Recall:

Linear model

(Data) $\{y_1, y_2, \dots y_n\}$ - n observations

$\{\beta_1, \beta_2, \dots, \beta_m\}$ - parameters

y_i
- $E[y_i] = \sum_{j=1}^{m} x_{ij}\beta_j$
- uncorrelated
- $Var[y_i] = \sigma^2$

$$y_i = \sum_{j=1}^{m} x_{ij}\beta_j + \varepsilon_i \qquad 1 \le i \le n$$

$\{\varepsilon_i\}_{i=1}^{n}$ are uncorrelated r.v.

$E[\varepsilon_i] = 0$ and $Var(\varepsilon_i) = \sigma^2$

$$\underline{y} = (y_1, \dots, y_n)^T, \quad \underline{\beta} = (\beta_1, \dots, \beta_m)^T$$

$$X = (x_{ij})_{\substack{1 \le i \le n \\ 1 \le j \le m}} \qquad \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$$

$$\underline{y} = X\underline{\beta} + \varepsilon$$

linear model : $(\underbrace{\underline{y}}_{\text{Data}}, \underbrace{X\beta}_{\mathbb{E}[\underline{y}]}, \underbrace{\sigma^2 I_{n \times n}}_{\text{Variance Covariance matrix}})$

Examples:

① one way classification model

$$y_{ij} = \mu + \mu_i + \varepsilon_{ij} \qquad \begin{array}{l} 1 \le j \le n_i \\ 1 \le i \le k \end{array}$$

② Nested Classification model

$$y_{ijk} = \mu + \mu_i + \delta_{ij} + \varepsilon_{ijk}$$

$$\begin{array}{l} 1 \le k \le n_{ij} \\ 1 \le j \le m_i \\ 1 \le i \le k \end{array}$$

③ Two way classification model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$$\begin{array}{l} 1 \le k \le n_{ij} \\ 1 \le j \le J \\ 1 \le i \le I \end{array}$$

Ex: Express ①, ②, ③ as

$$(\underline{y},\ X\beta,\ \sigma^2 I_{*\times *})$$

Least Square Estimation :

Suppose $(\underline{y}, X\beta, \sigma^2 I_{n\times n})$ be a linear model.

Definition: A (random) vector $\hat{\underline{\beta}} \in \mathbb{R}^m$ is called a least square estimate of β if

$$(*) \quad \|\underline{y} - X\hat{\underline{\beta}}\|_2 = \min_{\beta\in\mathbb{R}^m} \|\underline{Y} - X\underline{\beta}\|_2$$

Recall: $z \in \mathbb{R}^n$ then $\|z\|_2 = \sqrt{\sum_{i=1}^{n} z_i^2}$ ← in $\|\cdot\|_2$ distance

Remark: $(*)$ requires that $X\hat{\beta}$ is closest to $\underline{y}$ among all possible choices of $X\beta$ with $\beta \in \mathbb{R}^m$

- **Parametric estimate** : linear parametric function $p^T\beta$, a least square estimate is $p^T\hat{\beta}$.

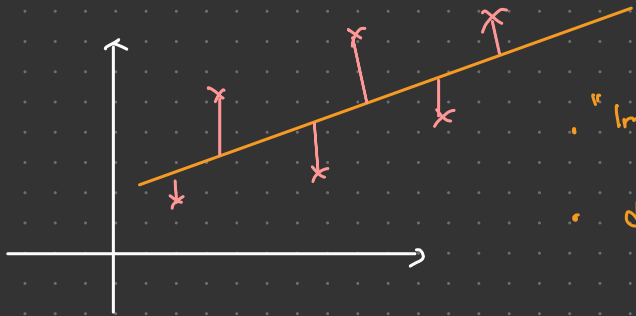
Theorem: Consider a linear model $(\underline{y}, X\beta, \sigma^2 I)$. Then any least square estimate $\hat{\beta}$ of β satisfies

$$(X^T X)\hat{\underline{\beta}} = X^T\underline{y} \quad \cdots \quad (1)$$

Remark: (1) is called Normal equation.

