Lecture 14

§1 Variables 与 error terms 的矩阵表示

1. x_i 与 y_i 的矩阵表示

令 $\{x_1, \cdots, x_n\}$ 与 $\{y_1, \cdots, y_n\}$ 为两组 random variables，则其可表示为

① column random vector $\mathbf{x}$：

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

② column random vector $\mathbf{y}$：

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

2、对 random vector 的 operations

① Expectation

$$E(\mathbf{x}) = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_n) \end{bmatrix}, \quad E(\mathbf{x}+\mathbf{y}) = E(\mathbf{x}) + E(\mathbf{y})$$

② Constant (matrix) scaling

$$a\mathbf{x} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix}, \quad E(a\mathbf{x}) = aE(\mathbf{x}), \quad E(A\mathbf{x}) = AE(\mathbf{x})$$

③ Transpose

$$\mathbf{x}^T \text{ (或 } \mathbf{x}') = [x_1, x_2, \cdots, x_n]$$

$$\mathbf{a}^T\mathbf{x} = a_1x_1 + \cdots + a_nx_n, \quad E(\mathbf{a}^T\mathbf{x}) = \mathbf{a}^T E(\mathbf{x})$$

3. The covariance matrix

定义 $n\times n$ (variance -) covariance matrix 为：

$$Var(\mathbf{x}) = E[(\mathbf{x}-E(\mathbf{x}))(\mathbf{x}-E(\mathbf{x}))^T] = E[\mathbf{x}\mathbf{x}^T] - E[\mathbf{x}]E[\mathbf{x}^T]$$

$$= \begin{bmatrix} Var(x_1) & cov(x_1,x_2) & \cdots & cov(x_1,x_n) \\ cov(x_2,x_1) & Var(x_2) & \cdots & cov(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_n,x_1) & cov(x_n,x_2) & \cdots & Var(x_n) \end{bmatrix}$$

① ith diagonal element 为 $var(x_i)$

② $\{i,j\}$ th element 为 $cov(x_i, x_j)$

③ 由于 $cov(x_i,x_j) = cov(x_j,x_i)$，因此 variance - covariance matrix 为 symmetric

④ $var(A\mathbf{x}) = A \cdot var(\mathbf{x}) A^T$

证明：

$$Var(A\mathbf{x}) = E[(A\mathbf{x} - E[A\mathbf{x}])(A\mathbf{x} - E[A\mathbf{x}])^T]$$

$$= E[A\mathbf{x}\mathbf{x}^TA^T - \underbrace{E[A\mathbf{x}]}_{=AE[\mathbf{x}]}\mathbf{x}^TA^T - A\mathbf{x}E[\mathbf{x}^T]A^T + E[A\mathbf{x}]E[A\mathbf{x}]^T$$

$= A E[\mathbf{x}\mathbf{x}^T]A^T - AE[\mathbf{x}]E[\mathbf{x}^T]A^T - AE[\mathbf{x}]E[\mathbf{x}^T]A^T + AE[\mathbf{x}]E[\mathbf{x}^T]A^T$

$= A\underbrace{(E[\mathbf{x}\mathbf{x}^T] - E[\mathbf{x}]E[\mathbf{x}^T])}_{= Var(\mathbf{x})}A^T$

$= A \cdot var(\mathbf{x}) A^T$

4. The model error term e_i 的矩阵表示

线性回归模型中的 e_i's 可以写作

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

且有

$$E(\mathbf{e}) = \mathbf{0}, \quad var(\mathbf{e}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

上述表示刻画了 Gauss-Markov conditions，即

① $E(e_i) = 0$

② $var(e_i) = constant$

③ e_i's 之间 uncorrelated

§2 SLR 点估计量的矩阵表示

1. 基本参数的矩阵表示

定义以下 vectors 与 matrices：

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

可以将 $y_i = \beta_0 + \beta_1 X_i + e_i$ 写作

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{e}$$

且 Gauss-Markov conditions 可由 $E(\mathbf{e})$ 与 $var(\mathbf{e})$ 刻画

2. $\mathbf{y}$ 的分布

① Multivariate normal distribution

1° 联合概率密度函数

若 $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$，则

$$f(y_1, \cdots, y_n) = (2\pi)^{-\frac{n}{2}} \cdot |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y}-\boldsymbol{\mu})\right\},$$

其中 $\boldsymbol{\mu} \in R^n$, $\Sigma \in R^{n\times n}$ 且 symmetric & P.D.

