№1 Khejrkhabarov Elman

Gr 1

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

$$\hat y=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}\qquad \underbrace{\hat y}_{n\times 1}=X\hat\beta,\quad X=\underbrace{\begin{pmatrix}x_1&x_1\\x_2&x_2\\\vdots&\vdots\\x_n&x_n\end{pmatrix}}_{n\times 2}\quad \hat\beta=\begin{pmatrix}\hat\beta_1\\\hat\beta_2\end{pmatrix}$$

a) $d\,loss(\hat\beta)=d((y-\hat y)^T)\cdot(y-\hat y)+(y-\hat y)^T\cdot d(y-\hat y)+d(\lambda\hat\beta^T\cdot\hat\beta+\hat\beta^T d\hat\beta)=$

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

$+2\lambda\hat\beta^T d\hat\beta=-2(y-X\hat\beta)^T d(X\hat\beta)+2\lambda\hat\beta^T d\hat\beta=0\,|:d\hat\beta$ transpose of scalar equal to scalar.

$-2y^TX=-2\lambda\hat\beta^T-2\hat\beta^T\cdot X^T\cdot X=\hat\beta^T(-2\lambda I-2X^TX)$

$\cancel{\hat\beta(X^T\cdot X+\lambda I)=X^Ty}$ $\quad(X^T\cdot X+\lambda I)^{-1}\cdot X^Ty=\hat\beta$

$$\begin{pmatrix}\Sigma x_i^2+\lambda & \Sigma x_i^2\\ \Sigma x_i^2 & \Sigma x_i^2+\lambda\end{pmatrix}$$ => with $\lambda\neq 0$ $\det(X^tX+\lambda I)\neq 0$ => it is non-sing square matrix => it is invertible

Inverse of matrix is: $\dfrac{1}{\underbrace{(\Sigma x_i^2+\lambda)^2-(\Sigma x_i^2)^2}_{=\lambda^2+2\lambda\Sigma x_i^2}}\begin{pmatrix}\Sigma x_i^2+\lambda & -\Sigma x_i^2\\ -\Sigma x_i^2 & \Sigma x_i^2+\lambda\end{pmatrix}$

$$X^Ty=\begin{pmatrix}\Sigma x_iy_i\\ \Sigma x_iy_i\end{pmatrix}=\Sigma x_iy_i\begin{pmatrix}1\\1\end{pmatrix}$$

$$\hat\beta=\frac{1}{\lambda^2+2\lambda\Sigma x_i^2}\begin{pmatrix}\Sigma x_i^2+\lambda & -\Sigma x_i^2\\ -\Sigma x_i^2 & \Sigma x_i^2+\lambda\end{pmatrix}\cdot\Sigma x_iy_i\begin{pmatrix}1\\1\end{pmatrix}=\frac{\Sigma x_iy_i}{\cancel{\lambda^2}+2\cancel{\lambda}\Sigma x_i^2}\cdot\cancel{\lambda}\begin{pmatrix}1\\1\end{pmatrix}=$$

$$\begin{pmatrix}\dfrac{\Sigma x_iy_i}{\lambda+2\Sigma x_i^2}\\[2mm] \dfrac{\Sigma x_iy_i}{\lambda+2\Sigma x_i^2}\end{pmatrix}\Rightarrow \hat\beta_1=\hat\beta_2=\frac{\Sigma x_iy_i}{\lambda+2\Sigma x_i^2}$$

b) As $\lambda\to\infty$ $\hat\beta_1$ and $\hat\beta_2\to 0$

c) As $\lambda\to 0$ $\hat\beta_1$ and $\hat\beta_2\to\dfrac{\Sigma x_iy_i}{2\Sigma x_i^2}\Rightarrow \hat\beta_1+\hat\beta_2\to\dfrac{\Sigma x_iy_i\cdot 2}{2\Sigma x_i^2}$

№2

$y = X\beta + u$, $E(u|X) = 0$, $Var(u|X) = \sigma^2 \cdot W$, $\hat\beta$-OLS

$\hat\beta = (X^T \cdot X)^{-1} \cdot X^T \cdot y$, X-non-random

a) $E(\hat\beta|X) = E((X^T \cdot X)^{-1} \cdot X^T y|X) = (X^T \cdot X)^{-1} \cdot X^T \cdot E(y|X) =$

$(X^T \cdot X)^{-1} \cdot X^T \cdot E(\underbrace{X\beta}_{\text{fixed}} + u|X) = (X^T X)^{-1} X^T \cdot (X\beta + \underbrace{E(u|X)}_{=0}) = (X^T \cdot X)^{-1} X^T \cdot X \cdot \beta \equiv \beta$

$E(\hat\beta) = E(E(\hat\beta|X)) = E(\beta) = \beta$

b) $Var(\hat\beta|X) = Var((X^T X)^{-1} X^T \cdot y|X) = (X^T \cdot X)^{-1} X^T \cdot Var(y|X) \cdot ((X^T \cdot X)^{-1} \cdot X^T)^T =$

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

$\sigma^2 (X^T X)^{-1} \cdot X^T \cdot W \cdot X (X^T \cdot X)^{-1}$

c) As now W can be not identity matrix => covariance between different error terms can be non-zero => correlated, which breakes the assumption of homoscedastic and uncorrelated error terms necessary for standart CI application.

