Overview of Principal Component Analysis Example

Dataset Link:

<https://www.kaggle.com/akram24/wine-pca>

Dataset Content:

This dataset contains fourteen columns. They contain the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

Problem:

We would like to reduce the number of independent variables we are going to use in order to apply our classification algorithm easier.

Solution:

We want to reduce the number of independent variables using Principal Components Analysis. Then we are going to use the components to train a Logistic Regression classification algorithm in order to classify wines into three categories based on the components we found from PCA.

Variables:

Independent variable 🡪 All variables of the dataset except the target variable.

Dependent variable 🡪 Non applicable in PCA