Reducing Dimensionality

Principal Component Analysis

PCA which assuming the original data set has been drawn from a Gaussian distribution allows to obtain an ordered list of components that account for the largest amount of the variance from the data in terms of least square errors: The amount of variance captured by the first component is larger than the amount of variance on the second component and so on.

Consider a data matrix, $\boldsymbol{X}$, with column-wise *zero empirical mean* (the sample mean of each column has been shifted to zero), where each of the n rows represents a different repetition of the experiment, and each of the p columns gives a particular kind of feature. Mathematically, the transformation is defined by a set of p-dimensional vectors of weights or loadings $\boldsymbol{w}_{(k)} = (w_1, \cdots, w_p)_{(k)}$ that map each row vector $\boldsymbol{x}_{(i)}$ of $\boldsymbol{X}$ to a new vector of principal component scores given by $\boldsymbol{t}_{(i)} = (t_1, \cdots, t_m)_{(i)}$, given by

$$t_{k(i)} = \boldsymbol{x}_{(i)} \cdot \boldsymbol{w}_{(k)}$$

where each loading vector $\boldsymbol{w}$ constrained to be a *unit vector*.

First component

The first loading vector satisfies:

$$\boldsymbol{w}_{(1)} = \arg\max \left\{ \frac{\boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{w}} \right\}$$

Further components

The k^{th} component can be found by subtracting the first $k-1$ principal components from $\boldsymbol{X}$:

$$\hat{\boldsymbol{X}}_k = \boldsymbol{X} - \sum_{s=1}^{k-1} \boldsymbol{X} \boldsymbol{w}_{(s)} \boldsymbol{w}_{(s)}^T$$

and then finding the loading vector which extracts the maximum variance from this new data matrix

$$\boldsymbol{w}_{(k)} = \arg\max \left\{ \frac{\boldsymbol{w}^T \hat{\boldsymbol{X}}_k^T \hat{\boldsymbol{X}}_k \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{w}} \right\}$$

The full principal components decomposition of $\boldsymbol{X}$ can therefore be given as

$$\boldsymbol{T} = \boldsymbol{X} \boldsymbol{W}$$

where $\boldsymbol{W}$ is a p-by-p matrix whose columns are the eigenvectors of $\boldsymbol{X}^T \boldsymbol{X}$. The transpose of $\boldsymbol{W}$ is sometimes called the *whitening* or *sphering transformation.*

Dimensionality reduction

Keeping only the first L principal components, produced by using only the first L loading vectors, gives the truncated transformation

$$\boldsymbol{T}_L = \boldsymbol{X}\boldsymbol{W}_L$$

where the matrix $\boldsymbol{T}_L$ now has n rows but only L columns.

Singular Value Decomposition

In linear algebra, the singular-value decomposition (SVD) is a factorization of a real or complex matrix. It is the generalization of the eigendecomposition of a *positive semidefinite normal matrix* (for example, a symmetric matrix with positive eigenvalues) to any $m \times n$ matrix via an extension of the polar decomposition

$$\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\dagger$$

with the properties:

- The left-singular vectors of $\boldsymbol{M}$ are a set of orthonormal eigenvectors of $\boldsymbol{M}\boldsymbol{M}^\dagger$.
- The right-singular vectors of $\boldsymbol{M}$ are a set of orthonormal eigenvectors of $\boldsymbol{M}^\dagger\boldsymbol{M}$.
- The non-zero singular values of $\boldsymbol{M}$ (found on the diagonal entries of $\boldsymbol{\Sigma}$) are the square roots of the non-zero eigenvalues of both $\boldsymbol{M}^\dagger\boldsymbol{M}$ and $\boldsymbol{M}\boldsymbol{M}^\dagger$.

It has many useful applications in signal processing and statistics.