Lecture 2:

Recall day 1



- "lm - function in" R
- distance function to find best line

Ex. Try to find the relationship of above with least square estimate

Proof of Theorem:

$$R_0^2 = \min_{\underline{\beta} \in \mathbb{R}^n} \| \underline{Y} - X \underline{\beta} \|_2^2$$

We will show $R_0^2 = \| \underline{Y} - X \hat{\underline{\beta}} \|_2^2$ where

$\hat{\underline{\beta}}$ solves $(X^T X) \hat{\underline{\beta}} = X^T \underline{y}$

[Remarks :- • Solution set may not be unique
• it exists is all that will be shown]

First of all: $X^T \underline{y} \in \mathcal{C}(X^T) \overset{\text{Ex.}}{=} \mathcal{C}(X^T X)$

$\Rightarrow \exists\, \hat{\beta} \in \mathbb{R}^m$ s.t $(X^T X) \hat{\beta} = X^T Y$

$\underline{\beta} \in \mathbb{R}^m$

$[\, z \in \mathbb{R}^n \,;\ \|z\|_2^2 = \sum_{i=1}^{n} z_i^2 \,]$

$$\|Y - X\underline{\beta}\|_2^2$$

$$= \|Y - X\hat{\beta} + X\hat{\beta} - X\beta\|_2^2$$

$$= \|Y - X\hat{\beta} + X(\hat{\beta} - \beta)\|_2^2$$

<u>Fact:</u> $\|z + w\|_2^2 = \sum_{i=1}^{n} (z_i + w_i)^2 = \sum_{i=1}^{n} z_i^2 + \sum_{i=1}^{n} w_i^2 + 2\sum_{i=1}^{n} z_i w_i$

$z, w \in \mathbb{R}^n$ $\qquad = \|z\|_2^2 + \|w\|_2^2 + 2\langle z, w\rangle$

where $\langle z, w\rangle = z^T w$

$$= \|Y - X\hat{\beta}\|_2^2 + \|X(\hat{\beta} - \beta)\|_2^2 + 2\langle Y - X\hat{\beta}, X(\hat{\beta} - \beta)\rangle \quad (+)$$

lets understand ; $\langle Y - X\hat{\beta}, X(\hat{\beta} - \beta)\rangle$

$$\langle Y - X\hat{\beta}, X(\hat{\beta} - \beta)\rangle = (Y - X\hat{\beta})^T X(\hat{\beta} - \beta)$$

$$= (\hat{\beta} - \beta)^T X^T (Y - X\hat{\beta})$$

One can take transpose since its a scalar

$$= (\hat{\beta} - \beta)^T (X^T Y - X^T X\hat{\beta})$$

$$= 0 \quad (\text{By choice of } \hat{\beta})$$

$(++)$

From (+) & (++)

$$\|Y - X\beta\|_2^2 = \|Y - X\hat{\beta}\|_2^2 + \|X(\hat{\beta} - \beta)\|_2^2$$

$$\geq \|Y - X\hat{\beta}\|_2^2$$

and further $= \|Y - X\hat{\beta}\|_2^2$ iff $\hat{\beta} = \beta$

$$\therefore \; R_0^2 = \|Y - X\hat{\beta}\|_2^2 \qquad \square$$

Lecture 3

Recall: $(\underline{y}, X\beta, \sigma^2 I_{n\times n})$ - linear model

$\underline{p} \in \mathbb{R}^m$, the linear parameter function

$\underline{p}^T \underline{\beta}$ is said to be estimable if

$\exists\; l \in \mathbb{R}^n$ st $E(l^T \underline{y}) = p^T \underline{\beta}$

$\forall\; \beta \in \mathbb{R}^n$. (unbiased estimate for $l^T \underline{y}$)

Theorem: let $(\underline{y}, X\beta, \sigma^2 I_{n\times n})$ be a linear model and $p \in \mathbb{R}^n$.

