Dimensionality Reduction

1. Background

In many fields of applications, we have to collect a huge amount of data with multi–variables, which is called high–dimensional dataset. Some of these variables may note be important or relative to our analysis. Some may have a lot of noises. In order to speed the analysis and get rid of useless information, we have to reduce dimension of dataset.

2. Dimensionality Reduction Methods

> ***2.1 Feature Selection***
>
> *Find major variables.*
>
> - *Lasso*
> - *Elastic Net*
>
> ***2.2 Linear Dimensionality Reduction***
>
> *Variables are always linear correlated, transform data from the high–dimensional space to lower–dimensional space through projection.*
>
> - *PCA*
> - *Kernel PCA*
> - *LDA*
> - *MDS*
>
> ***2.3 Non-linear Dimensionality Reduction***
>
> - *Manifold learning*

3. Matrix Form

> ***3.1 Data in matrix form***
>
> *Normally, a single sample is a **column vector**.*

$$X = (x_1, x_2, \ldots, x_N)^T_{N\times P},\ x_i \in \mathbb{R}^P,\ i = 1, 2, \ldots, N$$

$$= \begin{pmatrix} x_1^T \\ \\ . \\ \\ . \\ \\ . \\ \\ x_N^T \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & . & . & x_{1P} \\ \\ x_{21} & . & . & . & . \\ \\ . & . & . & . & . \\ \\ . & . & . & . & . \\ \\ x_{N1} & . & . & . & x_{NP} \end{pmatrix}_{N\times P}$$

3.2 Sample Mean Matrix

We define a vector called $\mathbb{1}_N$*:*

$$\mathbb{1}_N = \begin{pmatrix} 1 \\ \\ 1 \\ \\ \ldots \\ \\ 1 \end{pmatrix}_{N\times 1}$$

Iterate all samples to get P *means:*

$$\bar{X}_{P\times 1} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$= \frac{1}{N}\underbrace{\begin{pmatrix} x_1 & x_2 & . & . & . & x_N \end{pmatrix}}_{X^T}\begin{pmatrix} 1 \\ \\ 1 \\ \\ \ldots \\ \\ 1 \end{pmatrix}_{N\times 1}$$

$$= \frac{1}{N}X^T\mathbb{1}_N$$

3.3 Sample Covariance Matrix

In 1–D space, the variance is:

$$S = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

While in higher dimension, the covariance matrix is:

$$S_{P\times P} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T$$

$$= \frac{1}{N}\begin{pmatrix} x_1 - \bar{x} & x_2 - \bar{x} & \dots & x_N - \bar{x} \end{pmatrix}\begin{pmatrix} (x_1 - \bar{x})^T \\ (x_2 - \bar{x})^T \\ \dots \\ (x_N - \bar{x})^T \end{pmatrix}$$

$$A.\ \begin{pmatrix} x_1 - \bar{x} & x_2 - \bar{x} & \dots & x_N - \bar{x} \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \dots & x_N \end{pmatrix} - \begin{pmatrix} \bar{x} & \bar{x} & \dots & \bar{x} \end{pmatrix}$$

$$= X^T - \bar{X}\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}$$

$$= X^T - \bar{X}\mathbb{1}_N^T$$

$$= X^T - \frac{1}{N}X^T\mathbb{1}_N\mathbb{1}_N^T$$

$$= X^T(I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T)$$

$$B.\ \begin{pmatrix} (x_1 - \bar{x})^T \\ (x_2 - \bar{x})^T \\ \dots \\ (x_N - \bar{x})^T \end{pmatrix} = \begin{pmatrix} x_1^T \\ x_2^T \\ \dots \\ x_N^T \end{pmatrix} - \begin{pmatrix} \bar{x}^T \\ \bar{x}^T \\ \dots \\ \bar{x}^T \end{pmatrix}$$

$$= (X^T(I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T))^T$$

$$= (I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T)^T X$$

$$Set\ Centering\ Matrix\ H_N = I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T,$$

$$
\begin{aligned}
Then\ S_{P\times P} &= \frac{1}{N}AB \\
&= \frac{1}{N}X^T(I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T)(I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T)^T X \\
&= \frac{1}{N}X^T H_N H_N^T X \\
&= \frac{1}{N}X^T H X
\end{aligned}
$$

3.4 Centering Matrix

Zero-centering, subtract from mean.

$$
\begin{aligned}
H_N &= H_N^T = I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T \\
H_N^2 &= H^T \cdot H \\
&= (I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T)(I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T) \\
&= I_N - \frac{2}{N}\mathbb{1}_N\mathbb{1}_N^T - \frac{1}{N^2}\mathbb{1}_N\mathbb{1}_N^T\mathbb{1}_N\mathbb{1}_N^T \\
&= I_N - \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^T \\
&= H \\
H^n &= H
\end{aligned}
$$