Tsay Sofya, group 1

Home Assignment 3, theory.

#1

$$loss(\hat\beta) = (y-\hat y)^T(y-\hat y) + \lambda \hat\beta^T\hat\beta \quad , \quad \hat y = \hat\beta_1 x + \hat\beta_2 x$$

x - standardized

a) $\hat\beta_1, \hat\beta_2 = ?$

$$\hat\beta = \begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} \qquad \hat y = \hat\beta_1 x + \hat\beta_2 x = X\hat\beta = (x \quad x)\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}$$

$$loss(\hat\beta) = (y - X\hat\beta)^T(y - X\hat\beta) + \lambda\hat\beta^T\hat\beta = y^Ty - y^TX\hat\beta -$$

$$-\hat\beta^TX^Ty + \underline{\hat\beta^TX^TX\hat\beta} + \underline{\lambda\hat\beta^T\hat\beta} = \underbrace{y^Ty}_{const} - y^TX\hat\beta - \hat\beta^TX^Ty + \hat\beta^T(X^TX+\lambda I)\hat\beta$$

$$(X^TX+\lambda I)^T = X^TX+\lambda I$$

$$d\,loss(\hat\beta) = -y^TX\cdot d\hat\beta - d\hat\beta^T\cdot X^Ty + d\,\hat\beta^T(X^TX+\lambda I)\hat\beta \Leftrightarrow$$

$$\left\{ F = x^T(Ax) \Rightarrow dF = dx^T\cdot Ax + x^Td(Ax) = dx^T\cdot Ax + x^T(\cancel{dA\cdot x}^{\,0} + A\,dx) = \right.$$

$$= \underbrace{dx^T\cdot Ax}_{scalar\ =(dx^T\cdot Ax)^T = x^TA^Tdx} + x^TA\cdot dx = x^TA^Tdx + x^TAdx = x^T(A^T+A)\,dx$$

$$\Leftrightarrow -y^TX\cdot d\hat\beta - \underbrace{d\hat\beta^T\cdot X^Ty}_{scalar} + 2\hat\beta^T(X^TX+\lambda I)\cdot d\hat\beta \Leftrightarrow$$

$$(d\hat\beta^T\cdot X^Ty)^T = y^TX\cdot d\hat\beta$$

$$\Leftrightarrow -2y^TX\cdot d\hat\beta + 2\hat\beta^T(X^TX+\lambda I)\,d\hat\beta$$

FOC $\quad d\,loss(\hat\beta) = 0 \Leftrightarrow -y^TX + \hat\beta^T(X^TX+\lambda I) = 0$

$$\hat\beta^T(X^TX+\lambda I) = y^TX \quad \text{) transpose}$$

$$(X^TX+\lambda I)\hat\beta = X^Ty$$

$$\boxed{\hat\beta = (X^TX+\lambda I)^{-1}(X^Ty)}$$ matrix form

In our problem:

$$X = \begin{pmatrix} x_1 & x_1\\ \vdots & \vdots\\ x_n & x_n\end{pmatrix} = \begin{pmatrix} | & |\\ x & x\\ | & |\end{pmatrix}$$

$$X^T = \begin{pmatrix} x_1 & \dots & x_n\\ x_1 & \dots & x_n\end{pmatrix} = \begin{pmatrix} — & x^T & —\\ — & x^T & —\end{pmatrix}$$

$$X^TX = \underset{[2\times n]}{\begin{pmatrix} — & x^T & —\\ — & x^T & —\end{pmatrix}}\underset{[n\times 2]}{\begin{pmatrix} | & |\\ x & x\\ | & |\end{pmatrix}} = \begin{pmatrix} x^Tx & x^Tx\\ x^Tx & x^Tx\end{pmatrix} = \begin{pmatrix}\sum x_i^2 & \sum x_i^2\\ \sum x_i^2 & \sum x_i^2\end{pmatrix}$$

$$X^T y = \underset{[2\times n]}{\begin{pmatrix} - & x^T & - \\ - & x^T & - \end{pmatrix}} \underset{[n\times 1]}{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}} = \begin{pmatrix} \sum x_i y_i \\ \sum x_i y_i \end{pmatrix}$$

$$(X^T X + \lambda I) = \begin{pmatrix} \sum x_i^2 & \sum x_i^2 \\ \sum x_i^2 & \sum x_i^2 \end{pmatrix} + \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} \lambda + \sum x_i^2 & \sum x_i^2 \\ \sum x_i^2 & \lambda + \sum x_i^2 \end{pmatrix}$$

