Projection Matrix.

$$X = (\underset{\in \mathbb{R}^n}{X^1}, \cdots, X^d) : n\times d \text{ matrix.} \qquad (d \leq n)$$

$$\begin{aligned} Col(X) &= span\{X^1, \cdots, X^d\} \leq \mathbb{R}^n \\ &= \left\{ \sum_{j=1}^{d} \beta_j \cdot X^j \mid \beta_1, \cdots, \beta_d \in \mathbb{R} \right\} \\ &= \{ X\beta \mid \beta \in \mathbb{R}^d \} \end{aligned}$$

Consider $\Pi_X := X(X^tX)^{-1}X^t$

(assume $X^1 \sim X^d$ are lin. indep so that X^tX is full rank)

Observations

① $\Pi_X \cdot X = X(X^tX)^{-1}X^tX = X$, $X^t\Pi_X = X^t$

①-1 $\Pi_X \cdot X^j = X^j$ for $j = 1, 2, \cdots, d$

①-2 $\forall\, v \in Col(X)$, $\Pi_X v = v$

② $\Pi_X^2 = X(X^tX)^{-1}\cancel{X^t X(X^tX)^{-1}}X^t = \Pi_X$

②-1 $(I - \Pi_X)^2 = I - 2\Pi_X + \Pi_X^2 = I - \Pi_X$

③ $\forall\, v \in \mathbb{R}^n$, $\Pi_X v = X\left((X^tX)^{-1}X^t v\right) \in Col(X)$

$X^t(I - \Pi_X)v = 0 \Rightarrow (I - \Pi_X)v \in (Col(X))^\perp$

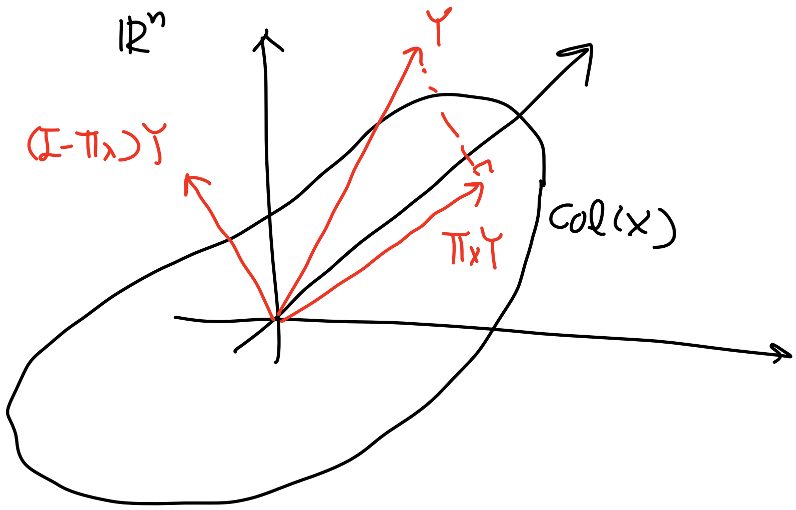
④ $tr(\Pi_X) = tr(X(X^tX)^{-1}X^t)$

$= tr((X^tX)^{-1}X^tX) = tr(I_d) = d$

$\Rightarrow tr(I - \Pi_X) = tr(I_n) - tr(\Pi_X) = n - d$

$$v = (I - \Pi_X + \Pi_X) v$$

$$= \underbrace{\Pi_X v}_{\in \operatorname{col}(X)} + \underbrace{(I - \Pi_X) v}_{\in (\operatorname{col}(X))^\perp}$$



$\therefore$ Π_X : projection matrix to column space of X

$I - \Pi_X$: " " " " " " "

orthogonal complement of

$\checkmark$ if $X = \mathbb{1}_n = \underbrace{(1, \cdots, 1)}_{n \text{ times}}^t \in \mathbb{R}^{n \times 1}$

$$\Rightarrow \Pi_X = \mathbb{1}\,(\underbrace{\mathbb{1}^t \mathbb{1}}_{=n})^{-1} \mathbb{1}^t = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ & \vdots & \\ 1 & \cdots & 1 \end{pmatrix}$$

$$\Rightarrow \Pi_X y = \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} = \mathbb{1}_n \cdot \bar{y} \qquad \left(\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i\right)$$

$$\& \ (I - \Pi_X) y = y - \Pi_X y = (y_1 - \bar{y}, \cdots, y_n - \bar{y})^t$$

$X_1 \in \mathbb{R}^{n \times d_1}$, $X_2 \in \mathbb{R}^{n \times d_2}$

if $col(X_1) \leqslant col(X_2)$, $\Pi_{X_2}\Pi_{X_1} = \Pi_{X_1}\Pi_{X_2} = \Pi_{X_2}$

