DSE HA № 3 Shvachko Elizaveta (use all 4 горшочков)

Problem № 1:

loss f-n: $loss(\hat{\beta}) = (y - \hat{y})^T (y - \hat{y}) + \lambda \hat{\beta}^T \hat{\beta}$, $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_2 x$

$\lambda \geq 0$

a) $\hat{\beta}_1$ & $\hat{\beta}_2$ for fixed x

as regressors x and x are identical $\Rightarrow$ their contributions to $\hat{y}$ are the same

$$\hat{y} = \hat{\beta}_1 x + \hat{\beta}_2 x = \underbrace{(\hat{\beta}_1 + \hat{\beta}_2)}_{= \beta} x = \beta x \quad \Rightarrow$$

$$\Rightarrow \quad loss(\beta) = \underbrace{(y - \beta x)^T (y - \beta x)}_{\text{how well model fits data}} + \underbrace{\lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)}_{\text{penalty term}}$$

As $\hat{\beta}_1 + \hat{\beta}_2 = \beta \Rightarrow \hat{\beta}_1 = \hat{\beta}_2 = \frac{\beta}{2} \Rightarrow \hat{\beta}_1^2 + \hat{\beta}_2^2 = \left(\frac{\beta}{2}\right)^2 + \left(\frac{\beta}{2}\right)^2 = \frac{\beta^2 + \beta^2}{4} = \frac{2\beta^2}{4} = \frac{\beta^2}{2}$

$$(y - \beta x)^T (y - \beta x) = y^T y - 2\beta y^T x + \beta^2 x^T x$$

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

min loss $\Rightarrow$ take derivative $loss(\beta)'_\beta = 0$

$$loss(\beta)'_\beta = -2y^T x + 2\beta x^T x + \lambda\beta = 0$$

$$\beta(2x^T x + \lambda) = 2y^T x$$

$$\boxed{\beta = \frac{2y^T x}{2x^T x + \lambda}}$$

$$\Rightarrow \boxed{\hat{\beta}_1 = \hat{\beta}_2 = \frac{\beta}{2} = \frac{y^T x}{2x^T x + \lambda}}$$

b) $\lambda \to \infty$

As $\uparrow \lambda \Rightarrow [\lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2)] \uparrow \Rightarrow \hat{\beta}_1$ & $\hat{\beta}_2$ shrinks to 0 to min penalty

As $\hat{\beta}_1 = \hat{\beta}_2 = \dfrac{y^T x}{2x^T x + \lambda}$

When $\lambda \to \infty \Rightarrow (2x^T x + \lambda) \to \infty \Rightarrow \hat{\beta}_1 \to 0$ & $\hat{\beta}_2 \to 0$

So, to avoid overfitting in ridge regression large λ hugely penalizes the coefficients

c) $\lambda \to 0$; $\hat{\beta}_1 + \hat{\beta}_2$ - ?

when $\lambda = 0 \Rightarrow$ regularization dissapears $\Rightarrow$ model reduces to OLS

$$\Rightarrow \hat{\beta}_1 + \hat{\beta}_2 = \beta = \frac{\cancel{2} y^T x}{\cancel{2} x^T x} = \frac{y^T x}{x^T x}$$

Problem №2

$$y = X\beta + u$$

a) $E(\hat{\beta}|X)$ & $E(\hat{\beta})$ - ?

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

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

$$\Rightarrow \hat{\beta} = \underbrace{\boxed{(X^T X)^{-1} X^T X}}_{\substack{\text{identity} \\ \text{matrix } (I)}} \beta + (X^T X)^{-1} X^T u = \beta + (X^T X)^{-1} X^T u$$

1) $E(\hat{\beta}|X) = E(\beta + (X^T X)^{-1} X^T u \mid X) = \underbrace{E(\beta|X)}_{\beta\text{, as } \beta \text{ - non - random}} + E((X^T X)^{-1} X^T u|X) =$

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

2) $E(\hat{\beta}) = E(\beta + (X^T X)^{-1} X^T u) \underset{\text{Tower prop.}}{=} E(\beta) + \underbrace{E(X^T X)^{-1} X^T u)}_{0} =$

$= \beta$

As $E(\hat{\beta}) = \beta \Rightarrow$ OLS is unbiased estimator of β

(b) $Var(\hat{\beta}|X)$ – ?

$\hat{\beta} = \underset{\text{non-random}}{\beta} + (X^T X)^{-1} X^T u \Rightarrow$

$\Rightarrow Var(\hat{\beta}|X) = Var((X^T X)^{-1} X^T u)$

As $u \sim N(0, \sigma^2 W) \Rightarrow Var(u|X) = \sigma^2 W$

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

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

(c) Standard confidence interval assumes $\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$

$\Rightarrow W = I$, but in this case $W \neq I$

from standard confidence intervals: $Var(\hat{\beta}|X) = \sigma^2 (X^T X)^{-1}$

from (b): $Var(\hat{\beta}|X) = \sigma^2 (X^T X)^{-1} X^T \underline{\underline{W}} X (X^T X)^{-1}$

As $W \neq I \Rightarrow$ variance is not constant across observ.

