How PCA Works?

Principal Component Analysis (PCA) is a technique used to reduce the dimensionality of a dataset while preserving as much of its variability as possible. **It works by transforming the original variables into a new set of variables, called principal components**, that are **linear combinations of the original variables**.

Here are the steps involved in performing PCA:

Standardize the data: PCA assumes that the variables are centered (mean of 0) and have a similar scale. Therefore, it is recommended to standardize the data before applying PCA.

Calculate the covariance matrix: PCA is based on the covariance matrix of the variables. The covariance matrix is calculated by taking the dot product of the standardized data matrix and its transpose.
— shows how much two or more variables change together …

Calculate the eigenvectors and eigenvalues: The eigenvectors and eigenvalues of the covariance matrix are calculated. **The eigenvectors represent the direction of the maximum variance in the data**, while the **eigenvalues** represent the magnitude of the variance along each eigenvector.

Choose the principal components: The principal components are chosen based on the **eigenvalues, with the highest eigenvalues representing the directions of the maximum variance in the data**. The number of principal components to retain is determined based on the amount of variability that needs to be preserved.

Transform the data: The original data is transformed into the new coordinate system defined by the principal components. This is done by taking the dot product of the standardized data matrix and the matrix of eigenvectors.

Interpret the results: The transformed data can be analyzed and visualized in the new coordinate system. The contribution of each original variable to each principal component can also be determined.

Overall, PCA is a powerful technique for dimensionality reduction and data exploration that can be applied in a variety of fields, including finance, biology, and social sciences.

Covariance matrix: A covariance matrix is a matrix that shows how much two or more variables change together. It gives an indication of whether two variables are positively or negatively related to each other. The diagonal of the matrix shows the variance of each variable.

Dot product: The dot product is a mathematical operation that takes two vectors and produces a scalar value. It is calculated by multiplying the corresponding components of the two vectors and adding up the products. The dot product is used in various applications, such as calculating work done by a force, finding the angle between two vectors, and calculating projections.

Transpose: The transpose of a matrix is obtained by interchanging its rows and columns. The resulting matrix has the same values as the original matrix, but the rows and columns are switched. The transpose of a matrix is useful in various applications, such as solving systems of linear equations, calculating determinants, and performing matrix operations.