(pf) $\forall v \in \mathbb{R}^n$ $\quad \Pi_{X_2} v \in col(X_2) \leqslant col(X_1)$

$$\therefore \Pi_{X_1}(\Pi_{X_2} v) = \Pi X_2 v$$

$$\therefore \Pi_{X_1}\Pi_{X_2} = \Pi_{X_2},$$

$\Pi_{X_2}\Pi_{X_1} = \Pi_{X_1}\Pi_{X_2}$ because $\Pi_{X_1}\Pi_{X_2} = \Pi_{X_2}$ is symmetric

$\hat{\beta}_{OLS} = \operatorname{argmin} \|y - X\beta\|_2^2$ (X is full rank)

. vector calculus approach

$$\|y - X\beta\|_2^2 = \beta^t X^t X \beta - 2y^t X\beta + y^t y$$

$$\frac{\partial}{\partial \beta} \bigcirc = 2(X^t X)\beta - 2X^t y$$

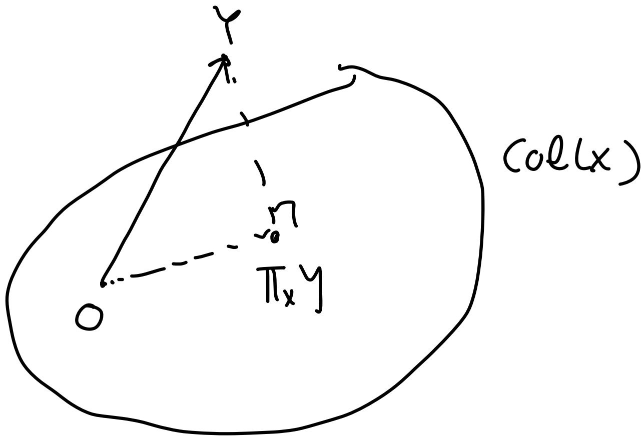
$$\frac{\partial^2}{\partial \beta^2} \bigcirc = 2(X^t X) \succ 0$$

$\therefore \hat{\beta}$ solves $\quad 2(X^t X)\beta - 2X^t y = 0 \Rightarrow \hat{\beta}_{OLS} = (X^t X)^{-1} X^t y$

minimize $\|y - X\beta\|_2$

$\Rightarrow$ find $X\beta \in col(X)$ that has closest distance from y

: orthogonal projection



$$\Pi_x y = X(X^tX)^{-1}X^t y = X\hat{\beta}$$

Statistical properties : next time.

- Hint for HW6 Pb4.

$$\delta^{JS}(Y) = \left(1 - \frac{d-2}{\|Y\|_2^2}\right)Y$$: works well when $\theta = 0$

$\Rightarrow \theta \in col(0)$

$$\delta^{(1)}(Y) = \bar{Y}\mathbb{1}_d + \left(1 - \frac{d-3}{\|Y - \bar{Y}\mathbb{1}_d\|_2^2}\right)(Y - \bar{Y})$$

: will work well when $\theta_1 = \cdots = \theta_d$

$\Rightarrow \theta \in col(\mathbb{1}_d)$

$\delta^{(2)}(Y) = ??$. want to work well when $\theta = X\beta \in col(X)$

$$\delta^{JS} : \quad \underset{\in col(0)}{Y = 0} + \underset{\in (col(0))^\perp}{Y}$$

$$\delta^{JS} = 0 + (1-\alpha)Y$$

$$\delta^{(1)} : \quad Y = \underset{\in col(\mathbb{1})}{\Pi_{\mathbb{1}} Y} + \underset{\in (col(\mathbb{1}))^\perp}{(I - \Pi_{\mathbb{1}})Y}$$

$$\delta^{(1)} = \Pi_{\mathbb{1}} Y + (1-\alpha)(I-\Pi_{\mathbb{1}})Y$$

$\Rightarrow$ we are leaving orthogonal projection part the same

and shrinking orthogonal complement part

what next??