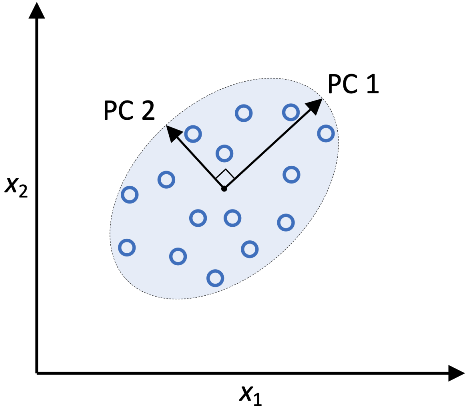
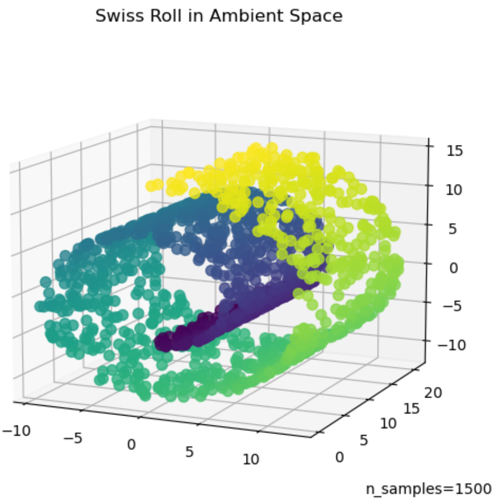
Principal Component Analysis

Marcin Kuta

Principal Component Analysis



Principal Component Analysis



Assumptions

- normality of data
- signal to noise ratio is meaningful

Full Singular Value Decompostion

$X \in \mathbb{R}^{n \times m}$
We assume X contains centered data.

$$X = \underset{n \times n}{U} \underset{n \times m}{\Sigma} \underset{m \times m}{V^T} \tag{1}$$

$U \in \mathbb{R}^{n \times n}$
$\Sigma \in \mathbb{R}^{n \times m}$
$V \in \mathbb{R}^{m \times m}$
$U^T U = U U^T = I_n$
$V^T V = V V^T = I_m$

Reduced Singular Value Decompostion

$X \in \mathbb{R}^{n \times m}$

$$X = \underset{n \times m}{U} \underset{m \times m}{\Sigma} \underset{m \times m}{V^T} \tag{2}$$

$U \in \mathbb{R}^{n \times m}$
$\Sigma \in \mathbb{R}^{m \times m}$
$V \in \mathbb{R}^{m \times m}$
$U^T U = I_m, \; UU^T = I_n$
$V^T V = VV^T = I_m$

Truncated Singular Value Decompostion

$X \in \mathbb{R}^{n \times m}$
$k < m$

$$X_k = \underset{n \times k}{U_k} \underset{k \times k}{\Sigma_k} \underset{k \times k}{V_k^T} \tag{3}$$

$U \in \mathbb{R}^{n \times k}$
$\Sigma \in \mathbb{R}^{k \times k}$
$V \in \mathbb{R}^{k \times k}$
$U_k^T U_k = I_k, \;\; U_k U_k^T = I_n$
$V_k^T V_k = I_k$

Principal Components

V - Columns of V are the *principal axes* of X

Principal Components (scores):

$$\underset{n\times m}{Z} = \underset{n\times m}{X} \underset{m\times m}{V} \tag{4}$$

$$Z = U\Sigma V^T V = U\Sigma \tag{5}$$

Z - coordinates of data in the new basis

$$X = ZV^T \tag{6}$$

Principal Component Analysis Learning

Learn the PCA model

Data: Training data $\mathcal{T} = \{\mathbf{x}_i\}_{i=1}^n$

Result: Principal axes $\mathbf{V}$ and scores $\mathbf{Z}_0$

1. Compute the mean vector $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$
2. Center the data, $\mathbf{x}_{0,i} = \mathbf{x}_i - \bar{\mathbf{x}}$, for $i = 1, \ldots, n$
3. Construct the data matrix $\mathbf{X}_0$ according to (10.36)
4. Perform SVD on $\mathbf{X}_0$ to obtain the factorization $\mathbf{X}_0 = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$
5. Compute principal components $\mathbf{Z}_0 = \mathbf{U}\mathbf{\Sigma}$

Method 10.5: Principal component analysis

Sample Covariance Matrix

Covariance matrix

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{n}X^TX = \tag{7}$$

$$= \frac{1}{n}V\Sigma^TU^TU\Sigma V^T = V(\frac{1}{n}\Sigma^T\Sigma)V^T = V\Lambda V^T$$

$\Lambda_{ii} = \sigma_i^2/n$

Variance Explained

Covariance matrix

$$\frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_m}, k \in 1, \ldots, m \tag{8}$$

Variants of PCA

- Kernel PCA
- Incremental PCA
- Randomized PCA
  - PCA: $O(nm^2 + m^3)$
  - Randomized PCA: $O(nd^2 + d^3)$

References

[1] Andreas Lindholm, Niklas Wahlström, Fredrik Lindsten, Thomas B. Schön
Machine Learning. A First Course for Engineers and Scientists, 2021

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

[3] https://github.com/rasbt/machine-learning-book/tree/main/ch05