Home Assignment 3.

Kupreeva Ekaterina
group 2.

№1.

$\text{loss}(\hat\beta) \to \min_{\hat\beta}$; $\hat\beta = \begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}$, $\hat y = \hat\beta_1 x + \hat\beta_2 x = \hat\beta^T \begin{pmatrix}x\\ x\end{pmatrix} =$

$X = \begin{pmatrix}x\\ x\end{pmatrix}$ $\qquad = (\hat\beta_1 + \hat\beta_2)x = \hat\beta^T\begin{pmatrix}1\\ 1\end{pmatrix}\cdot x.$

a) $\text{loss}(\hat\beta) = (y - \hat\beta_1 x - \hat\beta_2 x)^T (y - \hat\beta_1 x - \hat\beta_2 x) + \lambda \hat\beta^T \hat\beta =$

$= (y - X\hat\beta)^T (y - X\hat\beta) + \lambda\hat\beta^T\hat\beta.$

$d\text{loss}(\hat\beta) = 2(y - X\hat\beta)^T d(y - X\hat\beta) + 2\lambda\hat\beta^T d\hat\beta =$

$= -2(y - X\hat\beta)^T X\, d\hat\beta + 2\lambda\hat\beta^T d\hat\beta =$

$= [-2(y - X\hat\beta)^T X + 2\lambda\hat\beta^T]\, d\hat\beta = 0 \to$

$\to -(y - X\hat\beta)^T X + \lambda\hat\beta^T = 0 \to \lambda\hat\beta^{TT} = X^T(y - X\hat\beta) = X^T y - X^T X\hat\beta \to$

$\to (\lambda - X^T X)\hat\beta = X^T y \to \hat\beta^* = (\lambda - X^T X)^{-1} X^T y$ - optimal values of $\hat\beta_1$ & $\hat\beta_2$ is a linear combination

b) $\lim_{\lambda\to\infty} \hat\beta^* = \lim_{\lambda\to\infty} (\lambda - X^T X)^{-1} X^T y = 0 \to (\hat\beta_1, \hat\beta_2) = (0, 0).$

It helps to overcome multicollinearity caused by identical predictors.

c) $\lim_{\lambda\to 0} \hat\beta^* = \lim_{\lambda\to 0} (\lambda - X^T X)^{-1} X^T y = \beta^{OLS} \to$ The sum of estimators will converge to some constant depending on y. It is not unique since there are two identical predictors.

№2.

$y = X\beta + u$, $\mathbb{E}(u|X) = 0$, $X - [n \times k]$, $rk(X) = k$, $Var(u|X) = \sigma^2 W$, $W \neq I$.

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

a) $\mathbb{E}(\hat{\beta}|X) = \mathbb{E}(\underbrace{(X^T X)^{-1} X^T}_{\text{known}} y \,|\, X) = (X^T X)^{-1} X^T \mathbb{E}(X\beta + u|X) =$

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

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

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

$= (X^T X)^{-1} X^T \underbrace{Var(\underbrace{X\beta}_{\text{const}} + u|X)}_{= Var(u|X)} X (X^T X)^{-1} = (X^T X)^{-1} X^T \sigma^2 W X (X^T X)^{-1} =$

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

c) In this case, the standard confidence interval for β is not valid. Since $Var(u|X) = \sigma^2 W$, $W \neq I$, errors are heteroscedastic. Thus, we can't use standard formulas of CI. Error terms ~~are not~~ don't have constant variance & may exhibit correlation across observations.

d) $cov(y, \hat{\beta}|X) = cov(y, (X^T X)^{-1} X^T y \,|\, X) = cov(y, y|X)(X^T X)^{-1} X^T =$

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

N3.

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

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

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

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

$$\rightarrow \quad 36 - 12\lambda + \lambda^2 - 1 = \lambda^2 - 12\lambda + 35 = (\lambda - 5)(\lambda - 7) = 0$$

$\lambda_1 = 7$, $\lambda_2 = 5$. – eigenvalues. Thus, diagonal is : $D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$

$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ – eigenvector 1

– $\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \rightarrow -1a + 1b = 0$

$v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ – eigenvector 2

– $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \rightarrow 1a + 1b = 0$

b) SVD for X : $X = U \cdot D \cdot V^T$ s.t. $\underset{\substack{[n\times n] \\ = [3\times 3]}}{UU^T = I}$ & $\underset{\substack{[k\times k] \\ = [2\times 2]}}{V^TV = I}$

$$X^T X = V D^T \underbrace{U^T U}_{I} D V^T = V D^T D V^T$$

- $v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$

From a: $D = \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}$; Normalizing eigenvectors:

- $v_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$

$$V^* = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$XV = U \cdot D : \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 3/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 3/\sqrt{2} \\ 2/\sqrt{2} & 0 \end{pmatrix} \rightarrow U = \begin{pmatrix} 3/\sqrt{14} & -1/\sqrt{10} \\ 1/\sqrt{14} & 3/\sqrt{10} \\ 2/\sqrt{14} & 0 \end{pmatrix}$$

c) $X_{approx} = \underbrace{\underset{[3\times 1]}{U^*} \cdot \underset{[1\times 1]}{D^*} \cdot \underset{[1\times 2]}{(V^T)^*}}_{[3\times 2]}$

$$X = \begin{pmatrix} 3/\sqrt{14} & -1/\sqrt{10} \\ 1/\sqrt{14} & 3/\sqrt{10} \\ 2/\sqrt{14} & 0 \end{pmatrix} \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & +1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$U^* = \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}; \quad D^* = (\sqrt{7}); \quad (V^T)^* = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$X_{approx} = \begin{pmatrix} 3/\sqrt{2} \\ 1/\sqrt{2} \\ 2/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 3/2 & 3/2 \\ 1/2 & 1/2 \\ 2/2 & 2/2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix} \rightarrow rk = 1.$$

№4.

a) $\begin{cases} \|Xw\|^2 \to \max\limits_{w} \\ \text{s.t.} \\ \|w\|^2 = 1 \end{cases}$ $\qquad\qquad$ $\begin{cases} U^T U = I \\ D^* D = I \\ V^T V = I \end{cases}$

$$\mathcal{L} = \|Xw\|^2 - \lambda[\|w\|^2 - 1] = \|UDV^T w\|^2 - \lambda[\|w\|^2 - 1]$$

b) FOC: $d\big((UDV^T w)(UDV^T w)^T\big) - \lambda\, d(ww^T) =$

$= 2(UDV^T w)^T d(UDV^T w) - 2\lambda w^T dw$

$= 2w^T V D^T \underbrace{U^T U}_{I} D V^T dw - 2\lambda w^T dw =$

$= 2w^T dw - 2\lambda w^T dw = 2w^T dw(1-\lambda) = 0$

$\to$ ~~$2\Sigma w_i$~~ $\Leftrightarrow \lambda = 1, \quad \Sigma w_i = 1.$

c) $W = V^T w \;\to\; w = VW$

$\Rightarrow$ FOC: $d(UDW)(UDW)^T - \lambda \cdot d(VW) =$

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

$= 2W^T D^T U^T U D\, dW - \cancel{2}\lambda V dW =$

$= 2W^T dW - \cancel{2}\lambda V dW = 0$

$\to 2W^T = \lambda V \;\to\; W^T = \dfrac{\lambda V}{2}.$

$W^T = w^T V = \dfrac{\lambda V}{2} \;\to\; \underline{w^T = \dfrac{\lambda}{2} V V^{-1}}.$