Subspace Methods

Shadi Albarqouni, M.Sc.
Graduate Research Assistant | PhD Candidate
shadi.albarqouni@tum.de

Outline

What we need subspace methods?

Data lives in high dimensional space

Notation

- $\mathcal{X}^T = \{x_1, x_2, ..., x_N\}^T \in \mathbb{R}^{d \times N}$ is the data set.
- d is the feature dimension of x_i.
- N is the number of instances.

Objective

Find a subspace that maximizing the variance among the data.

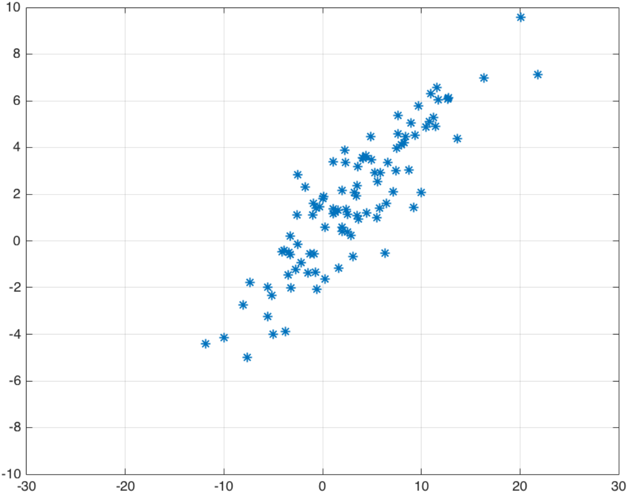
Principle Component Analysis(PCA) I

Objective: find a subspace that maximize the variance/covariance among the point cloud. Then, the point cloud $\mathcal{X}^T \in \mathbb{R}^{d\times N}$ can be projected to a lower dimensional space (subspace), i.e. $\mathcal{X}_{proj}^T \in \mathbb{R}^{r\times N}$, where $r < d$.

$$\begin{aligned}
\hat{v} &= \operatorname{argmax} \|Xv\|^2, \quad \text{s.t.} \quad \|v\| = 1, \\
&= \operatorname{argmax}(v^T X^T X v), \quad \text{s.t.} \quad v^T v = 1, \\
&= \operatorname{argmax} \frac{v^T X^T X v}{v^T v},
\end{aligned} \tag{1}$$

where $\hat{v}$ is the first principle component, i.e. eigenvector or loadings.

Principle Component Analysis(PCA) II



Principle Component Analysis(PCA) III

Singular Value Decomposition (SVD)[1]

Given a data matrix $X \in \mathbb{R}^{N \times d}$, where N is the number of samples (observations) and d is the feature dimension, the singular value decomposition (SVD) can be computed as follows:

$$X = U \Sigma V^T, \tag{2}$$

where $U \in \mathbb{R}^{N \times N}$ is the left-singular vectors, the diagonal elements of $\Sigma \in \mathbb{R}^{N \times d}$ are the singular values, and $V \in \mathbb{R}^{d \times d}$ is the right-singular vector. The eigenvectors are the same as the right-singular vector, where the eigenvalues are the diagonal elements of $\Sigma^T \Sigma$.

Principle Component Analysis(PCA) IV

Eigen-decomposition of Covariance Matrix

Given a covariance matrix $C \in \mathbb{R}^{d \times d}$, which can be computed from the data matrix, i.e. $C = X^T X$, the eigenvectors and eigenvalues can be computed as follows:

$$CV = \Lambda V, \tag{3}$$

where $V \in \mathbb{R}^{d \times d}$ is the eigenvectors matrix and the diagonal elements of $\Lambda \in \mathbb{R}^{d \times d}$ represent the eigenvalues.

Principle Component Analysis(PCA) V

Connection between Covariance and SVD

Let's start from Eq.(3), and substitute C with $X^T X$ as follows:

$$\begin{aligned} CV &= \Lambda V, \\ C &= V \Lambda V^T, \\ X^T X &= V \Lambda V^T, \\ (U \Sigma V^T)^T U \Sigma V^T &= V \Lambda V^T, \\ V \Sigma^T \Sigma V^T &= V \Lambda V^T, \end{aligned} \tag{4}$$

where $U^T U = V^T V = I$, and $\Sigma^T \Sigma = \Lambda$.

It should be noted that data matrix X has column zero mean (features) and the projected data can be obtained by $X_{proj}^T = V^T X^T$. Note: To get consistent results from SVD and Covariance, i.e. for SVD: divide X^T by the $\sqrt{(N-1)}$, COV: divide the $X^T X$ by the $(N-1)$.

[1] Nice submission from our chair to the MICCAI Educational Initiative, https://www.youtube.com/watch?v=7hqDUxAgaFU

Variations

- Kernel PCA
- Linear Discriminative Analysis
- Independent Component Analysis
- Laplacian Eigenmap

Applications

Eigenfaces
Statistical Shape Model (SSM)
Appearance Shape Model (ASM)
Regularization

The next slides from Fausto's lecture in WS 15/16.

References I

Christopher M Bishop.
Pattern recognition.
Machine Learning, 2006.

Ian Jolliffe.
Principal component analysis.
Wiley Online Library, 2002.