回顾：多重共线性，即 $\exists\ \lambda_{min}(X^TX)$ 非常接近于 0

一、主成分分析

Principle Component Analysis

对原始线性模型：

$$E(y) = X\beta \qquad cov(y) = \sigma^2 I$$

其中 $X \in R^{n\times p}$ $\quad \beta \in \mathbb{R}^p$

记 $\underline{X^TX = \Phi\Lambda\Phi^T}$，其中 $\Phi = [\phi_1, \phi_2, \cdots, \phi_p] \in \mathbb{R}^{p\times p}$ 为正交阵，而 $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_p)$ 满足 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$。令

$$Z = X\Phi \qquad \alpha = \Phi^T\beta$$

于是有主成分分析下的模型：

$$E(y) = Z\alpha \qquad cov(y) = \sigma^2 I$$

其中 Designed Matrix：$Z = (Z_{(1)}, Z_{(2)}, \cdots, Z_{(p)}) = (X\phi_1, X\phi_2, \cdots, X\phi_p)$

我们称由 X 列向量线性组合的 $\{Z_{(j)} : j = 1, 2, \cdots, p\}$ 为主成分。

其中对应 X^TX 第 j 大的特征值 λ_j 的 $Z_{(j)}$ 称为第 j 主成分。

如此，我们可对 $E(y) = Z\alpha$ 使用回归

Remark：若假设自变量已中心标准化，即 $1^TX = 0$

$\Rightarrow$ $1^TZ = 1^TX\Phi = 0$，即 Z 各列平均值 $\bar{Z}_{(j)} = 0$

$$\Rightarrow \quad \sum_{i=1}^{n}(z_{ij}-\bar{z}_{(j)})^2 = z_{(j)}^T z_{(j)} = \phi_j^T X^T X \phi_j = \phi_j^T \Phi \Lambda \Phi^T \phi_j = \lambda_j$$

说明：X^TX 的特征值 λ_j 衡量了主成分 $z_{(j)}$ 的波动大小。若 λ_j 比较小，可以考虑删去 $z_{(j)}$ 后进行回归

二、参数估计

Parameters Estimation of PCA

将下列矩阵写成分块形式：

$$\Lambda = \mathrm{diag}(\Lambda_1, \Lambda_2)$$

$$\alpha = (\alpha_1^T, \alpha_2^T)^T$$

$$Z = (Z_1, Z_2)$$

$$\Phi = (\Phi_1, \Phi_2)$$

其中： $\Lambda_1 \in \mathbb{R}^{r\times r}$ $\quad \alpha_1 \in \mathbb{R}^r$ $\quad Z_1 \in \mathbb{R}^{n\times r}$ $\quad \Phi_1 \in \mathbb{R}^{p\times r}$

剔除后 $(p-r)$ 个非主成分，得到 r 个主成分模型：

$$E(y) \approx Z_1\alpha_1 \qquad \mathrm{cov}(y) = \sigma^2 I$$

直接使用 OLSE 得：

$$\begin{aligned}\hat{\alpha}_1 &= (Z_1^T Z_1)^{-1} Z_1^T y \\ &= (\Phi_1^T X^T X \Phi_1)^{-1} Z_1^T y = (\Phi_1^T \Phi \Lambda \Phi^T \Phi_1)^{-1} Z_1^T y \\ &= \Lambda_1^{-1} Z_1^T y\end{aligned}$$

认为 $\hat{\alpha}_2 = 0$，于是得到 β 的主成分估计：

$$\tilde{\beta} = \Phi(\hat{\alpha}_1^T, \hat{\alpha}_2^T)^T = \Phi_1 \Lambda_1^{-1} \Phi_1^T X^T y$$

$Z=(Z_1\ Z_2)=X\Phi$
$=(X\Phi_1,\ X\Phi_2)$

$\alpha=\Phi^T\beta \Rightarrow \bar\beta=\Phi\alpha=\Phi\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}=\Phi_1\alpha_1+\Phi_2\alpha_2=\Phi_1\Lambda_1^{-1}Z_1^Ty$

三、主成分估计的性质

Properties of PCA estimate

有偏，但减小了方差，一定条件下优化了 MSE

3.1. 有偏性

$E(\bar\beta)=E(\Phi_1\Lambda_1^{-1}Z_1^Ty)=\Phi_1\Lambda_1^{-1}Z_1^T\cdot Z_1\alpha_1$
$=\Phi_1\alpha_1$

$$E(\bar\beta)=\Phi_1\alpha_1$$

若 $\alpha_2\neq 0$，则 $E(\bar\beta)\neq\beta$

3.2. MSE 改进

$\mathrm{cov}(\bar\beta)=\mathrm{cov}(\Phi_1\Lambda_1^{-1}Z_1^Ty)=\sigma^2\Phi_1\Lambda_1^{-1}Z_1^TZ_1\Lambda_1^{-1}\Phi_1^T$
$=\sigma^2\Phi_1\Lambda_1^{-1}\Phi_1^T$

$\mathrm{tr}(\Phi_1\Lambda_1^{-1}\Phi_1^T)$
$=\mathrm{tr}(\Lambda_1^{-1})$

若 $\|\alpha_2\|^2<\sigma^2\sum_{j=r+1}^{p}\lambda_j^{-1}$

则 $MSE(\bar\beta)<MSE(\hat\beta)$

其中 $\hat\beta=(X^TX)^{-1}X^Ty$

$MSE(\hat\beta)=\mathrm{tr}\{\mathrm{cov}(\hat\beta)\}$
$=\mathrm{tr}\{\mathrm{cov}((X^TX)^{-1}X^Ty)\}=\mathrm{tr}\{(X^TX)^{-1}X^T\cdot\sigma^2 I\cdot X(X^TX)^{-1}\}$
$=\mathrm{tr}\{(X^TX)^{-1}\}\sigma^2=\sigma^2\mathrm{tr}(\Lambda^{-1})$

Remark：由于 $\alpha_2=\Phi_2^T\beta$，故条件 $\|\alpha_2\|^2<\sigma^2\sum_{j=r+1}^{p}\lambda_j^{-1}$ 可写作：

$$\left\|\frac{\alpha_2}{\sigma}\right\|^2=\left(\frac{\beta}{\sigma}\right)^T\Phi_2\Phi_2^T\left(\frac{\beta}{\sigma}\right)<\mathrm{tr}(\Lambda_2^{-1})$$

直观地理解，PCA 在 MSE 上优于 OLSE 的条件：

(a). 固定 β 和 σ^2，矩阵 X^TX 的后 $(p-r)$ 个特征值足够小

(b). 固定 Λ_2 和 Φ_2（即固定 X^TX），模长 $\|\frac{\beta}{\sigma}\|$ 足够小，即模型信噪比 $\beta^TX^TX\beta/\sigma^2$ 足够低，自身波动太大

四、主成分分析应用

Implementation of PCA

4.1、预设贡献率

使 $\dfrac{\sum_{j=1}^{r}\lambda_j}{\sum_{i=1}^{p}\lambda_i} > rate$ 的 r

rate 为预设水平，如 75%，80% ···

4.2、可解释性

$Z = X\Phi$，为 X 的线性组合，欲解释其含义。

⇒ 转化回原本 $X\beta$ 的形式.

补：

$$\mathrm{diag}(\Lambda_1, \Lambda_2) = \Lambda = \Phi^T\Phi\Lambda\Phi^T\Phi = \Phi^T X^T X\Phi = Z^T Z = \begin{bmatrix} z_1^T z_1 & z_1^T z_2 \\ z_2^T z_1 & z_2^T z_2 \end{bmatrix}$$

$$= \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}$$