Overview of Principal Component Analysis Example

Dataset Link:

<https://www.kaggle.com/ronitf/heart-disease-uci>

Dataset Content:

This dataset contains fourteen columns. They contain information about various health related measurements (e.g. blood pressure, electrocardiographic results etc.). The target column contains values (0 or 1) which refer to the presence of heart disease in the patient.

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 if a patient has heart disease or not based on the components resulted from PCA.

Variables:

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

Dependent variable 🡪 Non applicable in PCA