• Problem 1

$loss(\hat{\beta}) = (y-\hat{y})^T(y-\hat{y}) + \lambda\hat{\beta}^T\hat{\beta}$, $\hat{y} = (\hat{\beta}_1 + \hat{\beta}_2)x$, $\underbrace{\hat{\beta} = \begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix},\ \hat{\beta}^T = (\beta_1, \beta_2)}_{\hat{\beta}^T\hat{\beta} = \beta_1^2 + \beta_2^2}$

a) $loss(\hat{\beta}) = (y - (\hat{\beta}_1 + \hat{\beta}_2)x)^T(y - (\hat{\beta}_1 + \hat{\beta}_2)x) + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)$

since x and x are identical and equality of coefficients we can calculate $\hat{\beta}_1^2 + \hat{\beta}_2^2 = 2\hat{\beta}^2 = \frac{(\hat{\beta}_1 + \hat{\beta}_2)^2}{2}$

$loss(\hat{\beta}) = (y - (\hat{\beta}_1 + \hat{\beta}_2)x)^T(y - (\hat{\beta}_1 + \hat{\beta}_2)x) + \lambda\frac{(\hat{\beta}_1 + \hat{\beta}_2)^2}{2} =$

$= y^Ty - 2(\hat{\beta}_1 + \hat{\beta}_2)\cdot y^T\cdot x + (\hat{\beta}_1 + \hat{\beta}_2)^2 x^Tx + \lambda\cdot\frac{(\hat{\beta}_1 + \hat{\beta}_2)^2}{2} =$

$= y^Ty - 2(\hat{\beta}_1 + \hat{\beta}_2)\cdot y^T\cdot x + (\hat{\beta}_1 + \hat{\beta}_2)^2(x^Tx + \frac{\lambda}{2})$

Let's $(\hat{\beta}_1 + \hat{\beta}_2) = k \Rightarrow loss(\hat{\beta}) \xrightarrow[k]{} min$

$-2y^Tx + 2k(x^Tx + \frac{\lambda}{2}) = 0$

$k = \frac{y^Tx}{(x^Tx + \frac{\lambda}{2})}$, since we know $\hat{\beta}_1 = \hat{\beta}_2 = \frac{\hat{\beta}_1 + \hat{\beta}_2}{2}$, we get

$\hat{\beta}_1 = \hat{\beta}_2 = \frac{y^Tx}{2\cdot(x^Tx + \frac{\lambda}{2})}$

b) As we can see from k and $\hat{\beta}_1 = \hat{\beta}_2$, as $\lambda\uparrow$, in other words, as $\lambda \to \infty$ our $\hat{\beta}_1, \hat{\beta}_2 \to 0$. So, estimates tend to zero

c) $k = \hat{\beta}_1 + \hat{\beta}_2 = \frac{y^Tx}{x^Tx + \frac{\lambda}{2}}$, as $\lambda \to 0$, $k \to \frac{y^Tx}{x^Tx + 0} = \frac{y^Tx}{x^Tx}$.

So as λ tends to zero, sum $\hat{\beta}_1 + \hat{\beta}_2$ tends to $\frac{y^Tx}{x^Tx}$; it approaches classic OLS estimate for x and y.

• Problem 2

a) $\hat{\beta} = (x^Tx)^{-1}x^Ty \Rightarrow \hat{\beta} = (x^Tx)^{-1}x^T(x\beta + u) = \underbrace{(x^Tx)^{-1}x^Tx}\beta + (x^Tx)^{-1}x^Tu =$

$= \beta + (x^Tx)^{-1}x^Tu \Rightarrow E(\beta + (x^Tx)^{-1}x^Tu \mid x) = E(\beta \mid x) + (x^Tx)^{-1}x^T\cdot \underbrace{E(u \mid x)}_{0} =$

$= E(\beta \mid x) + 0 = E(\beta \mid x) = \beta \Rightarrow \underline{\underline{E(\hat{\beta} \mid x) = \beta}}$

As for $E(\hat{\beta})$, knowing "Tower's rule" from lectures: $E(\hat{\beta}) = E(E(\hat{\beta} \mid x)) = E(\beta) = \beta$

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

b) Since $\hat{\beta}=\beta+(X^TX)^{-1}X^Tu$ from a) and $Var(Ax)=A\,Var(x)\cdot A^T$,

we get: $Var(\hat{\beta}|X)=Var(\underbrace{\beta}_{const}+(X^TX)^{-1}X^Tu|X)=Var((X^TX)^{-1}X^Tu|X)=$

$$=(X^TX)^{-1}X^T\cdot Var(u|X)\cdot((X^TX)^{-1}X^T)^T=\left\{\begin{array}{l}((X^TX)^{-1}X^T)^T=(X^T)^T\cdot((X^TX)^{-1})^T=\\ =X\cdot(X^TX)^{-1}\end{array}\right\}=$$

$=(X^TX)^{-1}X^T\,Var(u|X)\cdot X\cdot(X^TX)^{-1}$, since we know $Var(u|X)=\sigma^2W$,

we get: $Var(\hat{\beta}|X)=(X^TX)^{-1}\cdot X^T\cdot\sigma^2\cdot W\cdot X\cdot(X^TX)^{-1}$

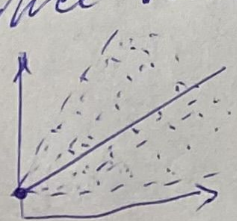
c) I think standard CI for β will not be valid in this case, since $Var(u|X)=\sigma^2W$, where $W\neq I$, meaning u (disturbance term) is heteroscedastic, which fails ~~the assum~~ one of 6 assumptions of Gauss-Markov BLUE estimators of model A with non-stochastic regressors.

