

Bank Note Authentication

PCA for dimensionality reduction

The project is about preparing to build classification model to predict the class of a bank note. There are two classes, counterfeit (value 1) and genuine (Value 0).

The data for this project is available at http://archive.ics.uci.edu/ml/machine-learning-databases/00267/data_banknote_authentication.txt

This data is about the characteristics of the image of the bank notes. Images of 1372 banknotes were taken, some counterfeit and some genuine. Wavelet transformation tools were used to extract the following descriptive features of the images: Variance, Skewness, Kurtosis, and Entropy. We also have the true label for whether or not the banknote is genuine (Yes = 1, No = 0)

Steps for PCA :-

1. Download the banknote data from the given website and load it into pandas dataframe.
2. Standardize the data to zscores
3. Generate the covariance matrix and print the covariance matrix to ensure the covariance is non zero
4. Using Eigen decomposition, break factor the matrix into eigen vectors and eigen values
5. Visually display the cumulative information content of eigen vectors starting with the eigen vector with largest eigen value
6. Choose the number of eigen vectors that give a cumulative of 95% information content
7. Transform the original dataset X via the projection matrix to obtain a data in the eigen vector space