2° multivariate normal distribution 的标准化

$$\Sigma^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu}) \sim N(0, I_n)$$

3° 关于 multivariate normal distribution 的重要结论

若 $\mathbf{x}, \mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$，则 $\mathbf{x}$ 与 $\mathbf{y}$ independent $\iff$ $Cov(\mathbf{x}, \mathbf{y}) = 0$

② $\mathbf{y}$ 的分布

1° $\mathbf{y}$ 的期望

$$E(\mathbf{y}) = E(X\boldsymbol{\beta}) + E(\mathbf{e}) = X\boldsymbol{\beta}$$

2° $\mathbf{y}$ 的方差

$$Var(\mathbf{y}) = Var(X\boldsymbol{\beta} + \mathbf{e}) = Var(\mathbf{e}) = \sigma^2 I$$

3° $\mathbf{y}$ 的分布

$$\mathbf{y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$$

3. 一些 gradient lemmas（此处均采用分母布局）

① 令 $\mathbf{c} = [c_1, \cdots, c_n]^T$，$f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} = \sum_{i=1}^{n} c_i x_i$，则

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{c}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{c}$$

② 令 $A \in R^{n\times n}$ 为 symmetric，$f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$，则

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^T A \mathbf{x}}{\partial \mathbf{x}} = 2A\mathbf{x}$$

注：$\dfrac{d(\mathbf{w}^T X \mathbf{w})}{d\mathbf{w}} = (X + X^T)\mathbf{w}$　（若 X 不为 symmetric）

③ 令 $A \in R^{n\times n}$，$f(\mathbf{x}) = A^T\mathbf{x}$，则

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial A^T\mathbf{x}}{\partial \mathbf{x}} = A$$

注：$\dfrac{d(\mathbf{y}^T X \mathbf{w})}{d\mathbf{w}} = X^T\mathbf{y}$

4. $\boldsymbol{\beta}$ 的求解

求出RSS：

$$\begin{aligned} RSS(\boldsymbol{\beta}) &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \\ &= (\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}) \\ &= \mathbf{y}^T\mathbf{y} - \boldsymbol{\beta}^T X^T \mathbf{y} - \mathbf{y}^T X \boldsymbol{\beta} + \boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \\ &= \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T X^T \mathbf{y} + \boldsymbol{\beta}^T X^T X \boldsymbol{\beta} \end{aligned}$$

令 RSS 对 $\boldsymbol{\beta}$ 求导得：

$$\frac{\partial RSS(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 - 2X^T\mathbf{y} + 2X^T X \mathbf{y} = 0$$

$$\Rightarrow X^T X \boldsymbol{\beta} = X^T \mathbf{y}$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$$　（若 $X^T X$ 可逆）

① 化简 X^TX

$$X^TX = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$= \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

② 化简 $(X^TX)^{-1}$

$$(X^TX)^{-1} = \frac{1}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix} \quad \left(利用 \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \right)$$

$$= \frac{1}{n S_{xx}} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$= \frac{1}{S_{xx}} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

③ X^TX 是否可逆？

(1) X^TX 可逆 $\iff$ $\det(X^TX) = 0$

$\iff$ X^TX 的 columns / rows 为 linearly independent

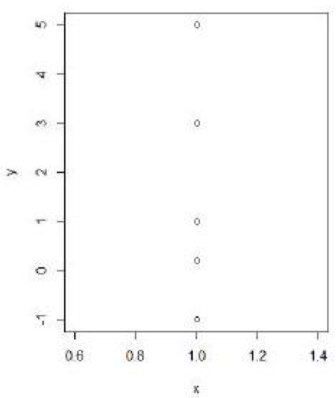
$\iff$ $\text{rank}(X^TX) = 2$

(2) X^TX 不可逆 ($\text{rank}(X^TX) \neq 2$) 的情况

注意到 $\text{rank}(AB) \le \min(\text{rank}(A), \text{rank}(B))$，且 $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$

考虑 $\text{rank}(X) \neq 2$ 的情况 (此时 $\text{rank}(X^TX)$ 一定 $\neq 2$)

则 $x_1 = x_2 = \cdots = x_n$



注：一般我们均会假设 X^TX 可逆

§3 $\hat{\beta}$ 的期望与方差 (部分结论适用于 MLR)

1. $\hat{\beta}$ 的期望

$\hat{\beta}$ 的期望为

$$E[\hat{\beta}] = E[(X^TX)^{-1}X^Ty]$$

$$= (X^TX)^{-1}X^T E[y]$$
$$= (X^TX)^{-1}X^T X \beta$$
$$= \beta$$

因此 $\hat{\beta}$ 无偏，$E(\hat{\beta}_1) = \beta_1$，$E(\hat{\beta}_0) = \beta_0$.

2、$\hat{\beta}$ 的方差

$\hat{\beta}$ 的方差为

$$Var(\hat{\beta}) = Var((X^TX)^{-1}X^Ty)$$
$$= (X^TX)^{-1}X^T Var(y) X(X^TX)^{-1} \quad (Var(Ax) = A \cdot var(x) A^T)$$
$$= (X^TX)^{-1}X^T \sigma^2 I X(X^TX)^{-1}$$
$$= \sigma^2 (X^TX)^{-1}X^TX(X^TX)^{-1}$$
$$= \sigma^2 (X^TX)^{-1}$$
$$= \frac{\sigma^2}{S_{xx}} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

因此，

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{S_{xx}} \cdot \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{\sigma^2}{S_{xx}} \cdot \frac{1}{n}(S_{xx} - n\bar{x}^2) = \sigma^2\left(\frac{1}{n} - \frac{\bar{x}^2}{S_{xx}}\right)$$
$$Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$
$$Cov(\beta_0, \beta_1) = -\frac{\sigma^2\bar{x}}{S_{xx}}$$