Principal Component Analysis (PCA)

Principal component analysis (PCA) is an essential algorithm in machine learning. It is a mathematical method for evaluating the principal components of a dataset. The principal components are a set of vectors in high-dimensional space that capture the variance (i.e., spread) or variability of the feature space.

The goal of computing principal components is to find a low-dimensional feature sub-space that captures as much information as possible from the original higher-dimensional features of the dataset.

PCA is particularly useful for simplifying data visualization of high-dimensional features by reducing the dimensions of the dataset to a lower sub-space. For example, since we can easily visualize relationships on a 2-D plane using scatter diagrams, it will be useful to condense an n-dimensional space into two dimensions that retain as much information as possible in the n-dimensional dataset. This technique is popularly called dimensionality reduction.

How Are Principal Components Computed

The mathematical details for computing principal components are somewhat involved. This section will instead provide a conceptual but solid overview of this process.

The first step is to find the covariance matrix of the dataset. The covariance matrix captures the linear relationship between variables or features in the dataset. In a covariance matrix, an increasingly positive number represents a growing relationship, while the converse is represented by an increasingly negative number. Numbers around zero indicate a non-linear relationship between the variables. The covariance matrix is a square matrix (that means it has the same rows and columns). Hence, given a dataset with m rows and p columns, the covariance matrix will be a $m \times p$ matrix.