PCA A data-driven features extraction method

1. Data Presentation

Data: $X = (x_1, x_2, \dots, x_n)^T_{N\times p} = \begin{pmatrix} \boxed{x_{11} & x_{12} & \cdots & x_{1p}} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}_{N\times p}$ $\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{matrix}$

$x_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix}_{p\times 1}$

Sample Mean: $\bar{X}_{p\times 1} = \frac{1}{N}\sum_i^N x_i = \frac{1}{N} X^T 1_N$ $\begin{bmatrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1p} & \cdots & x_{np} \end{bmatrix}_{p\times n} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1}$

$x_1 \cdots x_n$

Sample Covariance: $S_{p\times p} = \frac{1}{N}\cdot\sum_i^N (x_i - \bar{x})\cdot(x_i - \bar{x})^T$

$= \frac{1}{N}\cdot \boxed{(x_1 - \bar{x} \quad x_2 - \bar{x} \quad \cdots \quad x_n - \bar{x})} \cdot \begin{pmatrix} (x_1 - \bar{x})^T \\ (x_2 - \bar{x})^T \\ \vdots \\ (x_n - \bar{x})^T \end{pmatrix}$

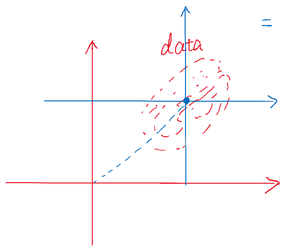
$= \frac{1}{N}\cdot \boxed{(x_1 \quad x_2 \quad \cdots \quad x_n) - (\bar{x} \quad \bar{x} \quad \cdots \quad \bar{x})}$

$= \frac{1}{N}\cdot \left(\underset{p\times N}{X^T} - \underset{p\times 1}{\bar{x}} \cdot \underset{1\times N}{1_N^T}\right) \cdot \boxed{\left(X^T - \bar{x}\cdot 1_N^T\right)^T}$

$\bar{x} = \frac{1}{N}X^T 1_N$ $\longrightarrow$ $= \frac{1}{N}\cdot(X^T - \bar{x}\cdot 1_N^T)\cdot(X - 1_N\cdot \bar{x}^T)$

$= \frac{1}{N}\cdot\left(\underline{X^T} - \frac{1}{N}\cdot \underline{X^T}\cdot 1_N \cdot 1_N^T\right)\cdot\left(X - 1_N\cdot\frac{1}{N}\cdot 1_N^T \cdot X\right)$

$= \frac{1}{N}\cdot X^T \underbrace{\left(I_N - \frac{1}{N}\cdot 1_N\cdot 1_N^T\right)}\cdot\left(I_N - \frac{1}{N}\cdot 1_N\cdot 1_N^T\right)^T\cdot X$



$\begin{bmatrix} 1-\frac{1}{N} & 1-\frac{1}{N} & \cdots & 1-\frac{1}{N} \\ \vdots & & \ddots & \vdots \\ 1-\frac{1}{N} & \cdots & & 1-\frac{1}{N} \end{bmatrix}_{N\times N}$ $\Rightarrow$ centering matrix H

$= \frac{1}{N}\cdot X^T H\cdot H^T\cdot X$ $(H^n = H, \; H^T = H)$

$= \frac{1}{N}\cdot X^T\cdot H\cdot X$

$H^2 = (I_N - \frac{1}{N}\cdot 1_N\cdot 1_N^T)\cdot(I_N - \frac{1}{N}\cdot 1_N\cdot 1_N^T)$

$= I_N - \frac{2}{N} 1_N\cdot 1_N^T + \frac{1}{N^2}\cdot 1_N\cdot 1_N^T\cdot 1_N\cdot 1_N^T$

$= I_N - \frac{2}{N} 1_N\cdot 1_N^T + \frac{1}{N^2}\cdot N\cdot 1_N\cdot 1_N^T$

$= I_N - \frac{1}{N} 1_N\cdot 1_N^T$

$= H$

Sample Mean: $\bar{x} = \frac{1}{N}\cdot X\cdot 1_N$

Sample Covariance: $S = \frac{1}{N}\cdot X^T\cdot H\cdot X$

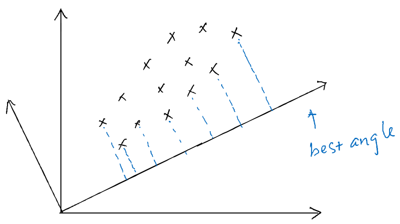
2. The logic of PCA

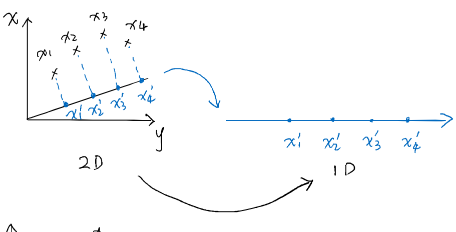
The original data dimension is p. We can lower p by removing the relevant dimensions and leave the irrelevant ones.

