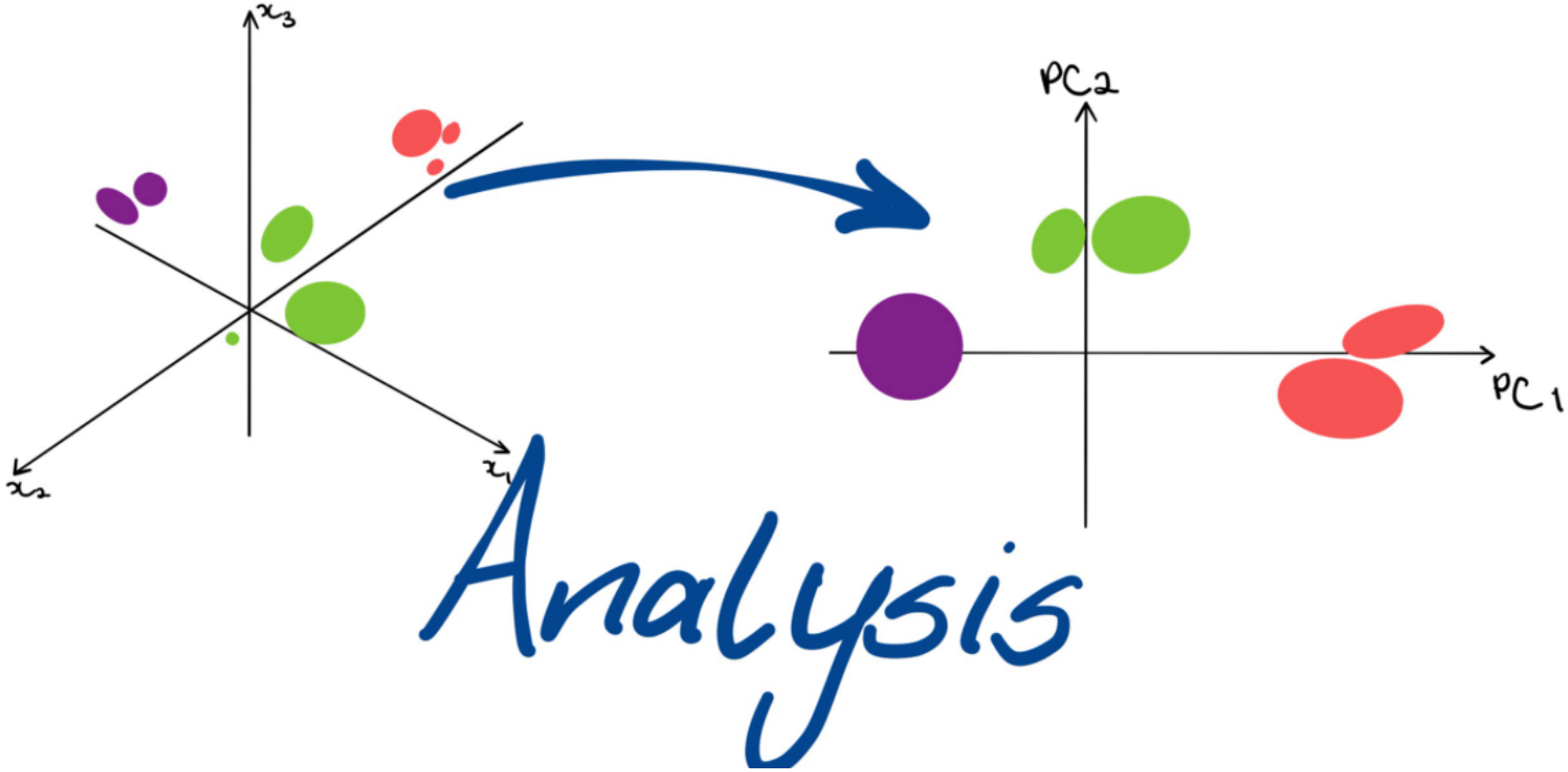
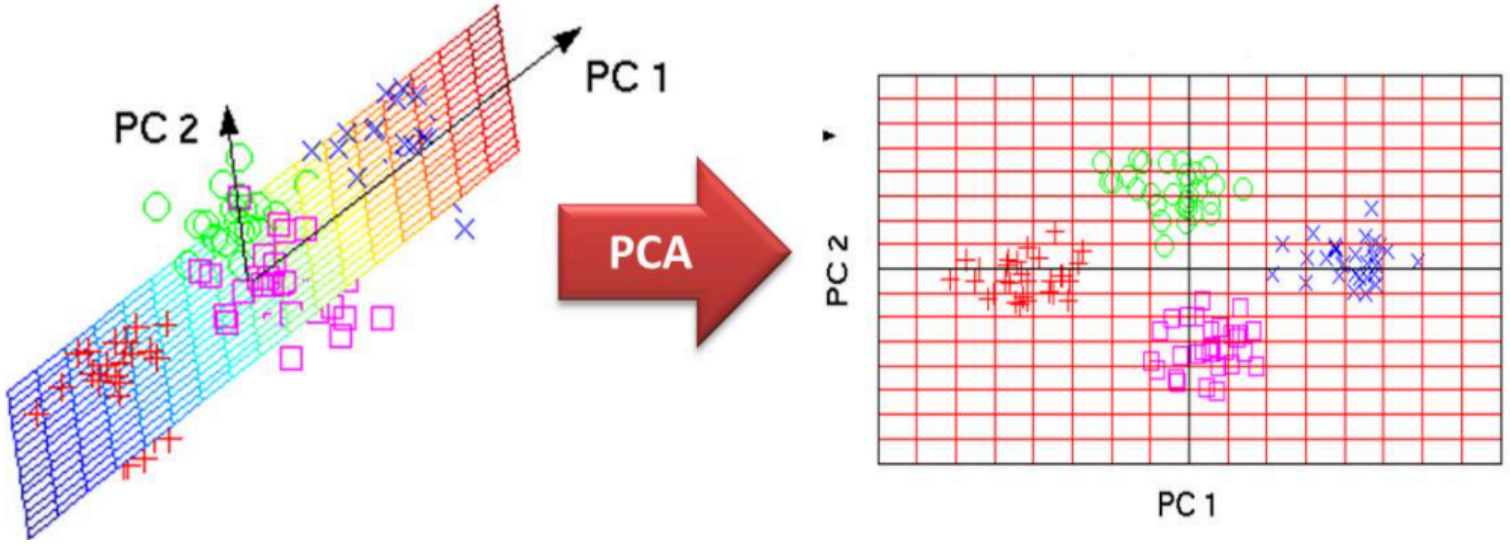
