Polina Bobyleva, group 4. Home Assignment 3.

№1.

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

$$\hat\beta = \begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}, \quad X = \begin{pmatrix}X\\ X\end{pmatrix}$$

a) to find optimal $\hat\beta_1, \hat\beta_2$, we should minimize loss function:

$$\min_{\hat\beta} (y-\hat y)^T(y-\hat y) + \lambda\hat\beta^T\hat\beta$$

$$d\left((y-\hat y)^T(y-\hat y) + \lambda\hat\beta^T\hat\beta\right) = \underset{①}{d((y-\hat y)^T(y-\hat y))} + \underset{②}{d(\lambda\hat\beta^T\hat\beta)} = [-2(y-X\hat\beta)^T X + 2\lambda\hat\beta^T]\,d\hat\beta = 0$$

$$\begin{cases} ①\ d\left((y-\hat y)^T\cdot(y-\hat y)\right) = 2(y-\hat y)^T\cdot d(y-\hat y) = -2(y-\hat y)^T d\hat y = -2(y-\hat y)^T d(X\hat\beta) = \\ = -2(y-\hat y)^T\cdot(X d\hat\beta + \underbrace{dX}_{0}\cdot\hat\beta) = -2(y-X\hat\beta)^T\cdot X\,d\hat\beta \\ ②\ d(\lambda\hat\beta^T\hat\beta) = \lambda\hat\beta^T d\hat\beta + \lambda(d\hat\beta)^T\cdot\hat\beta = 2\lambda\hat\beta^T d\hat\beta \end{cases}$$

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

$$\rightarrow \hat\beta^* = (\lambda + X^TX)^{-1}\cdot X^Ty$$ – linear combination – optimal values of $\hat\beta_1, \hat\beta_2$ ($\hat\beta_1, \hat\beta_2$ are identical)

b) $$\lim_{\lambda\to\infty}\hat\beta^* = \lim_{\lambda\to\infty}(\lambda + X^TX)^{-1}\cdot X^Ty = 0 \rightarrow (\hat\beta_1, \hat\beta_2) = (0,0)$$ → estimates become zero vector

$\left\{\frac{1}{\lambda}\Big|_{\lambda\to\infty} \approx 0\right\}$

there are two identical predictors → it helps to solve multicollinearity problem

c) $$\lim_{\lambda\to 0}(\hat\beta_1 + \hat\beta_2) = \lim_{\lambda\to 0}(\lambda + X^TX)^{-1}\cdot X^Ty = (X^TX)^{-1}X^Ty = \hat\beta^{OLS}$$ → Sum of estimators will converge to a an OLS estimator.

№2.

$y = X\beta + u$; $E(u|X)=0$; $\underset{[n \times k]}{X}$; $\operatorname{rank} X = k$; $Var(u|X) = \sigma^2 W$ $(W \neq I)$

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

$\{$Since $\hat{\beta}_{OLS} = (X^TX)^{-1} \cdot X^Ty\}$

According to Tower property: $E(\hat{\beta}) = E(\beta) = E(E(\hat{\beta}|X)) = E(\hat{\beta}|X) = \beta$

b) $Var(\hat{\beta}|X) = Var((X^TX)^{-1}X^Ty|X) = (X^TX)^{-1}X^T Var(y|X) \cdot [(X^TX)^{-1}X^T]^T = (X^TX)^{-1}X^T \underset{Var(u|X)}{\underset{\parallel}{Var(X\beta+u|X)}} \cdot$

$\cdot X(X^TX)^{-1} = (X^TX)^{-1}X^T\sigma^2 WX(X^TX)^{-1} = \sigma^2(X^TX)^{-1}X^TWX(X^TX)^{-1}$

c) I think in this case the standard confidence interval for β will not be valid. As $W \neq I$, $Var(u|X) = \sigma^2 W \neq \sigma^2$, disturbance terms are heteroscedastic, so standard formula for error terms will result in biased estimation of confidence interval. There can be correlation across observations, as disturbance terms have non-constant variance.

d) $Cov(y, \hat{\beta}|X) = Cov(y, (X^TX)^{-1}X^Ty|X) = (X^TX)^{-1}X^T Cov(y, y|X) = (X^TX)^{-1}X^T Var(y|X) =$

$= \sigma^2 W(X^TX)^{-1}X^T$ → indeed, $Cov(y, \hat{\beta}|X) \neq 0$, as it would have been for $W = I$

№3.

