

Feature Reduction with PCA (Principal Component Analysis)

OBJECTIVE

Feature Reduction (or dimension reduction) is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data. Principal component analysis (PCA) is a popular method for feature reduction.

Principal component analysis (PCA) invented in 1901 by Karl Pearson is a statistical method that is to represent the data by an orthogonal transformation which can convert a set of features into a smaller set of values of linearly uncorrelated variables called **principal components**. The number of distinct principal components will be less than the number of original features. PCA is sensitive to data preprocessing like the relative scaling of the original features. The objective of this lab is to implement PCA algorithm for feature reduction.

PREREQUISITES

Check the lecture slides Part II Module Data Pre processing ETL and be familiar with PCA algorithm below.

PCA Algorithm

- Step 1. Centralized the features
- Step 2. Calculate the covariance matrix of the centered features
- Step 3. Perform the eigendecomposition of the covariance matrix into respective eigenvectors and eigenvalues
- Step 4. Sort the eigenvalue and eigenvector pairs such that the eigenvalues are in descending order
- Step 5. Select top K eigenvectors
- Step 6. Construct K PCs

INSTRUCTIONS

- Implement PCA algorithm with Python
- Test PCA on **IRIS** dataset by selecting first two PCs. The data can be downloaded from <https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>.

- Plot the data with the two PCs.



DEEP
LEARNING
INSTITUTE

