

We have:

$$D_{\text{ridge}}(\beta_0, \beta) = \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2$$

We get:

$$\begin{aligned} \frac{\partial D_{\text{ridge}}(\beta_0, \beta)}{\partial \beta_0} &= 2 \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij}) \cdot (-1) \\ &= -2 \sum_{i=1}^N (y_i - \beta_0) + 2 \sum_{i=1}^N \sum_{j=1}^p \beta_j x_{ij} \\ &= -2 \sum_{i=1}^N (y_i - \beta_0) + 2 \sum_{j=1}^p \beta_j \underbrace{\sum_{i=1}^N x_{ij}}_{=0 \text{ because of standardization}} \\ &= -2 \sum_{i=1}^N (y_i - \beta_0) = 0 \\ &\rightarrow \sum_{i=1}^N y_i - N\beta_0 = 0 \rightarrow \beta_0 = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y} \end{aligned}$$

For $l = 1, \dots, p$:

$$\begin{aligned} \frac{\partial D_{\text{ridge}}(\beta_0, \beta)}{\partial \beta_l} &= 2 \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij}) \cdot (-x_{il}) + 2\lambda \beta_l \\ &= -2 \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij}) x_{il} + 2\lambda \beta_l = 0 \\ &\rightarrow \lambda \beta_l + \sum_{i=1}^N \sum_{j=1}^p \beta_j x_{ij} x_{il} = \sum_{i=1}^N (y_i - \beta_0) x_{il} \\ &= \sum_{i=1}^N (y_i - \bar{y}) x_{il} \\ &= \sum_{i=1}^N y_i x_{il} - \bar{y} \underbrace{\sum_{i=1}^N x_{il}}_{=0} \\ &= \sum_{i=1}^N y_i x_{il} \end{aligned}$$

On matrix form:

$$\begin{aligned} D_{\text{ridge}}(\beta_0, \beta) &= (y - \beta_0 \mathbf{1} - X\beta)^T (y - \beta_0 \mathbf{1} - X\beta) + \lambda \beta^T \beta \\ &= (y - X\beta)^T (y - X\beta) - \beta_0 \mathbf{1}^T (y - X\beta) \\ &\quad - (y - X\beta)^T \beta_0 \mathbf{1} + \beta_0 \mathbf{1}^T \mathbf{1} \beta_0 + \lambda \beta^T \beta \\ &= y^T y - 2y^T X\beta + \beta^T X^T X \beta + \lambda \beta^T \beta \\ &\quad - 2y^T \beta_0 \mathbf{1} + 2\beta_0 \mathbf{1}^T X\beta + N\beta_0^2 \\ &\quad \underbrace{\beta_0 y^T \mathbf{1}}_{= \beta_0 \sum_{i=1}^N y_i} \quad \underbrace{N\beta_0^2}_{\bar{y}^2} \\ &= \beta_0 \sum_{i=1}^N y_i \\ &= \underbrace{\beta_0 \cdot N \cdot \bar{y}}_{= \bar{y}^2} \\ &= N \cdot \bar{y}^2 \end{aligned}$$

We have:

$$\mathbf{1}^T X = (1 \dots 1) \begin{pmatrix} x_{11} & \dots & x_{p1} \\ \vdots & & \vdots \\ x_{1N} & \dots & x_{pN} \end{pmatrix} = \begin{pmatrix} \underbrace{\sum_{i=1}^N x_{1i}}_{=0} & \dots & \underbrace{\sum_{i=1}^N x_{pi}}_{=0} \end{pmatrix} = (0 \dots 0)$$

$$\begin{aligned} \rightarrow D_{\text{ridge}}(\beta_0, \beta) &= y^T y - 2y^T X\beta + \beta^T X^T X \beta + \lambda \beta^T \beta - N\bar{y}^2 \\ &= y^T y - 2y^T X\beta + \beta^T (X^T X + \lambda \mathbf{I}) \beta - N\bar{y}^2 \end{aligned}$$