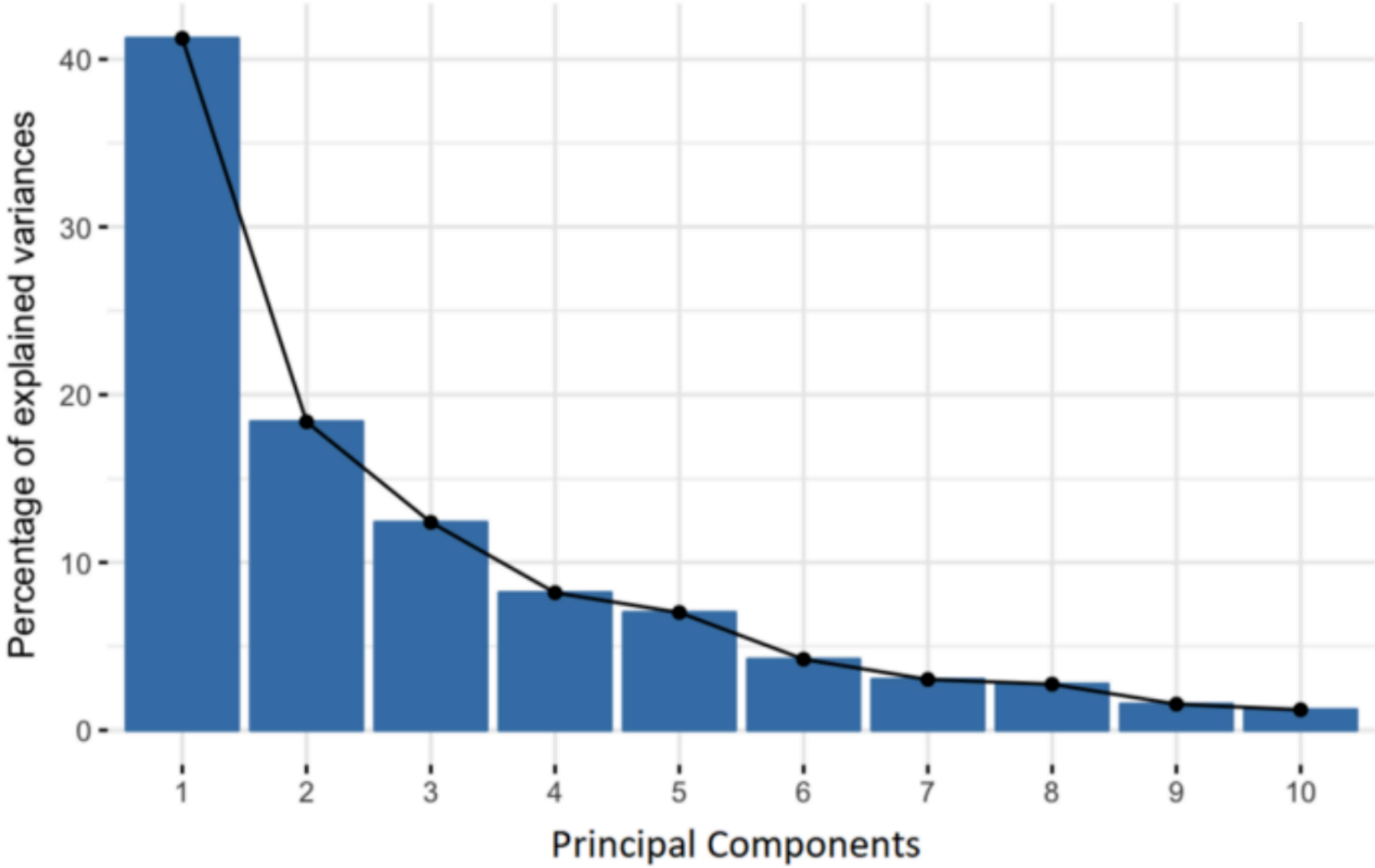