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

$Var(y|X) \cdot ((X^T \cdot X)^{-1} \cdot X^T)^T = Var(\beta X + u | X) \cdot X \cdot (X^T \cdot X)^{-1} = \sigma^2 W \cdot X (X^T \cdot X)^{-1}$

№3

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

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

$X^T X - \lambda I = \begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix}$

$\begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} = (6-\lambda)^2 - 1 = 36 - 12\lambda + \lambda^2 - 1 = \lambda^2 - 12\lambda + 35 = 0 \quad \lambda_{1,2} = 5, 7$

$\lambda = 5 \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 \end{pmatrix} \quad v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \sigma_1 = \sqrt{5} \quad \tilde{v}_1 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

$\lambda = 7 \quad \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \sigma_2 = \sqrt{7} \quad \tilde{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

$X^T \cdot X = P \cdot D \cdot P^{-1}$

$\begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

b) $u_1 = \frac{1}{\sigma_1} \cdot X \cdot \tilde{v}_1 = \frac{\sqrt{10}}{10} \cdot \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \frac{\sqrt{10}}{10} \cdot \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{\sqrt{10}}{10} \\ \frac{3\sqrt{10}}{10} \\ 0 \end{pmatrix}$

$u_2 = \frac{1}{\sigma_2} \cdot X \cdot \tilde{v}_2 = \frac{\sqrt{14}}{14} \cdot \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{3\sqrt{14}}{14} \\ \frac{\sqrt{14}}{14} \\ \frac{\sqrt{14}}{7} \end{pmatrix}$

$u_1 \cdot \sqrt{10} = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix}$; $u_2 \cdot \sqrt{14} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$

$\begin{pmatrix} -1 & 3 & 0 \\ 3 & 1 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -3 & 0 \\ 3 & 10 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -3 & 0 \\ 0 & 10 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -3 & 0 \\ 0 & 1 & 0.2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0.6 \\ 0 & 1 & 0.2 \end{pmatrix} \rightarrow$

$\tilde{u}_3 = \begin{pmatrix} -3 \\ -1 \\ 5 \end{pmatrix}$ $u_3 = \begin{pmatrix} -\frac{3}{\sqrt{35}} \\ -\frac{1}{\sqrt{35}} \\ \frac{5}{\sqrt{35}} \end{pmatrix}$

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

№4 $\quad X$ is $k\times n$, w is $n\times 1$

a) $L = \|Xw\|^2 + d(1-\|w\|^2) = \langle Xw, Xw\rangle + d(1-\langle w, w\rangle) =$
$(Xw)^T \cdot Xw - d(1 - w^t \cdot w) = w^T X^T \cdot X \cdot w + d(1 - w^T w)$

b) $\begin{cases} \frac{\partial L}{\partial w} = 0 \\ \frac{\partial L}{\partial d} = 0 \end{cases} \Rightarrow \begin{cases} dL = 0 \\ \|w\|^2 = 1 \end{cases} \Rightarrow \begin{cases} (1)\ d(w^t) \cdot X^T \cdot X \cdot w + w^t \cdot X^t \cdot X\, dw - d(dw^t w + w^t dw) = 0 \\ w^t \cdot w = 1 \end{cases}$

(1) $\overbrace{\underset{1\times n}{d(w^t)} \cdot \underset{n\times k}{X^T} \cdot \underset{k\times n}{X} \cdot \underset{n\times 1}{w}}^{\text{scalar}} + \underbrace{\underset{1\times n}{w^t} \cdot \underset{n\times k}{X^t} \cdot \underset{k\times n}{X} \cdot \underset{n\times 1}{dw}}_{\text{scalar}} - d(\underbrace{\underset{1\times n}{d(w^t)} \cdot \underset{n\times 1}{w}}_{\text{scalar}} + \underbrace{w^t dw}_{\text{scalar}}) = 0$

$2w^t X^t \cdot X \cdot dw - 2d\, w^t dw = 0$

$w^t X^t \cdot X + d\, w^t = 0 \Rightarrow w^t X^t X = d \cdot w^t \leftarrow$ transpose

$\boxed{X^t X w = d \cdot w}$

$X^t X = V \Sigma^t \Sigma V^t$

FOC: $\begin{cases} X^t \cdot X w = d w \ (1) \\ w^t w = 1 \end{cases}$

c) From (1) we can notice that w is the eigenvector of $X^t X$ with the eigenvalue d.

We also know that ~~e~~ w vector is normalised $\Rightarrow$ it is the column vector of V

So, in order to maximise $\|Xw\|^2 = w^t \underbrace{X^T X w}_{dw} = d \underbrace{w^t w}_{=1} = d$,
we need to choose eigenvector w, which corresponds to the highest eigenvalue, as $\underset{= \sqrt{d_{11}}}{d_{11}} > \underset{= \sqrt{d_{22}}}{d_{22}} > \dots > 0 \Rightarrow$ the first column of V corresponds to the highest eigenvalue $\Rightarrow$ it maximises the expression.