Machine Learning – 04

**Principal component analysis (PCA)**

PCA is used to reduce the dimensionality of a dataset (see overfitting – lack of generalization), using a transformation that preserves the most variance in the data using **the least amount of dimentions**

It in fact is assumed that the information is carried in the **variance** of the features, meaning that:

**Higher variation ⇒ more information**

We aim to follow these steps:

1. Construct the **covariance matrix** (see previous document)
2. Compute its eigenvalues
3. Use the eigenvectors to reconstruct data