# Principal Component Analysis (PCA):

**Definition:** Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms a high-dimensional dataset into a lower-dimensional form while retaining as much of the original variability as possible. It achieves this by identifying a set of new, uncorrelated variables, called principal components, which are linear combinations of the original features.

**Concept:** The main idea behind PCA is to find a new coordinate system in which the data variance is maximized along the new axes. The first principal component accounts for the most significant variance, the second principal component (orthogonal to the first) for the second most significant variance, and so on. By choosing a subset of these principal components, one can represent the data in a lower-dimensional space.

**Steps of PCA:**

1. **Standardization:**
    - Standardize the dataset to have a mean of 0 and a standard deviation of 1 for each feature. This step ensures that all features contribute equally to the analysis.
2. **Covariance Matrix Computation:**
    - Calculate the covariance matrix for the standardized dataset. The covariance matrix describes the relationships between different features.
3. **Eigenvalue and Eigenvector Computation:**
    - Find the eigenvalues and eigenvectors of the covariance matrix. Eigenvectors represent the directions of maximum variance, and eigenvalues indicate the magnitude of variance along those directions.
4. **Sorting Eigenvalues:**
    - Sort the eigenvalues in descending order. This determines the importance of each principal component.
5. **Selecting Principal Components:**
    - Choose the top k eigenvectors based on the desired number of dimensions in the new feature space (k). Higher eigenvalues indicate more important components.
6. **Projection:**
    - Form a new matrix using the selected eigenvectors as columns. Project the original data onto this lower-dimensional subspace to obtain the principal components.

7. **Interpretation:**
    - Interpret the results. Each principal component is a linear combination of the original features, and they are uncorrelated with each other.
8. **Explained Variance:**
    - Optionally, compute the explained variance ratio for each selected principal component. This ratio indicates the proportion of the total variance in the data captured by each component.

PCA is widely used for data preprocessing, visualization, noise reduction, and feature extraction in various fields such as machine learning, image analysis, and signal processing.