$$\det(X^T X + \lambda I) = (\lambda + \sum x_i^2)^2 - (\sum x_i^2)^2 =$$

$$= (\lambda + \sum x_i^2 - \sum x_i^2)(\lambda + 2\sum x_i^2) = \lambda(\lambda + 2\sum x_i^2)$$

$$(X^T X + \lambda I)^{-1} = \frac{1}{\lambda(\lambda + 2\sum x_i^2)} \begin{pmatrix} \lambda + \sum x_i^2 & -\sum x_i^2 \\ -\sum x_i^2 & \lambda + \sum x_i^2 \end{pmatrix}$$

$$\hat{\beta} = (X^T X + \lambda I)^{-1}(X^T y) = \frac{1}{\lambda(\lambda + 2\sum x_i^2)} \underset{[2\times 2]}{\begin{pmatrix} \lambda + \sum x_i^2 & -\sum x_i^2 \\ -\sum x_i^2 & \lambda + \sum x_i^2 \end{pmatrix}} \underset{[2\times 1]}{\begin{pmatrix} \sum x_i y_i \\ \sum x_i y_i \end{pmatrix}} =$$

$$= \frac{\sum x_i y_i}{\lambda(\lambda + 2\sum x_i^2)} \begin{pmatrix} \lambda + \sum x_i^2 & -\sum x_i^2 \\ -\sum x_i^2 & \lambda + \sum x_i^2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} =$$

$$= \frac{\sum x_i y_i}{\lambda(\lambda + 2\sum x_i^2)} \begin{pmatrix} \lambda + \cancel{\sum x_i^2} - \cancel{\sum x_i^2} \\ -\cancel{\sum x_i^2} + \lambda - \cancel{\sum x_i^2} \end{pmatrix} = \frac{\sum x_i y_i}{\lambda(\lambda + 2\sum x_i^2)} \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}$$

$$\Rightarrow \boxed{\hat{\beta}_1 = \hat{\beta}_2 = \frac{\sum x_i y_i}{\lambda + 2\sum x_i^2}}$$

b) as $\lambda \to \infty$: $\hat{\beta}_1 = \hat{\beta}_2 = \dfrac{\sum x_i y_i}{\lambda + 2\sum x_i^2} \sim \dfrac{1}{\lambda} \to 0$

c) as $\lambda \to 0$: $\hat{\beta}_1 + \hat{\beta}_2 = \dfrac{2\sum x_i y_i}{\lambda + 2\sum x_i^2} \to \dfrac{\cancel{2}\sum x_i y_i}{\cancel{2}\sum x_i^2} = \dfrac{\sum x_i y_i}{\sum x_i^2}$

$\sim$ regression model w/o intercept $y = \beta x + \varepsilon$

#2

$y = X\beta + u \qquad E(u|X) = 0 \qquad X - [n \times k]$ matrix, rank $X = k$

$Var(u|X) = \sigma^2 W, \quad W \neq I$

$\hat{\beta}$ – standard OLS estimator of β

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T y}$$

a) $E(\hat{\beta}|X) = E\left((X^TX)^{-1} X^T y \mid X\right) = (X^TX)^{-1} X^T E(y|X) =$

$= (X^TX)^{-1} X^T E(X\beta + u|X) = (X^TX)^{-1} X^T (X\beta + \cancel{E(u|X)}^{0}) =$

$= \cancel{(X^TX)^{-1}} \cancel{X^TX} \beta = \beta \Rightarrow \boxed{E(\hat{\beta}|X) = \beta}$

b) $Var(\hat{\beta}|X) = Var\left(\underbrace{(X^TX)^{-1} X^T}_{} y \mid X\right) = (X^TX)^{-1} X^T Var(y|X) \cdot X (X^TX)^{-1} \circledcirc$

$$\xleftarrow{\quad\quad} \quad \xrightarrow{T}$$

$Var(y|X) = Var(\underbrace{X\beta}_{\text{fixed}} + u|X) = Var(u|X) = \sigma^2 W$

$\circledcirc\ \sigma^2 (X^TX)^{-1} X^T W X (X^TX)^{-1} \Rightarrow \boxed{Var(\hat{\beta}|X) = \sigma^2 (X^TX)^{-1} X^T W X (X^TX)^{-1}}$

c) in case $\underline{W = I}$ we had $Var(\hat{\beta}|X) = \sigma^2 \cancel{(X^TX)^{-1}} \cancel{X^TX} (X^TX)^{-1} =$
$= \sigma^2 (X^TX)^{-1}$

and $CI_{1-\alpha}(\beta) = \hat{\beta} \pm \sigma \sqrt{(X^TX)^{-1}} \cdot z_{\alpha/2}$

But now $W \neq I$ and $Var(\hat{\beta}|X) \neq \sigma^2 (X^TX)^{-1}$,

thus applying standard CI for β is <u>invalid</u>.