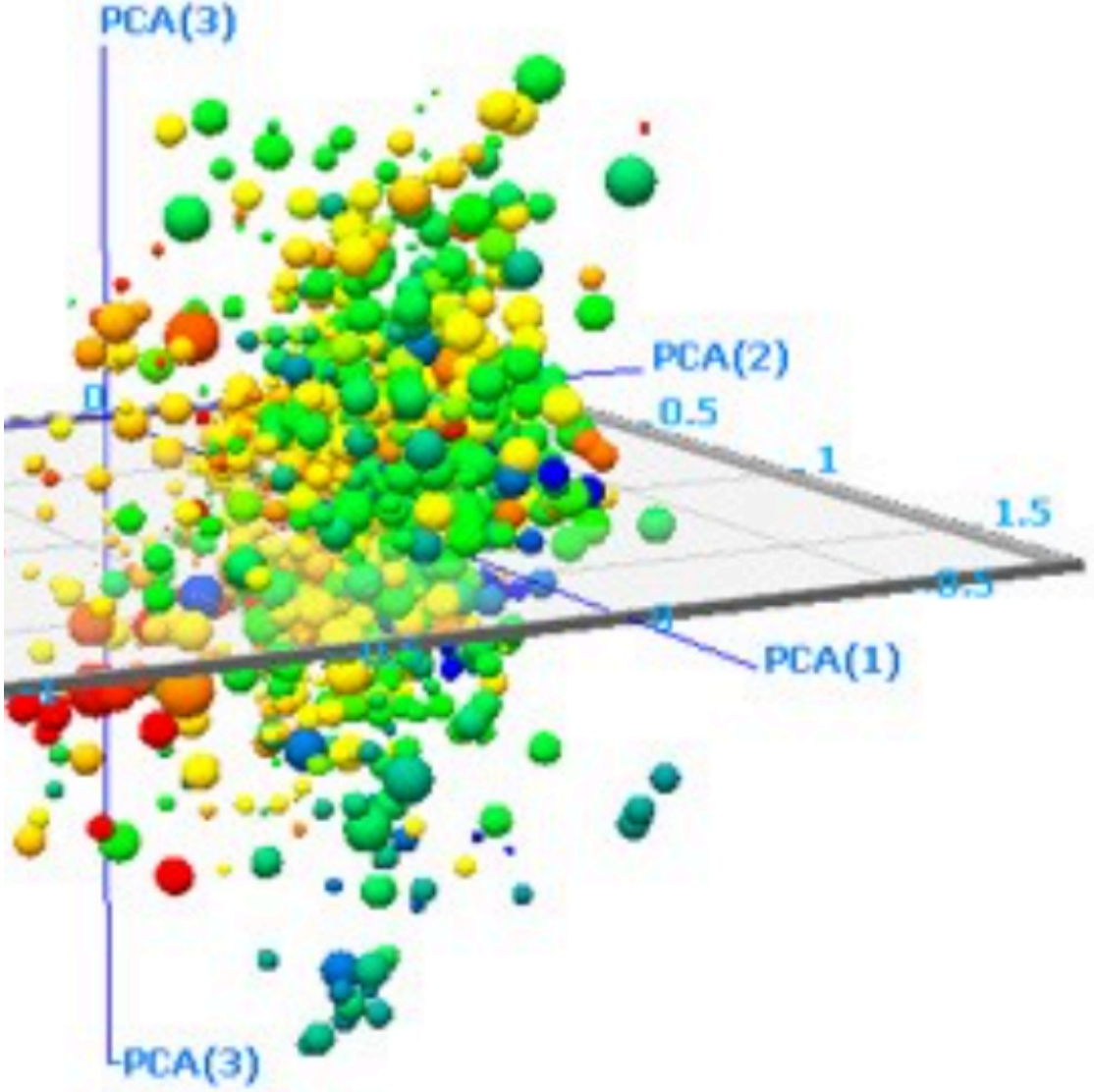
Principal Component Analysis



