BANKNOTE AUTHENTICATION USING K-NEAREST NEIGHBORS

CMSC 191 - MACHINE LEARNING

Harold R. Mansilla
Department of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Manila
hrmansilla@up.edu.ph

Virgilio M. Mendoza III
Department of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Manila
vmmendoza1@up.edu.ph

May 14, 2019

1 Introduction

1.1 Banknote Authentication

A banknote is a "a promissory note issued by a bank payable to bearer on demand without interest and acceptable as money" [1]. A major issue surrounding banknotes is counterfeit or inauthentic banknotes, which serve to imitate the real ones in point of trade. Thus, the process of banknote authentication is deemed an important practice for parties concerned, in order to suppress fraudulent acts surrounding these.

Banknote authentication involves inspecting the bills themselves while looking for signs of counterfeiting. This paper aims to show how machine learning applications can aid in this domain for automating this process.

1.2 K-Nearest Neighbors

K-nearest-neighbor (kNN) classification is one of the most fundamental and simple classification methods. Developed from the need to perform discriminant analysis when reliable parametric estimates are difficult to determine, it is one of the first choices for classification when there is little to no prior knowledge about the data [2].

2 The Dataset

The dataset was retrieved from [3]

Table 1: Features in the dataset

Feature	**Type**	**Description**
Variance	Numerical	Variance of Wavelet Transformed image
Skewness	Numerical	Skewness of Wavelet Transformed image
Kurtosis	Numerical	Kurtosis of Wavelet Transformed image
Entropy	Numerical	Entropy of image
Class	Categorical	Type of banknote

The target value is the `Class` column wherein the possible values are 0 for authentic and 1 for inauthentic. From [3], the number of observations for each class is not balanced. There are 1,372 observations in total.

3 KNN Classifier Construction

For purposes of experimentation with the effect of the parameter K to the overall accuracy of the derived classifier, Grid Search cross validation with 5 folds was used wherein the parameter grid consists of values of the `n_neighbors` parameter ranging from 1 to 25. These configurations are available in the `scikit-learn` Python library.

4 Results

Figure 1 shows the training and test accuracies of the KNN classifier against the values of `n_neighbors`. The highest test accuracy recorded was a perfect 100% at multiple values of `n_neighbors`, one of which is `n_neighbors = 1`. It can be noticeable how the scores fluctuate starting from `n_neighbors = 14`.

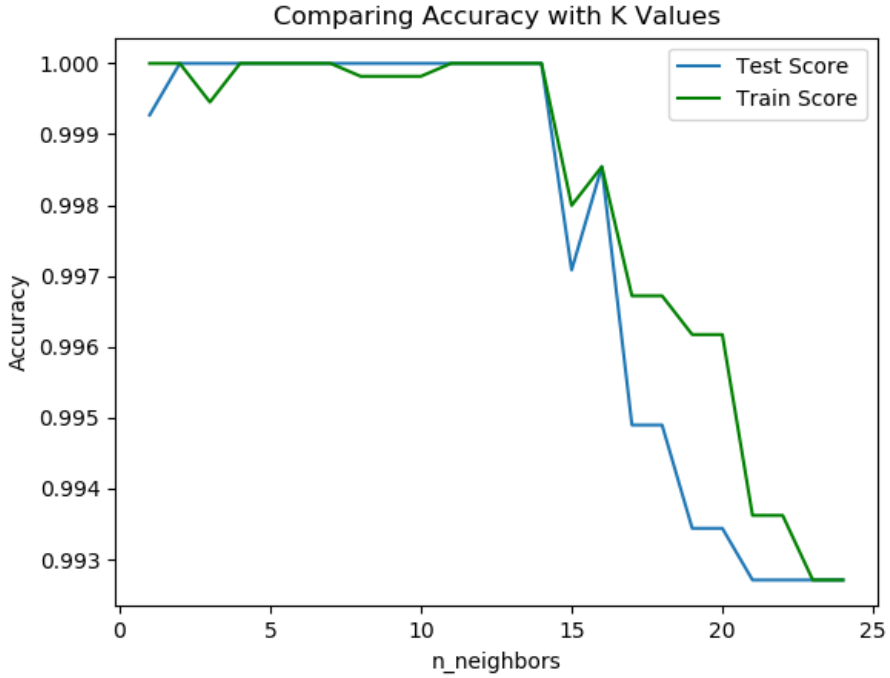

Figure 1: Plot of Accuracy against the `n_neighbors`

5 Conclusion

The effect of the number of neighbors in KNN classifier performance was experimented on. For this dataset, 0 to 14 neighbors proved optimal to the performance of the classifier, resulting in a perfect accuracy.However, increasing this number (particularly from 15 onwards), results in lower performance. However, with the range of accuracy values starting from around 0.993, it can be said that this is negligible.

References

[1] Merriam-Webster.com, *Banknote*. [Online]. Available: `https://www.merriam-webster.com/dictionary/banknote` (visited on 05/19/2019).

[2] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[3] J. Brownlee, *10 standard datasets for practicing applied machine learning*. [Online]. Available: `https://machinelearningmastery.com/standard-machine-learning-datasets/`.