$\hat{p}\beta$ is estimable $\iff$ its least square estimate (ie $p^T \hat{\underline{\beta}}$) is unique.

Proof :-

Suppose $\hat{p}^T \underline{\beta}$ is estimable

$\Rightarrow$ $\underline{p} \in \mathcal{C}(X^T X)$ [Recall earlier Theorem Feb 9: $p \in \mathcal{C}(X^T)$]

$\Rightarrow \exists\, \underline{b} \in \mathbb{R}^m : \quad X^T X \underline{b} = \underline{p} \quad - (I)$

Suppose $\hat{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_2$ are any two solutions of Normal Equation $(X^T X \underline{\beta} = X^T \underline{y}) \quad - (II)$

$$(I) \Rightarrow \quad \underline{p}^T \hat{\underline{\beta}}_1 = \underline{b}^T (X^T X)^T \hat{\underline{\beta}}_1$$

$$= \underline{b}^T X^T X \hat{\underline{\beta}}_1$$

$$\overset{(II)}{=} \underline{b}^T X^T \underline{y}$$

$$\overset{(II)}{=} b^T (X^T X \hat{\beta}_2)$$

$$= (b^T (X^T X)) \hat{\beta}_2$$

$$= p^T \hat{\beta}_2$$

$\therefore$ The least square estimate of $\hat{p}^T \underline{\beta}$ is unique.

$\Leftarrow$ • Suppose $\underline{p} \in \mathbb{R}^n$ is s.t the least square estimate of $\underline{p}^T \underline{\beta}$ is unique.

• Then for any two solutions $\hat{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_2$ of the Normal equation $(X^T X \underline{\beta} = X^T \underline{y})$

$$\underline{p}^T \hat{\underline{\beta}}_1 = \underline{p}^T \hat{\underline{\beta}}_2$$

• Ex: $\forall\ \underline{\lambda}$ is st $X^T X \underline{\lambda} = 0$

then $\underline{p}^T \underline{\lambda} = 0$

$\Rightarrow$ $\underline{p}$ is orthogonal to $\underline{\lambda}$

$\Rightarrow$ p is orthogonal to $N(X^T X)$

$\Rightarrow$ $p \in$ Row space $(X^T X)$

$\overset{\text{Ex.}}{\Rightarrow}$ $p \in$ Column space $(X^T X)$ $\square$

lecture 4 :

Proposition : $X \underline{\beta}$ is estimable and its least square estimate is given by $P_X \underline{y}$ where P_X is the orthogonal projection onto $\mathcal{C}(X)$

Proof :- $X\beta = \begin{pmatrix} p_1^T \\ \vdots \\ p_n^T \end{pmatrix}\beta$ when $X = \begin{pmatrix} p_1^T \\ \vdots \\ p_n^T \end{pmatrix}$

i.e. $p_i^T \in \mathcal{C}(X^T) \quad 1 \le i \le n$

$\iff X\beta$ is estimable
(Ex.)

Suppose $\hat{\beta}$ is a solution to Normal equation

$$R_0^2 = \|Y - X\hat{\beta}\|_2^2 = \min_{\beta \in \mathbb{R}^m} \|Y - X\beta\|_2^2$$

and moreover $X\hat{\beta} \in \mathcal{C}(X)$

$$\Rightarrow R_0^2 = \min_{\eta \in \mathcal{C}(X)} \|Y - \eta\|_2^2$$

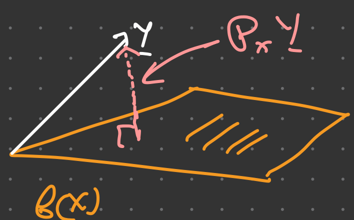
$$\overset{(Ex.)}{=} \|Y - P_X Y\|_2^2 \quad \text{where}$$

P_X is the orthogonal projection onto column span of X.