Insted, for example we could use method proposed by White $\widehat{Var}_{HC0}(u|X) = \begin{pmatrix} \hat{u}_1^2 & & & 0 \\ & \hat{u}_2^2 & & \\ & & \ddots & \\ 0 & & & \hat{u}_n^2 \end{pmatrix}$, $\hat{u}_i = y - \hat{y}_i$

or $\widehat{Var}_{HC3}(u|X) = \begin{pmatrix} \hat{u}_{1,cv}^2 & & & 0 \\ & \hat{u}_{2,cv}^2 & & \\ & & \ddots & \\ 0 & & & \hat{u}_{n,cv}^2 \end{pmatrix}$ CV – cross validated (LOO)

(it can be shown that $E(\hat{u}_{i,cv}^2|X) = Var(u_i|X)$
$E(\hat{u}_i^2|X) < Var(u_i|X)$)

or some other measures to fight heteroskedasticity.

* info about HC0 and HC3 is from Boris' Pandametrics 🐼 course

d) $\mathrm{cov}(y, \hat{\beta} \mid X) = \mathrm{cov}(y, \underbrace{(X^T X)^{-1} X^T}_{T} y \mid X) =$

$= \mathrm{Var}(y \mid X)\, X (X^T X)^{-1} = \boxed{\sigma^2 W X (X^T X)^{-1} = \mathrm{cov}(y \mid \hat{\beta})}$

$\mathrm{cov}(y, \hat{\beta} \mid X) = \mathrm{cov}(\underbrace{\cancel{X\beta}}_{\text{fixed}} + u, (X^T X)^{-1} X^T y \mid X) =$

$= \mathrm{cov}(u, (X^T X)^{-1} X^T y \mid X)$

$\hat{\beta} = (X^T X)^{-1} X^T (X\beta + u) = \beta + (X^T X)^{-1} X^T u = \beta + Au$

$\mathrm{cov}(u, \underbrace{\cancel{\beta}}_{\text{fixed}} + (X^T X)^{-1} X^T u \mid X) = \mathrm{cov}(u, \underbrace{(X^T X)^{-1} X^T}_{\text{fixed}} u \mid X) =$

$= \mathrm{Var}(u) \cdot ((X^T X)^{-1} X^T)^T = \sigma^2 W \cdot X (X^T X)^{-1}$

a) $E(\hat{\beta}) \overset{\text{tower property}}{=} E(E(\hat{\beta} \mid X)) = E(\beta) = \beta$

#3

$$X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$

a) $$X^T X = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$$

$$\det(X^T X - \lambda I) = (6-\lambda)^2 - 1^2 = (6-\lambda-1)(6-\lambda+1) = (5-\lambda)(7-\lambda) = 0$$

$$\lambda_1 = 5 : \quad (X^T X - 5I) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \rightarrow \boxed{v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}}$$

$$\lambda_2 = 7 : \quad (X^T X - 7I) = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \rightarrow \boxed{v_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}}$$

$$\boxed{X^T X = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix}\underbrace{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^T}_{= \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^{-1}}}$$

(orthogonal)

b) Find SVD of X

SVD: $X = U D V^T$ $\qquad$ D - diagonal

$\qquad\qquad\qquad\qquad\qquad$ U, V - orthogonal

$$X^T X = V D^T \cancel{U^T U} D V^T = V D^T D V^T = V \begin{pmatrix} d_{11}^2 & 0 \\ 0 & d_{22}^2 \end{pmatrix} V^T$$

from (a): $$\boxed{V = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}} \qquad d_{11} = \sqrt{7}, \quad d_{22} = \sqrt{5}$$

