Omarov Timur | DS for Economics | HW_3

№1

(a) $loss(\hat\beta) = [y - (\hat\beta_1 + \hat\beta_2)x]^T [y - (\hat\beta_1+\hat\beta_2)x] + \lambda \cdot \begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}^T \begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} =$

$$= [y^T - (\hat\beta_1+\hat\beta_2)x^T][y - (\hat\beta_1+\hat\beta_2)x] + \lambda \cdot (\hat\beta_1 \ \hat\beta_2)\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix} =$$

$$= y^Ty - (\hat\beta_1+\hat\beta_2)\underline{xy^T} - (\hat\beta_1+\hat\beta_2)\underline{x^Ty} - (\hat\beta_1+\hat\beta_2)^2 \underline{\underline{x^Tx}} + \lambda(\hat\beta_1^2 + \hat\beta_2^2)$$

x - standardized $\Rightarrow x^Tx = 1$

$$x^Ty = \begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}^T \begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix} = (x_1 \dots x_n)\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix} = \sum x_i y_i$$

$$xy^T = y^Tx = \begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}^T \begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix} = (y_1 \dots y_n)\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix} = \sum y_i x_i = \sum x_i y_i$$

$\Rightarrow$

$\Rightarrow x^Ty = xy^T$

$$loss(\hat\beta) = y^Ty - 2(\hat\beta_1+\hat\beta_2)y^Tx + (\hat\beta_1+\hat\beta_2)^2 + \lambda(\hat\beta_1^2+\hat\beta_2^2)$$

F.O.C.-s :

$$\frac{\partial\, loss(\hat\beta)}{\partial \hat\beta_1} = -2y^Tx + 2(\hat\beta_1+\hat\beta_2) + 2\lambda\hat\beta_1 = 0$$

$$\lambda\hat\beta_1 + \hat\beta_1 + \hat\beta_2 - y^Tx = 0$$

$$\boxed{\hat\beta_1(\lambda+1) + \hat\beta_2 = y^Tx} \quad (1)$$

$$\frac{\partial\, loss(\hat\beta)}{\partial \hat\beta_2} = -2y^T x + 2(\hat\beta_1 + \hat\beta_2) + 2\lambda\beta_2 = 0$$

$$\lambda\beta_2 + \hat\beta_1 + \hat\beta_2 - y^T x = 0$$

$$\boxed{\hat\beta_2(\lambda + 1) + \hat\beta_1 = y^T x} \quad (2)$$

$$RHS\,(1) = RHS\,(2) = y^T x \Rightarrow$$

$$\Rightarrow \hat\beta_1(\lambda+1) + \hat\beta_2 = \hat\beta_2(\lambda+1) + \hat\beta_1$$

$$\hat\beta_1\lambda + \cancel{\hat\beta_1} + \cancel{\hat\beta_2} = \hat\beta_2\lambda + \cancel{\hat\beta_2} + \cancel{\hat\beta_1}$$

$$\boxed{\hat\beta_1 = \hat\beta_2} \quad (3)$$

---

(3) ∪ (2):

$$\hat\beta_2(\lambda+1) + \hat\beta_2 = y^T x$$

$$\hat\beta_2(\lambda+1+1) = y^T x$$

$$\hat\beta_2(\lambda+2) = y^T x$$

$$\boxed{\hat\beta_2 = \frac{y^T x}{\lambda+2} = \hat\beta_1}$$ – optimal $\hat\beta_1 = \hat\beta_2$ for fixed λ

(b)

$$\hat\beta_1 = \hat\beta_2 = \frac{y^T x}{\lambda+2}$$

$$\lim_{\lambda\to\infty} \hat\beta_1 = \lim_{\lambda\to\infty} \hat\beta_2 = \frac{y^T x}{\lambda+2} = 0$$

So, $\hat\beta_1, \hat\beta_2 \to 0$ as $\lambda \to \infty$

(c)

$$\hat\beta_1 + \hat\beta_2 = \frac{y^T x}{\lambda+2} + \frac{y^T x}{\lambda+2} = \frac{2y^T x}{\lambda+2}$$

$$\lim_{\lambda\to 0}(\hat\beta_1 + \hat\beta_2) = \lim_{\lambda\to 0}\frac{2y^T x}{\lambda+2} = \frac{2y^T x}{0+2} =$$

$$= \frac{\cancel{2}y^T x}{\cancel{2}} = y^T x$$

So, $\hat\beta_1 + \hat\beta_2 \to y^T x$ as $\lambda \to 0$