$\Rightarrow$ assumptions of st. confidence intervals are violated

Ignoring of W leads to misestimation of standard errors of $\hat{\beta}$ $\Rightarrow$ invalid F-tests & t-tests $\Rightarrow$ as standard confidence intervals rely on homoscedesticity ($W = I$) they are not valid

(d) $Cov(y, \hat{\beta}|X) = E[(y - \underbrace{E(y|X)}_{X\beta})(\hat{\beta} - \underbrace{E(\hat{\beta}|X)}_{\beta})^T | X]$

$Cov(y, \hat{\beta}|X) = E[\underbrace{(y - X\beta)}_{y = X\beta + u \Rightarrow u}\underbrace{(\hat{\beta} - \beta)}_{(X^T X)^{-1} X^T u}^T | X] \Rightarrow$

$\Rightarrow Cov(y, \hat{\beta}|X) = E[u \cdot \underbrace{((X^T X)^{-1} X^T u)^T}_{u^T X (X^T X)^{-1}} | X] = E[u \cdot u^T \underbrace{X (X^T X)^{-1}}_{\text{not depend on } u} | X] =$

$= E(u \cdot u^T | X) \cdot X (X^T X)^{-1}$

$Var(u|X) = E(u \cdot u^T | X) = \sigma^2 W \Rightarrow Cov(y, \hat{\beta}|X) = \sigma^2 W X (X^T X)^{-1}$

№3

As $X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$, then $X^T = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}$

$$XX^T = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \end{pmatrix} = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

$n_{11} = 2\cdot 2 + 1\cdot 1 = 5$

$n_{12} = 2\cdot(-1) + 1\cdot 2 = 0$

$n_{13} = 2\cdot 1 + 1\cdot 1 = 3$

$n_{21} = (-1)2 + 2\cdot 1 = 0$

$n_{22} = (-1)\cdot(-1) + 2\cdot 2 = 4$

$n_{23} = (-1)\cdot 1 + 2\cdot 1 = 1$

$n_{31} = 1\cdot 2 + 1\cdot 1 = 3$

$n_{32} = (-1)\cdot 1 + 2\cdot 1 = 1$

$n_{33} = 1\cdot 1 + 1\cdot 1 = 2$

$$X^T X = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$$

$n_{11} = (2\cdot 2) + ((-1)\cdot(-1)) + (1\cdot 1) = 6$

$n_{12} = (2\cdot 1) + ((-1)\cdot 2) + (1\cdot 1) = 1$

$n_{21} = (1\cdot 2) + (2\cdot(-1)) + (1\cdot 1) = 1$

$n_{22} = (1\cdot 1) + (2\cdot 2) + (1\cdot 1) = 6$

Diagnolize $X^T X$:

- Eigenvalues:

$\det(X^T X - \lambda I) = 0 \Rightarrow \det\begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} = 0$

$(6-\lambda)(6-\lambda) - 1\cdot 1 = \lambda^2 - 12\lambda + 35 = 0$

$D = 12^2 - 4\cdot 35 = 144 - 140 = 4$

$\lambda = \dfrac{12 \pm 2}{2}$

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

8 for $\lambda = 7$: $(X^T X - 7 \cdot I) \cdot v = 0$

$$\begin{pmatrix} 6-7 & 1 \\ 1 & 6-7 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \Rightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} v_1 - v_2 = 0 \\ -v_1 + v_2 = 0 \end{cases} \Rightarrow v_1 = v_2 \Rightarrow v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

devide v by st. deviation to normalize $\Rightarrow V_n = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$

for $\lambda = 5$: $(X^T X - 5 \cdot I) \cdot v = 0$

$$\begin{pmatrix} 6-5 & 1 \\ 1 & 6-5 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \Rightarrow \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} v_1 + v_2 = 0 \\ v_1 + v_2 = 0 \end{cases} \Rightarrow v_1 = -v_2 \Rightarrow v = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

devide v by st. deviation to normalize $\Rightarrow V_n = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$

$$V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$ where columns of V are the normalized eigenvectors of $X^T X$ $\Leftrightarrow$

$$\Leftrightarrow V^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

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

(б) SVD: $X = U\Sigma V^T \Rightarrow X^T X = (U\Sigma V^T)' U\Sigma V^T = V\Sigma^T \underbrace{U^T U}_{\text{identity matrix}} \Sigma V^T = V\Sigma^T\Sigma V^T$

$XX^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T$

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

Analogically with (a):

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

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

$$= (5-\lambda)(\lambda^2 - 7\lambda + 9) + 3(-15 + 3\lambda) = -\lambda^3 + 12\lambda^2 - 44\lambda + 45 -$$

$$-45 + 9\lambda = \lambda^3 + 12\lambda^3 - 35\lambda = 0 \Rightarrow -\lambda(\lambda^2 - 12\lambda + 35) = 0$$

$$\Rightarrow \left[\begin{array}{l} \lambda = 0 \\ \lambda^2 - 12\lambda + 35 = 0 \end{array}\right.$$

$$D = (-12)^2 - 4\cdot 35 = 144 - 140 = 4$$

$$\lambda = \frac{12 \pm 2}{2}$$

$$\left[\begin{array}{l} \left[\begin{array}{l} \lambda = 7 \\ \lambda = 5 \end{array}\right. \\ \lambda = 0 \end{array}\right. \Rightarrow D = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