$$X X^T = U D \cancel{V^T V} D^T U^T = U D D^T U^T$$

$$X X^T = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

$$\det(X X^T - \lambda I) = \begin{pmatrix} 5-\lambda & 0 & 3 \\ 0 & 5-\lambda & 1 \\ 3 & 1 & 2-\lambda \end{pmatrix} = -\lambda(\lambda-5)(\lambda-7) = 0$$

$$\lambda_1 = 7 : \quad (X X^T - 7I) = \begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix} \rightarrow u_1 = \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}$$

$$\lambda_2 = 5 : \quad (X X^T - 5I) = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix} \rightarrow u_2 = \begin{pmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \\ 0 \end{pmatrix}$$

$$\lambda_3 = 0 : \quad (X X^T - 0I) = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \rightarrow u_3 = \begin{pmatrix} -3/\sqrt{35} \\ -1/\sqrt{35} \\ 5/\sqrt{35} \end{pmatrix}$$

$$U = \begin{pmatrix} 3/\sqrt{14} & -1/\sqrt{10} & -3/\sqrt{35} \\ 1/\sqrt{14} & 3/\sqrt{10} & -1/\sqrt{35} \\ 2/\sqrt{14} & 0 & 5/\sqrt{35} \end{pmatrix}$$

SVD of X: $$X = \begin{pmatrix} 3/\sqrt{14} & -1/\sqrt{10} & -3/\sqrt{35} \\ 1/\sqrt{14} & 3/\sqrt{10} & -1/\sqrt{35} \\ 2/\sqrt{14} & 0 & 5/\sqrt{35} \end{pmatrix} \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}^T$$

c) $$\hat{X} = u_1 d_{11} v_1^T = \sqrt{7} \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} =$$

$$= \sqrt{7} \begin{pmatrix} 3/\sqrt{28} & 3/\sqrt{28} \\ 1/\sqrt{28} & 1/\sqrt{28} \\ 2/\sqrt{28} & 2/\sqrt{28} \end{pmatrix} = \begin{pmatrix} 3/2 & 3/2 \\ 1/2 & 1/2 \\ 1 & 1 \end{pmatrix}$$

4

$X = UDV^T$ (SVD)

$$\begin{cases} \|Xw\|^2 \to \max\limits_w \\ \text{s.t. } \|w\|^2 = 1 \end{cases} \iff \begin{cases} (Xw)^T Xw \to \max\limits_w \\ \text{s.t. } w^T w = 1 \end{cases}$$

$$\begin{cases} w^T X^T X w \to \max\limits_w \\ \text{s.t. } w^T w = 1 \end{cases}$$

d) $L(w,\lambda) = w^T X^T X w - \lambda(w^T w - 1)$

e) $d\Big(w^T \overset{= X^TX = (X^TX)^T}{A} w - \lambda(w^T w)\Big) = d(w^T A w) - \lambda\, d(w^T w) =$

$d(w^T A w) = dw^T \cdot Aw + w^T d(Aw) = dw^T \cdot Aw + w^T(\cancel{dA}\cdot w + A\,dw) =$

$= dw^T \cdot Aw + w^T A\,dw = 2w^T A(dw)$

$d(w^T w) = dw^T \cdot w + w^T \cdot dw = 2w^T dw$

$= 2w^T A(dw) - 2\lambda w^T I(dw) = 2w^T(A - \lambda I)dw = 2w^T(X^TX - \lambda I)dw$

FOC $\begin{cases} w^T(X^TX - \lambda I) = 0 \\ w^T w = 1 \end{cases} \xrightarrow{\text{transpose}} \boxed{\begin{cases} (X^TX - \lambda I)w = 0 \\ w^T w = 1 \end{cases}}$

interpretation: λ - eigenvalue, w - eigenvector of matrix X^TX

f) From part (a) we know that

$$X^TX = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}^T$$

$w^*_{max} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ corresponds to highest eigenvalue

$w_{min} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$ for smallest eigenvalue

already normalized

Check: $\|Xw^*_{max}\| = {w^*_{max}}^T X^TX\, w^*_{max} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} =$

$= \frac{1}{2}\begin{pmatrix} 7 & 7 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2} \cdot 14 = 7$

$\|Xw_{min}\| = w_{min}^T X^TX\, w_{min} = \frac{1}{2}\begin{pmatrix} 1 & -1 \end{pmatrix}\begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} =$

$= \frac{1}{2}\begin{pmatrix} 5 & -5 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2} \cdot 10 = 5 < 7 \Rightarrow \boxed{w^*_{max} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}}$

1st column of V corresponds to d_{11}