Dimensionality Reduction
&
Principal Component Analysis

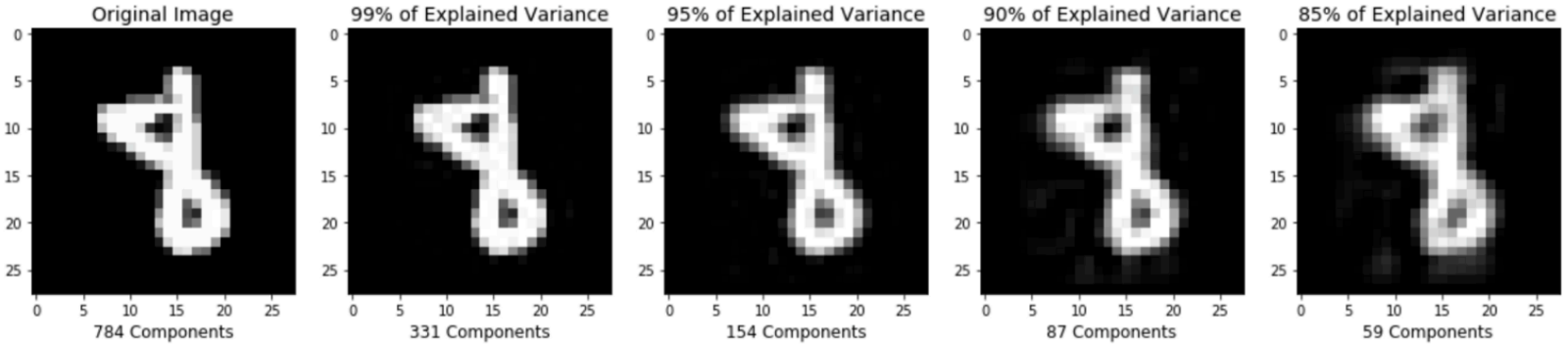



Percentage of explained variances
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
Principal Components


PCA(3)
PCA(2)
0.5
1
1.5
PCA(1)
PCA(3)

