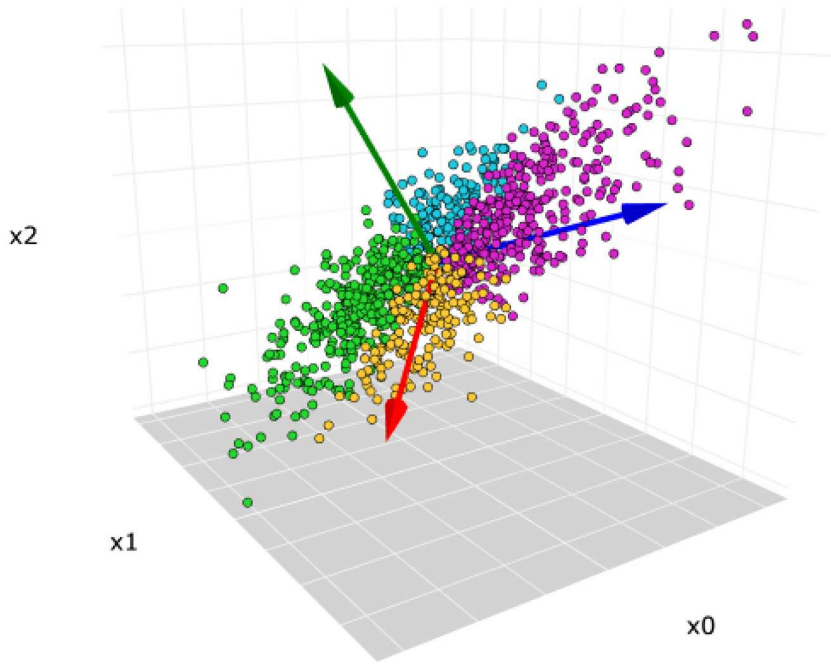
Principal Component Analysis

DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE

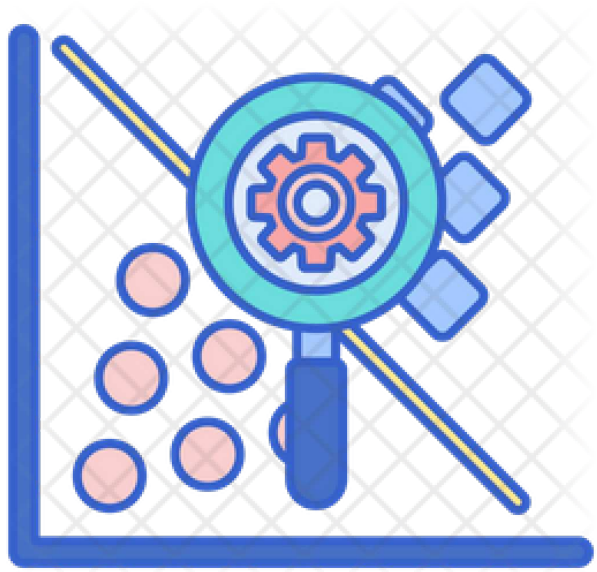
What is Principal Component Analysis?



Principal Component Analysis (PCA) is an unsupervised learning algorithm that is used for dimensionality reduction in machine learning.

PCA is used to reduce the dimensionality of a large dataset while retaining as much of the original variation as possible.

How PCA works?



PCA works by identifying a set of new variables, called principal components, that are linear combinations of the original variables, and then projecting the data onto these new variables.

The first principal component is chosen to explain the maximum amount of variation in the data, and each subsequent component is chosen to explain the remaining variation in order of decreasing importance. The principal components are chosen such that they are orthogonal (i.e., uncorrelated) to each other.

PCA Algorithm Steps

1. Standardize the data: PCA assumes that the data is standardized (i.e., centered at zero and scaled to have unit variance). This step involves subtracting the mean of each variable from each observation and dividing it by the standard deviation of each variable.

2. Compute the covariance matrix: The covariance matrix measures the pairwise covariances between all pairs of variables in the data. This matrix is used to compute the principal components.

PCA Algorithm Steps

3. Compute the eigenvectors and eigenvalues of the covariance matrix: The eigenvectors of the covariance matrix represent the directions in which the data varies the most, while the eigenvalues represent the amount of variance explained by each eigenvector.

PCA Algorithm Steps

4. Select the principal components: The eigenvectors are sorted in descending order of their corresponding eigenvalues, and the top k eigenvectors are selected as the principal components. The number of principal components to select depends on the amount of variance that needs to be explained and the desired level of dimensionality reduction.

PCA Algorithm Steps

5. Project the data onto the principal components: The original data is then projected onto the new set of k principal components. This involves computing the dot product of the original data matrix with the matrix of selected eigenvectors.

6. Analyze the results: The resulting matrix of projected data can be analyzed using standard statistical techniques. The principal components themselves can also be examined to gain insight into the underlying structure of the data.

Why it is useful?

PCA can be used for a variety of purposes, including data visualization, data compression, and data pre-processing for machine learning algorithms.

By reducing the dimensionality of the data, PCA can make it easier to analyze and interpret large datasets, and can also help to remove noise and redundancy in the data.

Advantages

1. Easy to calculate and compute.

2. Speeds up machine learning computing processes and algorithms.

3. Prevents predictive algorithms from data overfitting issues.

4. Increases performance of ML algorithms by eliminating unnecessary correlated variables.

5. Principal Component Analysis results in high variance and increases visualization.

6. Helps reduce noise that cannot be ignored automatically.

Disadvantages

1. Sometimes, PCA is difficult to interpret. In rare cases, you may feel difficult to identify the most important features even after computing the principal components.

2. You may face some difficulties in calculating the covariances and covariance matrices.

3. Sometimes, the computed principal components can be more difficult to read rather than the original set of components.

Follow **#DataRanch** on LinkedIn for more...

DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE

Follow **#DataRanch** on LinkedIn for more...

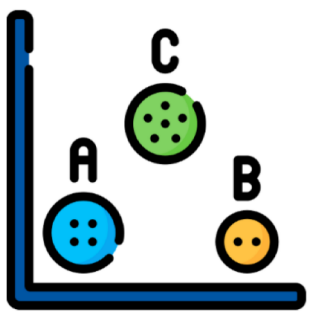
What is Supervised Learning?

DATARANCH.org

What is Unsupervised Learning?

DATARANCH.org

Clustering



DATARANCH.org

DATARANCH.org
VISUALIZE | ANALYZE | CAPITALIZE

DATARANCH.org

VISUALIZE | ANALYZE | CAPITALIZE

info@dataranch.org

linkedin.com/company/dataranch