II. Ridge Regularization

$$l(\beta) = (\vec{y} - \vec{x}\beta)^T(\vec{y} - \vec{x}\beta) + \alpha \beta^T \hat{I} \beta \quad , \hat{I} = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$$

1. Derive expression of $\frac{\partial l}{\partial \beta}$.

$$= \frac{\partial}{\partial \beta}\left[y^T y - \beta^T x^T y - y^T x \beta + \beta^T x^T x \beta + \alpha \beta^T \hat{I} \beta \right]$$

Each term gives (solving separately)

$$\frac{\partial\, y^T y}{\partial \beta} = 0 \quad ; \frac{\partial}{\partial \beta}[\beta^T x^T y] = [x^T y]^T = y^T x$$

$$\frac{\partial}{\partial \beta}[y^T x \beta] = y^T x$$

$$\frac{\partial}{\partial \beta}[\beta^T x^T x \beta] = \beta^T (x^T x + x^T x) = 2\beta^T (x^T x)$$

$$\frac{\partial}{\partial \beta}[\alpha \beta^T \hat{I} \beta] = \alpha \beta^T (\hat{I} + \hat{I}) = 2\alpha \beta^T \hat{I}$$

$$\therefore \frac{\partial}{\partial \beta}[l(\beta)] = 2x^T (x\beta^T - y) + 2\alpha \beta^T \hat{I}$$

$$= 2\beta^T (x^T x) - 2x^T x + 2\alpha \beta^T \hat{I}$$

2. Derive expression for $\text{argmin}_\beta\, l(\beta)$

minimizing $l(\beta)$, $\frac{\partial}{\partial \beta} l(\beta) = 0$

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

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

$$\Rightarrow \beta^T (x^T x) + \alpha \beta^T I = y^T x$$

Taking transpose,

$$\{\beta^T [[x^T x] + \alpha \hat{I}]\}^T = (y^T x)^T$$

$$\therefore \boxed{\begin{array}{l}\beta = (x^T x + \alpha I)^{-1} x^T y \\ \text{arg min of } l(\beta)\end{array}} \qquad [I^T = I]$$