№1 Khejrkhabarov Elman
Gr 1

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

$\hat y = \begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}$ $\underbrace{\hat y}_{n\times1} = X\hat\beta$, $X = \underbrace{\begin{pmatrix} x_1 & x_1\\ x_2 & x_2\\ \vdots & \vdots\\ x_n & x_n\end{pmatrix}}_{n\times2}$ $\hat\beta = \begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}$

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

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

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

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

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

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

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

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

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

$\begin{pmatrix}\frac{\sum x_iy_i}{\lambda+2\sum x_i^2}\\ \frac{\sum x_iy_i}{\lambda+2\sum x_i^2}\end{pmatrix} \Rightarrow \hat\beta_1 = \hat\beta_2 = \frac{\sum x_iy_i}{\lambda+2\sum x_i^2}$

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

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

№2

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

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

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

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

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

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

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

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

c) As now W can be not identity matrix => ~~cat~~ 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) $u_1 = \frac{1}{\sigma_1} \cdot X \cdot \tilde{v}_1 = \frac{\sqrt{10}}{10} \cdot \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \frac{\sqrt{10}}{10} \cdot \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{\sqrt{10}}{10} \\ \frac{3\sqrt{10}}{10} \\ 0 \end{pmatrix}$

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

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

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

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

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

+ Code + Markdown | Run All Restart Clear All Outputs | View data

```python
import numpy as np
import pandas
```

[12] ✓ 0.0s

```python
# standartised matrix
x=np.array([[4/3,-1/3],[-5/3,2/3],[1/3,-1/3]])
# svd decomposition
p=numpy.linalg.svd(x)
u1=(p[0].T[0]).reshape(-1,1)
vh=p[2][0].reshape(1,-1)
s1=p[1][0]
# Answer
s1*u1@vh
```

[71] ✓ 0.0s

```
... array([[ 1.28551159, -0.46541939],
           [-1.68691253,  0.61074657],
           [ 0.40140095, -0.14532719]])
```

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

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

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

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

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

$w^t \cdot X^t \cdot X + \lambda w^t = 0 \Rightarrow w^t \cdot X^t \cdot X = \lambda \cdot w^t$ ← transpose

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

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

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

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

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

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