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

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

$$X^T X = \underbrace{\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}}_{[2\times3]} \cdot \underbrace{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}}_{[3\times2]} = \underbrace{\begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}}_{[2\times2]} = A$$

Diagonalize:

$$\det(A - \lambda I) = \begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} = 36 + \lambda^2 - 12\lambda - 1 = (\lambda - 7)(\lambda - 5) = 0$$

$\lambda_1 = 7$; $\lambda_2 = 5$

$$D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$$

$-x_1 + x_2 = 0$ $\rightarrow x_1 = x_2$
$x_1 - x_2 = 0$

$$\lambda_1 = 7: \quad \underbrace{\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}}_{[2\times2]} \cdot \underbrace{v_1}_{[2\times1]} = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{[2\times1]} \rightarrow v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$x_1 + x_2 = 0$ $\rightarrow x_1 = -x_2$

$$\lambda_2 = 5: \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot v_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

b) SVD of X: $\quad \underbrace{U^T U = I}_{[3\times3]}$; $\quad \underbrace{V^T V = I}_{[2\times2]}$

$$X = U \cdot D \cdot V^T$$

$$D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \\ 0 & 0 \end{pmatrix}$$

matrix A

$$X^T X = V D^T U^T U D V^T = V D^T D V^T$$

$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow$ normalized $\quad v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

$v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rightarrow \quad v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$

$$\rightarrow V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}; \quad D = \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}$$

$$X \cdot X^T = \underbrace{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}}_{[3\times2]} \times \underbrace{\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}}_{[2\times3]} = \underbrace{\begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}}_{[3\times3]}$$

$$10-5\lambda-2\lambda+\lambda^2$$

$$\begin{vmatrix} 5-\lambda & 0 & 3 \\ 0 & 5-\lambda & 1 \\ 3 & 1 & 2-\lambda \end{vmatrix} = (5-\lambda)^2(2-\lambda)+0+0-9(5-\lambda)-0-(5-\lambda) = (5-\lambda)\big((5-\lambda)(2-\lambda)-9-1\big) =$$

$$= (5-\lambda)(\lambda^2-7\lambda) = \lambda(5-\lambda)(\lambda-7)=0 \rightarrow \lambda_1=0;\ \lambda_2=5;\ \lambda_3=7$$

$\lambda_3=0$: $\begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \cdot v_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $\begin{cases} 5x_1+3x_3=0 \\ 5x_2+x_3=0 \\ 3x_1+x_2+2x_3=0 \end{cases}$ $x_3=1$, $x_2=-\frac{1}{5}$, $x_1=-\frac{3}{5}$ $\quad v_3=\begin{pmatrix} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \end{pmatrix} \xrightarrow[|v_1|=\frac{\sqrt{35}}{5}]{\text{normalizing}} v_3=\begin{pmatrix} -\frac{3}{\sqrt{35}} \\ -\frac{1}{\sqrt{35}} \\ \frac{5}{\sqrt{35}} \end{pmatrix}$

$\lambda_2=5$: $\begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix} \cdot v_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $x_3=0$, $x_2=1$, $x_1=-\frac{1}{3}$ $\quad v_2=\begin{pmatrix} -\frac{1}{3} \\ 1 \\ 0 \end{pmatrix} \xrightarrow[|v_2|=\frac{\sqrt{10}}{3}]{\text{normalizing}} v_2=\begin{pmatrix} -\frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \\ 0 \end{pmatrix}$

