Exercise. 3.29

Suppose we fit a ridge regression with a given shrinkage parameter $\lambda \in \mathbb{R}^+$ on a single variable x_1. (Notice that x_1 is a $N \times 1$ vector.)

1. (Essential) Show that the coefficient must be $\frac{X^T y}{X^T X + \lambda}$ where $X = x_1$.
2. (Essential) We now include an exact copy $x_2 = x_1$, so our new design matrix would be $X = [x_1|x_2]$. Using this matrix, re-fit our ridge regression. Show that both coefficients are identical, and derive their value.
3. (Extra) Show in general that if m copies of a variable x_j, are included in a ridge regression, so X would be $[x_1|x_2|\cdots|x_m]$, their coefficients are all the same.

1\. $$\hat{\beta}^R = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \cdot \beta_j \Big)^2 + \lambda \cdot \sum_{j=1}^{p} \beta_j^2 \right\}^2$$

$$= \underset{\beta}{\operatorname{argmin}} \left\{ \|y - X\beta\|^2 + \lambda \cdot \|\beta\|^2 \right\}$$

$\Rightarrow$ β에 대해 미분하여 최소값 찾기

$$-2X'(y - X\hat{\beta}) + 2\lambda \cdot \hat{\beta} = 0$$

$$X'y = X'X\hat{\beta} + \lambda \cdot \underline{I} \cdot \hat{\beta} = (X'X + \lambda I) \cdot \hat{\beta}$$

$p \times p$ Identity matrix

$$\therefore \hat{\beta} = (X'X + \lambda I)^{-1} \cdot X' \cdot y$$ $\Leftarrow$ $p=1$일때 $(X'X + \lambda I)$는 scala

$$= \frac{X' \cdot y}{X'X + \lambda}$$

2\. when $p = 2$

$$\hat{\beta}^R = \underset{\beta_1, \beta_2}{\operatorname{argmin}} \left\{ \|X\beta_1 + X\beta_2 - y\|^2 + \lambda \cdot \|\beta_1\|^2 + \lambda \|\beta_2\|^2 \right\}$$

β에 대해 각각 편미분

$$\begin{cases} 2X'(X\beta_1 + X\beta_2 - y) + 2\lambda \cdot \beta_1 = 0 \\ 2X'(X\beta_1 + X\beta_2 - y) + 2\lambda \cdot \beta_2 = 0 \end{cases}$$

$$\Rightarrow \beta_1 = \beta_2$$

위의 결과를 위의 연립방정식에 대입해보면...

$$2X'(X\beta_1 + X\beta_1 - y) + 2\lambda \cdot \beta_1 = 0$$

$$2X'X\beta_1 - 2X'y + 2\lambda \cdot \beta_1 = 0$$

$$\beta_1(2X'X+\lambda) = X'X$$

$$\beta_1 = \beta_2 = \frac{X'X}{2X'X+\lambda} \qquad \therefore \beta_1 = \beta_2$$

3. when $p = m$

$$\hat{\beta}^R = \underset{\beta_1, \beta_2, \cdots, \beta_m}{\arg\min} \left\{ \|X\beta_1 + X\beta_2 + \cdots + X\beta_m - y\|^2 + \lambda\|\beta_1\|^2 + \cdots + \lambda\|\beta_m^2\| \right\}$$

즉 β에 대해 위의 식을 편미분하면,

$$2X'(X\beta_1 + X\beta_2 + \cdots + X\beta_m - y) + 2\lambda\beta_1 = 0$$

$$\vdots \qquad\qquad \vdots$$

$$2X'(X\beta_1 + X\beta_2 + \cdots + X\beta_m - y) + 2\lambda\beta_m = 0$$

$$\Rightarrow \ -2\lambda\beta_1 = -2\lambda\beta_2 = \cdots = -2\cdot\lambda\cdot\beta_m$$

위의 결과에서도 $\beta_1 = \beta_2 = \cdots = \beta_m$

이를 위의 연립방정식에 대입해보면

eq $\beta_1(mX'X+\lambda) = X'y$

$$\therefore \beta_1 = \beta_2 = \cdots = \beta_m = \frac{X'y}{m\cdot X'X+\lambda}$$