PCA for Exploratory Data Analysis

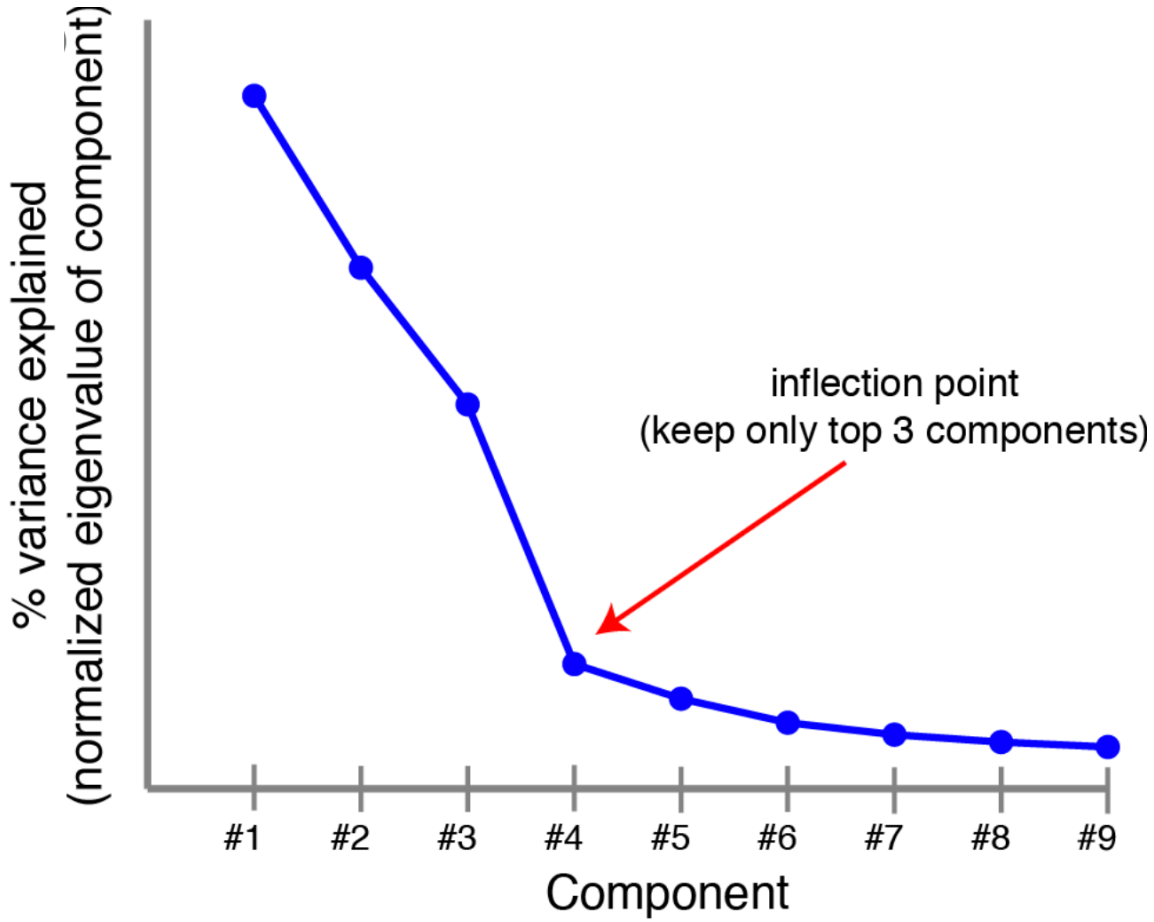
Principal Component Analysis



Original image (left) with Different Amounts of Variance Retained

PCA for Image Recognition

PCA for Machine Learning Algorithm Optimization



Principal Component Analysis

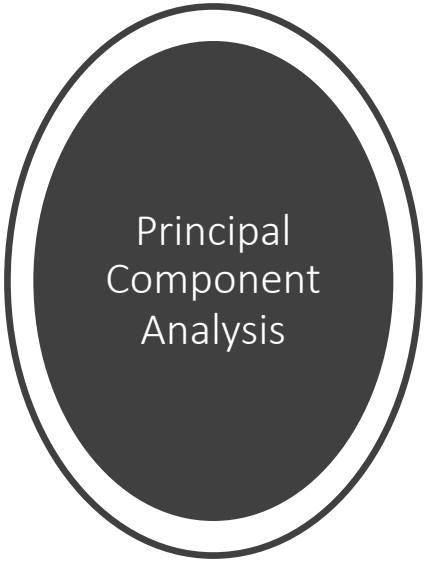
- Principal Component Analysis (PCA) is a **linear dimensionality reduction** technique that can be utilized for extracting information from a high-dimensional space by projecting it into a lower-dimensional sub-space.