$\lambda_1=7$: $\begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix} \cdot v_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $\begin{cases} -2x_1+3x_3=0 \\ -2x_2+x_3=0 \\ 3x_1+x_2-5x_3=0 \end{cases}$ $x_3=1$, $x_2=\frac{1}{2}$, $x_1=\frac{3}{2}$ $\quad v_1=\begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix} \xrightarrow[|v_3|=\frac{\sqrt{14}}{2}]{\text{normalizing}} v_1=\begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix}$

SVD of X:

$$X = U\cdot D\cdot V^T = \underbrace{\begin{pmatrix} \frac{3}{\sqrt{14}} & -\frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{35}} \\ \frac{1}{\sqrt{14}} & \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{35}} \\ \frac{2}{\sqrt{14}} & 0 & \frac{5}{\sqrt{35}} \end{pmatrix}}_{=U} \cdot \underbrace{\begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}}_{=D} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{=V^T} = \underbrace{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}}_{=X}$$

Check:

$$\begin{pmatrix} \frac{3\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{3\sqrt{2}}{2} \\ \sqrt{2} & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$

c) Take first column from U, one value from D and first row from V^t:

$$\underbrace{\begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix}}_{=U_{1c}\ [3\times1]} \cdot \underbrace{(\sqrt{7})}_{=D_1\ [1\times1]} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{=V^t_{1r}\ [1\times2]} = \begin{pmatrix} \frac{3}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \end{pmatrix}$$

– matrix approximation of X

↓

its rank ≠ 1, need to reduce it:

$$\begin{pmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \end{pmatrix} \xrightarrow{[1]-[2]\cdot 3} \begin{pmatrix} 0 & 0 \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \end{pmatrix} \xrightarrow{2\cdot[2]-[3]} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{pmatrix} \to$$ matrix approximation of X with rank $=1$ is $(1 \quad 1)$

№4.

$X = UDV^T \quad d_{11} > d_{22} > \ldots > 0 \quad \|w\|^2 = 1$

d)

$$\begin{cases} \|Xw\|^2 \to \max_w \\ \text{s.t.} \\ \|w\|^2 = 1 \end{cases} \qquad \|Xw\|^2 = (Xw)^T Xw = w^T X^T X w$$

$$\|w\|^2 = w^T w$$

$$\downarrow$$

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

$$\mathcal{L} = \|Xw\|^2 - \lambda(\|w\|^2 - 1) = w^T X^T X w - \lambda(w^T w - 1)$$

e) FOC:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} = (dw)^T X^T X w + w^T X^T X (dw) - \lambda (dw)^T w - \lambda w^T (dw) = 0 & (1) \\ \frac{\partial L}{\partial \lambda} = -w^T w + 1 = 0 & (2) \end{cases}$$

(1) $\frac{\partial \mathcal{L}}{\partial w} = 2w^T X^T X dw - 2\lambda w^T dw = 2w^T(X^T X - \lambda) dw = 2w^T(X^T X - \lambda I) dw$

$$\downarrow$$

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} = 2w^T(X^T X - \lambda I) = 0 \\ \frac{\partial L}{\partial \lambda} = 1 - w^T w = 0 \end{cases} \to \begin{cases} (X^T X - \lambda I) w = 0 \\ w^T w = 1 \end{cases}$$

f) $W = V^T w \rightarrow w = VW \rightarrow (X^T X - \lambda I) w \, dw = 0$

$$d(UDW)^T (UDW) - \lambda \cdot d(VW) = 0$$

$W = V^T w$

$$2(UDW)^T d(UDW) - \lambda V dW = 0$$

$$2 W^T \underbrace{D^T \underbrace{U^T U}_{I} D}_{I} dW - \lambda V dW = 0$$

$$2 W^T dW = \lambda V dW$$

$$2 W^T = \lambda V$$

$$W^T = \frac{\lambda V}{2}$$

$$W = \left(\frac{\lambda V}{2}\right)^T$$

$$V^T w = \frac{(\lambda V)^T}{2}$$

$$\underline{w = \frac{\lambda}{2} \cdot V^T \cdot (V^T)^{-1}}$$