Maxim Ivanov №1 group 1

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

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

$$= -2(y-\hat{y})^T \cdot d\hat{y} + 2\lambda\hat{\beta}^T \cdot d\hat{\beta} = -2(y-\hat{y})^T \cdot (X \cdot d\hat{\beta} + \overset{=0}{dX} \cdot \hat{\beta}) +$$

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

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

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

F.O.C. : $2(\hat{\beta}^T(X^T X + 2\lambda I) - y^T X) = 0$

$$\hat{\beta}^T(X^T X + 2\lambda I) - y^T X = 0$$

$$(\hat{\beta}^T(X^T X + 2\lambda I))^T = (y^T X)^T$$

$$(X^T X + 2\lambda I)\hat{\beta} = X^T y$$

$$\hat{\beta} = (X^T X + 2\lambda I)^{-1} X^T y$$ if $(X^T X + 2\lambda I)^{-1}$ exists

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

b) As $\lambda \to \infty$, estimates $\hat{\beta} \to 0$. High value coefficients shrink at a greater rate than low value coefficients.

c) As $\lambda \to 0$, sum $\hat{\beta}_1 + \hat{\beta}_2$ increases
When $\lambda = 0$, the ridge estimator reduces to ordinary least squares (OLS) as the constraint is no longer binding. So with $\lambda \to 0$ $\hat{\beta}_1 + \hat{\beta}_2 \to$
$\to \hat{\beta}_1^{OLS} + \hat{\beta}_2^{OLS}$ (increases untill reaches the sum of OLS).

N2

$y = X\beta + u$, β — non-random

$E(u|X) = 0$ $\quad X_{[n \times k]}$ $\quad \operatorname{rank} X = k$

$Var(u|X) = \sigma^2 W$, $W \neq I$ $\quad \hat{\beta}$ — standard OLS est. of β.

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

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

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

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

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

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

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

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

$= (X^T X)^{-1} X^T \, Var(y|X) \, [(X^T X)^{-1} X^T]^T = (X^T X)^{-1} X^T \, Var(X\beta + u | X) \cdot$

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

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

c) $CI(\beta) = \hat{\beta} \pm Z_{0.025} \sqrt{Var(\hat{\beta}|X)}$

I think that it will be valid in this case.

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

$= Cov(u, y | X) [(X^T X)^{-1} X^T] = Cov(u, u | X) (X^T X)^{-1} X^T =$

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

№3

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

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

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

$$= \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$$

$$|X^T X - \lambda I| = \begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} = (6-\lambda)^2 - 1 =$$

$$= (6-\lambda-1)(6-\lambda+1) = (5-\lambda)(7-\lambda) = 0$$

$\lambda_1 = 5 \quad \lambda_2 = 7$

$\lambda_1 = 5$: $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 \end{pmatrix} \quad v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$\lambda_2 = 7$: $\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$D=\begin{pmatrix}5&0\\0&7\end{pmatrix} \qquad P=\begin{pmatrix}v_1 & v_2\end{pmatrix}=\begin{pmatrix}1&1\\-1&1\end{pmatrix}$$

$$X^TX=PDP^{-1} \qquad P^{-1}=\frac{1}{\det(P)}\begin{pmatrix}1&-1\\1&1\end{pmatrix}=\frac{1}{2}\begin{pmatrix}1&-1\\1&1\end{pmatrix}$$

$$X^TX=\begin{pmatrix}1&1\\-1&1\end{pmatrix}\begin{pmatrix}5&0\\0&7\end{pmatrix}\begin{pmatrix}1/2&-1/2\\1/2&1/2\end{pmatrix}$$

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

$$|XX^T-\lambda I|=\begin{vmatrix}5-\lambda&0&3\\0&5-\lambda&1\\3&1&2-\lambda\end{vmatrix}=(5-\lambda)^2(2-\lambda)-9(5-\lambda)-(5-\lambda)=$$

$$=-\lambda^3+12\lambda^2-35\lambda=$$

$$=-\lambda(5-\lambda)(7-\lambda)$$

$\lambda_1=0 \quad \lambda_2=5 \quad \lambda_3=7$

$\lambda_1=0$: $\begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 3/5 \\ 0 & 1 & 1/5 \\ 0 & 1 & 1/5 \end{pmatrix} \to$

$\to \begin{pmatrix} 1 & 0 & 3/5 \\ 0 & 1 & 1/5 \end{pmatrix}$ $u_1 = \frac{1}{\sqrt{35}} \begin{pmatrix} 3 \\ 1 \\ -5 \end{pmatrix}$

$\lambda_2=5$: $\begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 1/3 & -1 \\ 0 & 0 & 1 \end{pmatrix} \to$

$\to \begin{pmatrix} 1 & 1/3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ $u_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ -3 \\ 0 \end{pmatrix}$

$\lambda_3=7$: $\begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -3/2 \\ 0 & 1 & -1/2 \\ 0 & 1 & -1/2 \end{pmatrix} \to$

$\to \begin{pmatrix} 1 & 0 & -3/2 \\ 0 & 1 & -1/2 \end{pmatrix}$ $u_3 = \frac{1}{\sqrt{14}} \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$

$$U = \begin{pmatrix} 3/\sqrt{35} & 1/\sqrt{10} & 3/\sqrt{14} \\ 1/\sqrt{35} & -3/\sqrt{10} & 1/\sqrt{14} \\ -5/\sqrt{35} & 0 & 2/\sqrt{14} \end{pmatrix}$$

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

$$X^T X = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} \quad \text{see (a)}$$

$$v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

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

$$X = UDV^T$$

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

c) $$X \approx \hat{X} = \begin{pmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \\ 0 \end{pmatrix} \begin{pmatrix} \sqrt{5} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} =$$

$$= \begin{pmatrix} 1/2 & -1/2 \\ -3/2 & 3/2 \end{pmatrix}$$

N4

$$\|Xw\|^2 \to \max_w \quad , \|w\|^2 = 1$$

a) $L = (Xw)^T(Xw) + \lambda(I - w^T w)$

b) $dL = 2(Xw)^T d(Xw) + \overbrace{\lambda\, dI}^{=0} - 2\lambda w^T dw =$

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

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

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

F.O.C

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

$$w^T(X^T X - \lambda) = 0$$

$$w^T X^T X - \lambda w^T = 0$$

$$(X^T X w)^T = (\lambda w)^T$$

$$X^T X w = \lambda w$$

w - eigenvector

λ - eigenvalue

c) $X^T X w = \lambda w$

w - eigenvector of $X^T X$

$V = (v_1, v_2, \dots, v_n)$, where columns are eigenvectors of $X^T X$

So, to maximize $\|Xw\|^2$ choose w equal to the column of V corresponding to highest eigenvalue. As $||Xw||^2 \to \max$

$$\Updownarrow$$

$||\lambda w||^2 \to \max$

$||\lambda w||^2 \to \max$ when λ is max.

So, choose $w = v_1$ (v_1 corresponds to d_{11})

($X^T X v_1 = d_{11} v_1$) v_1 - first coleumn of V

$V = (v_1, v_2, \ldots)$

Answer: $w = v_1$