The 6 assumptions are:

1) Regression must be linear in parameteres and correctly specified
2) There's some variation in regressor
3)* Disturbance term is homoscedastic – fails
4) Disturbance term has zero expectation
5) The values of disturbance term have independent distribution
6) The disturbance term has normal distribution

so, for point 3) $E((u_i-\mu_u)^2)$ must be equal to $E(u_i^2-2u_i\mu_u+\mu_u^2)$ $=E(u_i^2)-\mu_u^2\overset{?}{=}E(u_i^2)=\sigma_u^2$, where $E(u_i)=\mu_u$, $\mu_u=0$

since $Var(u|X)=\sigma^2\cdot W$ – it's heteroscedastic.



d) $cov(y, \hat{\beta}|X) = E(y \cdot \hat{\beta}|X) - E(y|X) \cdot E(\hat{\beta}|X) = E((X\beta+u) \cdot \hat{\beta}|X) -$
$- E(X\beta + u|X) \cdot E(\hat{\beta}|X) = E((X\beta+u) \cdot \hat{\beta}|X) - (X\beta + \underbrace{E(u|X)}_{0}) \cdot E(\hat{\beta}|X) =$
$= E((X\beta+u) \cdot \hat{\beta}|X) - X\beta \cdot E(\hat{\beta}|X)$

From previous points we know $E(\hat{\beta}|X) = \beta$, $\hat{\beta} = \beta + (X^TX)^{-1}X^Tu$, so

$cov(y, \hat{\beta}|X) = E(X\beta \cdot \hat{\beta} + u \cdot \hat{\beta}|X) - X\beta \cdot \beta = X\beta^2 + \underbrace{E(u \cdot (\beta + (X^TX)^{-1}X^Tu)|X)}_{0, \text{ since } E(u|X)=0} -$
$- X\beta^2 = \cancel{X\beta^2} + 0 - \cancel{X\beta^2} = 0$

So, $cov(y, \hat{\beta}|X) = 0$

Problem 3

a)
$X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}_{3\times2}$, $X^T = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}_{2\times3}$

$\underset{2\times2}{X^TX} = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$, so $\underset{2\times2}{X^TX} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$

Diagonalizing:

$\det\begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} = (6-\lambda)(6-\lambda) - 1 = (6-\lambda)^2 - 1$

$(6-\lambda)^2 - 1 = 0$, $6 - \lambda = \pm 1 \Rightarrow \underbrace{\lambda_1 = 7, \lambda_2 = 5}_{\text{eigenvalues}}$, so $D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$

for $\lambda_1 = 7$:

$(A - 7I) = \begin{pmatrix} 6-7 & 1 \\ 1 & 6-7 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ $-x + y = 0$
$-x = -y$
$x = y$

V_1 - eigenvalue, $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

for $\lambda_2 = 5$:

$(A - 6I) = \begin{pmatrix} 6-5 & 1 \\ 1 & 6-5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ $x + y = 0$
$x = -y$

$V_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$R = A \cdot D \cdot A^{-1} \Rightarrow R = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} \cdot \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$

$A = (V_1, V_2) \Rightarrow A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, $A^{-1} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}$

b)

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}, \quad A^T = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \Rightarrow A \cdot A^T = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

$$|A \cdot A' - \lambda I| = 0 \Rightarrow \begin{vmatrix} 5-\lambda & 0 & 3 \\ 0 & 5-\lambda & 1 \\ 3 & 1 & 2-\lambda \end{vmatrix} = 0 \Rightarrow$$ from this we get

$$-\lambda^3 + 12\lambda^2 - 35\lambda = 0 \Rightarrow \lambda_1 = 0,\ \lambda_2 = 5,\ \lambda_3 = 7$$

for $\lambda_1 = 7$, we get: $v_1 = \begin{pmatrix} 3/2 \\ 1/2 \\ 1 \end{pmatrix}$

for $\lambda_2 = 5$, we get: $v_2 = \begin{pmatrix} -1/3 \\ 1 \\ 0 \end{pmatrix}$

for $\lambda_3 = 0$, we get: $v_3 = \begin{pmatrix} -0{,}6 \\ -0{,}2 \\ 1 \end{pmatrix}$

$L_1 = \sqrt{1{,}5^2 + 0{,}5^2 + 1^2} = 1{,}87$

$L_2 = \sqrt{\left(-\frac{1}{3}\right)^2 + 1^2 + 0^2} = 1{,}054$

$L_3 = \sqrt{(-0{,}6)^2 + (0{,}2)^2 + 1^2} = 1{,}1832$

$u_1 = (0{,}80;\ 0{,}26;\ 0{,}53)$

$u_2 = (-0{,}31;\ 0{,}95;\ 0)$

$u_3 = (-0{,}50;\ -0{,}17;\ 0{,}84)$

$$V = \begin{pmatrix} 0{,}707 & -0{,}707 \\ 0{,}707 & 0{,}707 \end{pmatrix} \text{ or } V = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$U = \begin{pmatrix} 0{,}80 & 0{,}26 & 0{,}53 \\ -0{,}31 & 0{,}95 & 0 \\ -0{,}50 & -0{,}17 & 0{,}84 \end{pmatrix}$$

$$U \cdot S \cdot V^T = \begin{pmatrix} 0{,}80 & 0{,}26 & 0{,}53 \\ -0{,}31 & 0{,}95 & 0 \\ -0{,}50 & -0{,}17 & 0{,}84 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0{,}707 & 0{,}707 \\ -0{,}707 & 0{,}707 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$

c) Using Python I got best rank 1 approximation as:

$$B = \begin{pmatrix} 1{,}5 & 1{,}5 \\ 0{,}5 & 0{,}5 \\ 1 & 1 \end{pmatrix}$$