- It tries to preserve the essential parts that have more variation of the data and remove the non-essential parts with fewer variation.

- Dimensions are nothing but features that represent the data.

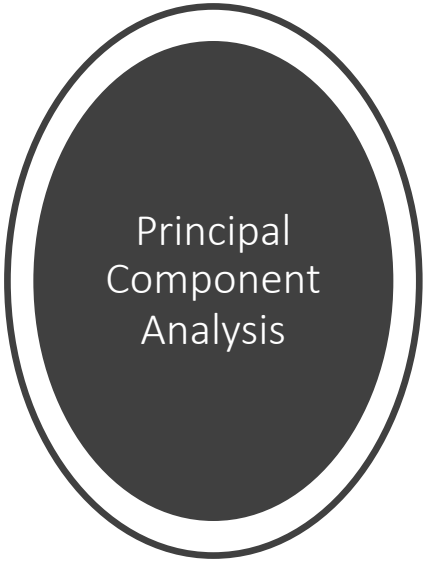
- For example, A 28 X 28 image has 784 picture elements (pixels) that are the dimensions or features which together represent that image.

https://www.datacamp.com/community/tutorials/principal-component-analysis-in-python



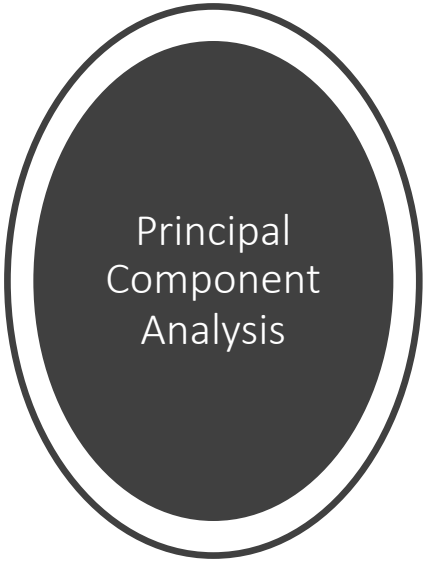
- PCA is that it is an **Unsupervised** dimensionality reduction technique, so, you can cluster the similar data points based on the feature correlation between them without any supervision (or labels).
- PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called **principal components**.
- Note: Features, Dimensions, and Variables are all referring to the same thing. You will find them being used interchangeably.

https://www.datacamp.com/community/tutorials/principal-component-analysis-in-python



- **Principal components** are the key to PCA; they represent what's underneath the hood of your data.
- When the data is projected into a lower dimension (assume three dimensions) from a higher space, the three dimensions are nothing but the three Principal Components that captures (or holds) most of the variance (information) of your data.
- Principal components have both direction (eigenvectors) and magnitude (eigenvalues).
- The direction represents across which principal axes the data is mostly spread out or has most variance.
- The magnitude signifies the amount of variance that Principal Component captures of the data when projected onto that axis.

https://www.datacamp.com/community/tutorials/principal-component-analysis-in-python



- The principal components are a straight line, and the first principal component holds the most variance in the data.
- Each subsequent principal component is orthogonal to the last and has a lesser variance. In this way, given a set of x correlated variables over y samples you achieve a set of u uncorrelated principal components over the same y samples.
- The reason you achieve uncorrelated principal components from the original features is that the correlated features contribute to the same principal component, thereby reducing the original data features into uncorrelated principal components; each representing a different set of correlated features with different amounts of variation.
- Each principal component represents a percentage of total variation captured from the data.

https://www.datacamp.com/community/tutorials/principal-component-analysis-in-python

Hands-on Exercises with Python

PCA WITH THE BREAST CANCER DATASET

DOWNLOAD PYTHON SCRIPT PCA-BREAST.PY

https://www.datacamp.com/community/tutorials/principal-component-analysis-in-python

Hands-on Exercises with Python

PCA WITH THE IRIS DATASET

DOWN LOAD PYTHON SCRIPT PCA-IRIS.PY

https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60

Hands-on Exercises with Python

PCA WITH THE DIGITS DATASET

DOWNLOAD PYTHON SCRIPT PCA-DIGITS

https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html

Hands-on Exercises with Python

PCA WITH THE FACES DATASET

DOWN LOAD PYTHON SCRIPT PCA-FACES.PY

https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html