$$\Rightarrow X\hat{\beta} = P_X Y. \qquad \square$$

Linear Algebra fact

Review

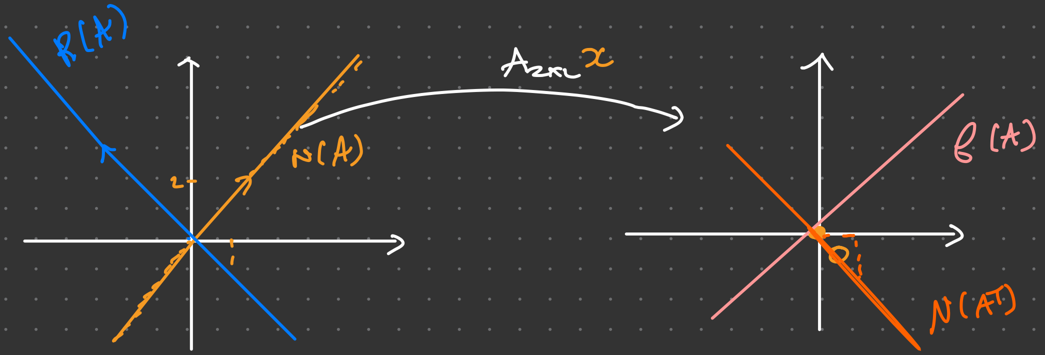


$\mathbb{R}^2$ $\qquad A_{2\times 2} = \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}$

Row space (A)

Column space (A)

Null space (A) $\equiv N(A)$



$$N(A) = \{ x \in \mathbb{R}^2 : Ax = 0 \}$$

$$x \in \mathbb{R}^2 : \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\iff 2x_1 - x_2 = 0$$

$$\boxed{\text{e.g. } x_1 = 1, \; x_2 = 2}$$

$$R(A) = \text{Span}\left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \end{bmatrix} \right\}$$

$$= \left\{ \alpha \begin{bmatrix} 2 \\ -1 \end{bmatrix} : \alpha \in \mathbb{R} \right\}$$

$$C(A) = \{ b_{2\times 1} \mid \exists x \in \mathbb{R}^2 : Ax = b \}$$

$$= \{ b_{2\times 1} \mid \exists x \in \mathbb{R}^2 \; b = x_1 A_{*1} + x_2 A_{*2} \}$$

Solve: $\begin{bmatrix} 2 & -1 & b_1 \\ 4 & -2 & b_2 \end{bmatrix}$ row reduce

$R_2 \to R_2 - 2R_1 \quad \begin{bmatrix} 2 & -1 & b_1 \\ 0 & 0 & b_2 - 2b_1 \end{bmatrix}$

$b \in C(A) \iff b_2 - 2b_1 = 0$

$b_2 = 2b_1 \quad$ e.g. $b_1 = 1, \; b_2 = 2$

$$N(A^T) = \{ y_{2\times 1} : A^T y = 0 \}$$

$$\begin{bmatrix} 2 & 4 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0$$

$$\iff -y_1 - 2y_2 = 0$$

$$y_1 = -2y_2$$

$$\text{Nullity}(A) = \dim(\text{Null}(A)) \qquad A_{n \times n}$$

$$+ \qquad\qquad +$$

$$\text{Rank}(A) = \dim(\text{Rowspace}(A))$$

$$n$$

- $R(A) \perp N(A)$

- $x \in N(A) \Rightarrow Ax = 0$

$$\Rightarrow A^T A x = 0$$

$$\Rightarrow x \in N(A^T A)$$

$x \in N(A^T A) \Rightarrow A^T A x = 0$

$$\Rightarrow x^T A^T A x = 0$$

$$\Rightarrow \langle Ax, Ax \rangle = 0$$

$$\Rightarrow \|Ax\|_2^2 = 0$$

$$\Rightarrow Ax = 0$$

$$\Rightarrow x \in N(A)$$

$$\therefore N(A) = N(A^T A)$$

$$\mathcal{C}(A^T) \subseteq \mathcal{C}(A^T A)$$

[in recording mentioned A instead of A^T]

Rank nullity Theorem $\Rightarrow$ $\rho(A^TA) = \rho(A^T)$