Principal component analysis (PCA) is a statistical procedure that reduces the number of dimensions (features) that is used during the machine learning process while maintaining almost the same results. Applying that process makes visualizing and analyzing data much easier.

That operation is divided to 5 steps:

1-Standardize the range of values of different features: Basically, making all values of all features lay within a common range to avoid any biased results. (Usually, ranges like: 0 to 100 or 0 to 1 are used)

2-Compute the covariance matrix to identify correlations: The purpose here is to understand how the features of the input data set are varying from each other. Sometimes, Features are highly correlated in such a way that they contain redundant information. Therefore, the correlations between features are identified in order to discard any expendable features. That is done by computing the covariance matrix.

3-Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components: Eigenvectors and eigenvalues are used to compute the covariance matrix in order to determine the principal components of the data. Every eigenvector has an eigenvalue. Their number is equal to the number of dimensions of the data. Covariance matrix contains the directions of the axes where there is the most variance (most information) and these axes are called Principal Components. The eigenvalues are the coefficients give the amount of variance carried in each Principal Component.

4-Create a feature vector to decide which principal components to keep: Here, the eigenvectors of the components that have the highest significance are used as columns for a matrix of vectors that we call Feature vector.

5-Recast the data along the principal components axes: in this step, the data from the original axes is reoriented to the ones represented by the principal components. This can be done by multiplying the transpose of the original data set by the transpose of the feature vector.