was correct in about 88% of cases. The model managed to identify honest notes (accuracy 89%) slightly better than forgeries (accuracy 86%). It might be a good trait, considering that the majority of notes taken in by the bank are honest. The model was trained and tested 20 times and achieved results each time were greatly similar (standard deviation of accuracy was about 10^{-4}) which indicates that the model converged (without manipulating the approach and / or the data set, the model is stable and improving it' s performance is impossible).

	precision	recall	f1-score	support
honest	0.88	0.90	0.89	762
forgery	0.87	0.85	0.86	610
avg / total	0.88	0.88	0.88	1372

Model test report

Recommendations

- Proposed model may be used as a part of banknote validation process, but shouldn' t be used fully autonomously as it clearly has its flaws
- In some cases, results of the model may be treated with good confidence, depending on the values of extracted features. As prepared analysis shows, cases from variable 1 normalized range (-0.04, 0.03) and simultaneously from variable 2 normalized range (-0.04, 0.03) should be treated with great caution and should be checked with additional procedures (e.g. expert's opinion), the rest can be classified automatically.
- Prepared model doesn't accommodate the fact that finding even one false note among the deposit of a person increases the probability that the rest of their notes are forged. Banks should examine with great caution the entire deposit even if only one note is classified as forgery with good confidence.
- More research, analysis and tests could significantly increase performance of the model and consequently increase its application potential.
- Supporting the author with additional real-life data could improve the quality of analysis.