Wavelet Analysis for Banknote Authentication

Introduction

This report evaluates the ability of a k-means clustering algorithm, with $k = 2$, to distinguish between genuine and counterfeit banknotes. The goal is to assess the potential of this approach as a basis for an automated authentication system. By leveraging features derived from wavelet transformations, the project focuses on understanding how effectively clustering can group similar banknotes.

Data Description

The dataset used in this project is a simplified version of the Banknote Authentication dataset. It consists of two numerical features:

- **Variance:** A measure of the spread of pixel intensity values obtained from the wavelet transformation of banknote images.
- **Skewness:** An indicator of asymmetry in the pixel intensity distribution.

The dataset includes 1,372 instances, equally split between genuine and counterfeit banknotes.

To explore the relationship between Variance and Skewness, a pairplot was created (Figure 1). This visualization reveals a clustering tendency, suggesting that the two features provide some separation between genuine and counterfeit banknotes.

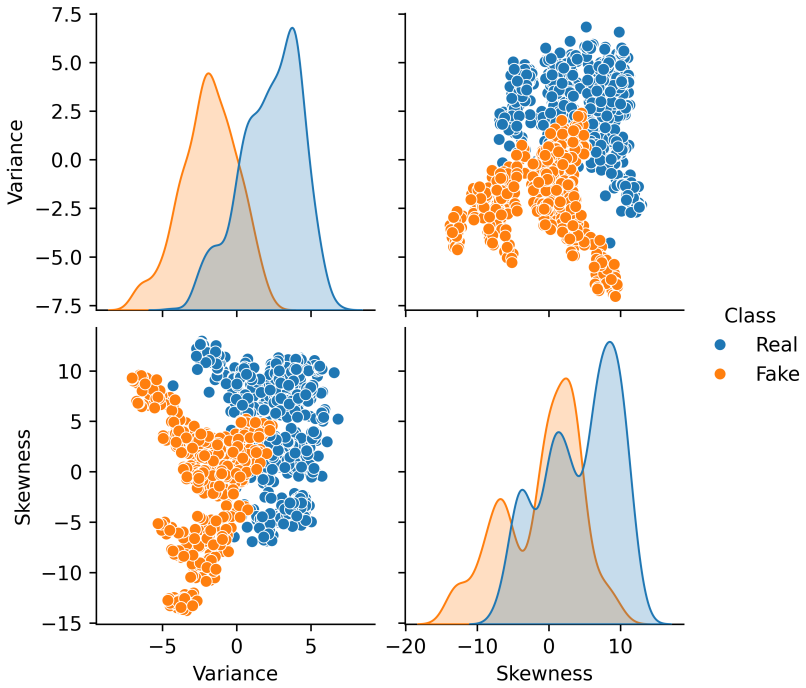

Figure 1: Pairplot of Variance and Skewness. The plot highlights the distribution of the two features and hints at potential clustering patterns. The separation is done using the validation dataset from openML.

Methods

The analysis process involved the following steps:

1. **Exploratory Data Analysis:** Variance and Skewness were visualized to understand their distributions and relationships.

2. **Wavelet Transformation:** Features were derived from wavelet-transformed banknote images, providing a compact representation.

3. **K-Means Clustering:** The algorithm was applied with $k = 2$, grouping the dataset into clusters that correspond to genuine and counterfeit banknotes.

4. **Validation:** The clustering results were evaluated using the OpenML benchmark, yielding an accuracy of 65%.

Results

The k-means clustering algorithm successfully formed two clusters, as shown in Figure 2. Data points in the scatterplot are color-coded according to their cluster assignments. While the clusters show reasonable separation, the OpenML validation score of 65% indicates that the model's accuracy is moderate and may need improvement.

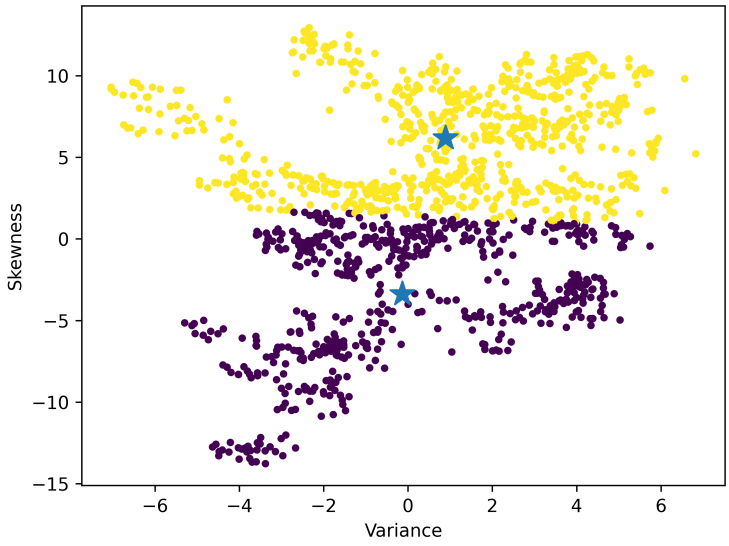

Figure 2: Scatterplot of Variance and Skewness. Points are color-coded by the cluster assigned by the k-means algorithm.

Discussion and Recommendations

The analysis demonstrates that Variance and Skewness provide some level of distinction between genuine and counterfeit banknotes. However, with an accuracy of 65%, the clustering model falls short of providing reliable predictions for practical use. To enhance performance, the following steps are recommended:

1. **Expand Feature Set:** Incorporate additional features, such as curtosis and entropy, to capture more nuanced distinctions between classes.

2. **Explore Alternative Models:** Consider supervised learning techniques like logistic regression or decision trees for improved accuracy.

3. **Validate Robustly:** Use metrics such as the Silhouette score and Davies-Bouldin Index to assess clustering quality quantitatively.

4. **Periodic Updates:** Continuously retrain the model with updated data to address evolving counterfeiting techniques.

5. **Combine with Security Measures:** Integrate the model with physical security features and manual checks for a multi-layered authentication approach.

Conclusion

This study explored the use of k-means clustering to classify banknotes using two features: Variance and Skewness. While initial results show potential, the moderate accuracy score highlights the need for further refinement and feature expansion. With additional improvements, this approach could contribute to a robust and automated system for counterfeit detection.