**Principal Components Analysis (PCA)**

PCA is linear algebra technique widely used for feature extraction and dimensionality reduction. This technique seeks to summarize the information content of a dataset by transforming it ont a new feature subspace of lower dimensionality that the original one. Other popular applications include exploratory data analyses, identification of correlated variables, de-noising of signals in stock market trading, as well as the analysis of genome data and gene expression levels in the field of bioinformatics.

The goal of PCA is to find a new set of dimensions (attributes) that better captures the variability of the data. More specifically, the first dimension is chosen to capture as much of the variability as possible. PCA aims to find the directions of maximum variance in high-dimensional data and project it onto a new subspace with equal of fewer dimensions than the original one.

**PCA has several appealing characteristics:**

* It tends to identify the strongest patterns in the data
* Most of the variability of the data can be captured by a small fraction of the total set of dimensions
* Since most of the noise in the data is weaker than the patterns, dimensionality reduction can eliminate much of the noise

**A goal of PCA is to find a transformation on the data that satisfies the following properties:**

* Each pair of new attributes has 0 covariance
* The attributes are ordered with respect to how much of the variance of the data each attribute captures
* The first attribute captures as much of the variance of the data as possible
* Subject to the orthogonality requirement, each successive attribute captures as much of the remaining variance as possible

**Extracting the principal components step by step:**

* Standardize the data
* Construct the covariance matrix
* Obtain the eigenvalues and eigenvectors of the covariance matrix
* Sort the eigenvalues by decreasing order to rank the eigenvectors