(1) for $\lambda = 7$: $(X^T X - 7\cdot I)\cdot u = 0$

$$\begin{pmatrix} 5-7 & 0 & 3 \\ 0 & 5-7 & 1 \\ 3 & 1 & 2-7 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = 0 \Rightarrow$$

$$\Rightarrow \begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = 0$$

$$\begin{cases} -2v_1 + 0 + 3v_3 = 0 \\ 0 - 2v_2 + v_3 = 0 \\ 3v_1 + v_2 - 5v_3 = 0 \end{cases} \Rightarrow v = \begin{pmatrix} 3/2 \\ 1/2 \\ 1 \end{pmatrix}$$

to normalize
devide v by st.deviation $\Rightarrow$ $v_n = \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}$

$$\sqrt{\left(\frac{3}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + 1^2} = \sqrt{\frac{14}{4}} = \frac{\sqrt{14}}{2}$$

2) for $\lambda = 5$: $(X^T X - 5 \cdot I) \cdot v = 0$

$$\begin{pmatrix} 5-5 & 0 & 3 \\ 0 & 5-5 & 1 \\ 3 & 1 & 2-5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Rightarrow$$

$$\Rightarrow \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0$$

$$\begin{cases} 0 + 0 + 3v_3 = 0 \\ 0 + 0 + v_3 = 0 \\ 3v_1 + v_2 - 3v_3 = 0 \end{cases} \Rightarrow \begin{cases} v_3 = 0 \\ 3v_1 + v_2 - 3v_3 = 0 \end{cases} \Rightarrow \begin{cases} v_3 = 0 \\ 3v_1 + v_2 = 0 \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} v_3 = 0 \\ v_2 = -3v_1 \end{cases} \Rightarrow v = \begin{pmatrix} -1/3 \\ 1 \\ 0 \end{pmatrix}$$

to normalize
devide v by st.deviation $\Rightarrow v_n = \begin{pmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \\ 0 \end{pmatrix}$

$$\sqrt{1^2 + (-3)^2 + 0^2} = \sqrt{10}$$

3) for $\lambda = 0$: $(X^T X - 0 \cdot I) \cdot v = 0$

$$\begin{pmatrix} 5-0 & 0 & 3 \\ 0 & 5-0 & 1 \\ 3 & 1 & 2-0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \Rightarrow$$

$$\Rightarrow \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = 0$$

$$\begin{cases} 5u_1 + 0 + 3u_3 = 0 \\ 0 + 5u_2 + u_3 = 0 \\ 3u_1 + u_2 + 2u_3 = 0 \end{cases} \Rightarrow \begin{cases} u_1 = -3 \\ u_2 = -1 \\ u_3 = 5 \end{cases} \Rightarrow u = \begin{pmatrix} -3 \\ -1 \\ 5 \end{pmatrix}$$

to normalize
devide v by st. deviation $\Rightarrow u_n = \begin{pmatrix} -3/\sqrt{35} \\ -1/\sqrt{35} \\ 5/\sqrt{35} \end{pmatrix}$

$\sqrt{5^2 + (-3)^2 + (-1)^2} = \sqrt{35}$

$$\Rightarrow XX^T = \begin{pmatrix} 3/\sqrt{14} & 1/\sqrt{10} & -3/\sqrt{35} \\ 1/\sqrt{14} & -3/\sqrt{10} & -1/\sqrt{35} \\ 2/\sqrt{14} & 0 & 5/\sqrt{35} \end{pmatrix} \begin{pmatrix} 7 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3/\sqrt{14} & 1/\sqrt{14} & 2/\sqrt{14} \\ 1/\sqrt{10} & -3/\sqrt{10} & 0 \\ -3/\sqrt{35} & -1/\sqrt{35} & 5/\sqrt{35} \end{pmatrix}$$

$$\Rightarrow \Sigma = \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}$$

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

c) The best rank-1 approximation of X is obtained from its SVD : $X = U \Sigma V^T$

rank-1 approxy : $X_1 = \sigma_1 u_1 v_1^T$

$\sigma_1 = \sqrt{7}$

$v_1^T = \frac{1}{\sqrt{2}} (1 \;\; 1) = \left(\frac{1}{\sqrt{2}} \;\; \frac{1}{\sqrt{2}} \right)$

$u_1 = \frac{1}{\sqrt{14}} \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}$

$$X_1 = \sqrt{7} \cdot \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix} \left(\frac{1}{\sqrt{2}} \;\; \frac{1}{\sqrt{2}} \right) = \begin{pmatrix} 3/2 & 3/2 \\ 1/2 & 1/2 \\ 1 & 1 \end{pmatrix}$$