General linear model

$\underset{n\times d}{Y} = \underset{n\times m}{X}\,\underset{m\times d}{G} + \underset{n\times d}{U}$ All matrices

kronecker product $\otimes$

vec operation

$$\text{vec}\left(\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

$$\text{Vec}(Y) = (I \otimes X)\,\text{vec}(G) + \text{vec}(U)$$

Multiple linear regression

$$\mathbb{R}^n = \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}}_{1} n = \mathbb{R}^{n\times 1}$$

$$\underset{\mathbb{R}^n}{\vec{Y}} = \underset{\mathbb{R}^{n\times m}}{X}\,\underset{\mathbb{R}^{m\times 1}}{\vec{G}} + \underset{\mathbb{R}^{n\times 1}}{\vec{U}}$$

input X, output $\vec{Y}$, unknow $\vec{G}$

Homoskedasticity $\vec{U} \sim N(0, \sigma^2 I_n)$ $\underbrace{\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}}_{n} n$

"classical" setting: $n >> m$

"big data"

reasonable to assume $\text{rank}(X) = m$

$\Leftrightarrow (X^T X)^{-1}$ exists

$X^T\vec{Y} = X^T X\vec{G} + X^T\vec{U}$

$\hat{G} = (X^TX)^{-1}X^T\vec{Y}$ if $\mathbb{E}\vec{U}=0$, then $\mathbb{E}\hat{G} = \vec{G}$

$\mathbb{E}(\hat{G}-\vec{G})(\hat{G}-\vec{G})^T = \sigma^2 (X^TX)^{-1}$ → estimation error

$\mathbb{R}^{m\times 1}$ $\mathbb{R}^{1\times m}$

$\mathbb{E}\,|\hat{G}-\vec{G}|^2 = \mathbb{E}(\hat{G}-\vec{G})^T(\hat{G}-\vec{G}) = \sigma^2\,\mathrm{trace}\,(X^TX)^{-1}$

$SSE := (\vec{Y}^T - \hat{G}^T X^T)\vec{Y}$

$\overset{(\triangle)}{=} \vec{U}^T(I_n - P)\vec{U}$

$P = X(X^TX)^{-1}X^T$

$PX = X$, $P^T = P$ orthogonal projection

$PP = P$ P is 0 or 1 $\lambda^2 = \lambda$

Thm If $A = A^T$, then $A\vec{v} = \lambda\vec{v}$ → $\lambda \in \mathbb{R}$

$\vec{v}$ forms a basis

if $\lambda_1 \neq \lambda_2$, then $\vec{v}_1 \cdot \vec{v}_2 = 0$

and $\exists Q$, Q^TQ

s.t. $Q^TAQ = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

if $\vec{U} \sim N(0, \sigma^2 I)$

and $Q^TQ = I$

$QU \sim N(0, \sigma^2 I)$