how? $\Big($

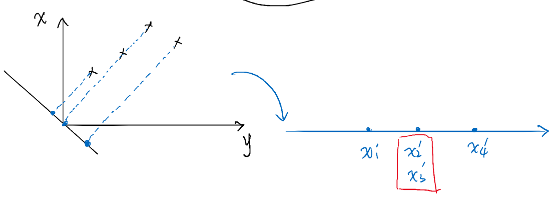
$\left\{\begin{array}{l} \text{maximize the projection variance} \\ \qquad\qquad \text{||} \\ \text{minimize the reconstruction cost} \end{array}\right.$

why? $\Big($



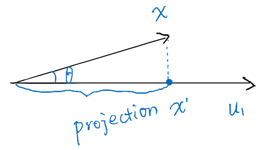


(good projection)



(not good enough)

Maximize Porjection Variance



let $\|u_1\| = 1$

$$x \cdot u_1 = \|x\| \cdot \|u_1\| \cdot \cos\theta$$
$$= \|x\| \cdot \cos\theta$$
$$= \|x'\|$$

$$J(u_1) = \frac{1}{N}\sum_{i}^{N}\left(\boxed{(x_i - \bar{x}) \cdot u_1}\right)^2$$ real number

$$= \frac{1}{N}\sum_{i}^{N}\left((x_i - \bar{x})\cdot u_1\right)^T\left((x_i - \bar{x})\cdot u_1\right)$$

$$= \frac{1}{N}\sum_{i}^{N} u_1^T \boxed{(x_i - \bar{x})^T \quad (x_i - \bar{x})} \cdot u_1$$ real number

$$= u_1^T \left(\frac{1}{N}\sum_{i}^{N}(x_i - \bar{x})^T(x_i - \bar{x})\right)\cdot u_1$$

$$= u_1^T \cdot S \cdot u_1 \qquad (\text{s.t. } u_1^T \cdot u_1 = 1)$$

$$\begin{cases} \underset{u_1}{\operatorname{argmax}}\ u_1^T S u_1 \\ u_1^T \cdot u_1 = 1 \end{cases}$$

Largrange : $\mathcal{L}(u_1, \lambda) = u_1^T S u_1 + \lambda(1 - u_1^T u_1)$

$$\frac{\partial \mathcal{L}}{\partial u_1} = 2 S u_1 - 2\lambda u_1 \stackrel{\Delta}{=} 0$$

$$S u_1 = \lambda u_1$$

eigen value

eigenvector

$$\underset{u_1}{\operatorname{argmax}}\ u_1^T S u_1 = \underset{u_1}{\operatorname{argmax}}\ u_1^T \lambda u_1$$
$$= \underset{u_1}{\operatorname{argmax}}\ \lambda u_1^T u_1$$
$$= \underset{u_1}{\operatorname{argmax}}\ \lambda$$

find the largest eigenvalue (principle component)

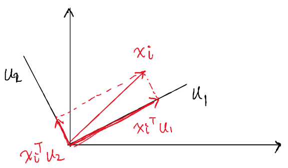
Minimize reconstruction cost

$$x_i \in \mathbb{R}^p, \quad x_i' = \sum_{k=1}^{p}(x_i^T u_k) u_k$$

reduce dimension

$$\hat{x}_i \in \mathbb{R}^q, \quad \hat{x}_i = \boxed{\sum_{k=1}^{q}}(x_i^T u_k) u_k$$

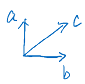
only use q basis, information may be lost



minimize reconstruction cost = minimize the information loss

$$J = \frac{1}{N}\sum_{i=1}^{N}\|x_i' - \hat{x}_i\|^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left\|\sum_{k=q+1}^{p}(x_i^T u_k) u_k\right\|^2$$



$\|\vec{a} + \vec{b}\|^2 = \|\vec{c}\|^2$

$\Updownarrow$

$a^2 + b^2 = c^2$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=q+1}^{p}(x_i^T u_k)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=q+1}^{p}\left((x_i - \bar{x})^T u_k\right)^2$$

$$= \sum_{k=q+1}^{p}\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^T u_k\right)^2$$ $u_k^T S u_k$

$$= \sum_{k=q+1}^{p} u_k^T S u_k$$

$$\begin{cases} \underset{u_k}{\operatorname{argmin}}\ \sum_{k=q+1}^{p} u_k^T S u_k \\ \text{s.t. } u_k^T u_k = 1 \end{cases}$$

Largrange :

$$S u_k = \lambda u_k$$

$$\underset{u_k}{\operatorname{argmin}}\sum_{k=q+1}^{p} u_k^T S u_k = \underset{u_k}{\operatorname{argmin}}\sum_{k=q+1}^{p} u_k^T \lambda u_k$$

$$= \underset{u_k}{\operatorname{argmin}}\sum_{k=q+1}^{p} \lambda u_k^T u_k$$ $= 1$