

We have:

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

We get:

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

We get:

$$-2 \sum_{i=1}^N (y_i - \hat{\beta}_0) = 0 \rightarrow \sum_{i=1}^N y_i = \underbrace{\sum_{i=1}^N \hat{\beta}_0}_{N \cdot \hat{\beta}_0} \rightarrow \hat{\beta}_0 = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y}$$

Further, for $l = 1, \dots, p$:

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

We get:

$$\begin{aligned} &-2 \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \sum_{j=1}^p \hat{\beta}_j x_{ij} \right) \cdot x_{il} + 2\lambda \hat{\beta}_l = 0 \\ \rightarrow &-2 \sum_{i=1}^N (y_i - \hat{\beta}_0) x_{il} + 2 \sum_{i=1}^N \sum_{j=1}^p \hat{\beta}_j x_{ij} x_{il} + 2\lambda \hat{\beta}_l = 0 \\ \rightarrow &\sum_{i=1}^N \sum_{j=1}^p \hat{\beta}_j x_{ij} x_{il} + \lambda \hat{\beta}_l = \sum_{i=1}^N (y_i - \hat{\beta}_0) x_{il} = \sum_{i=1}^N (y_i - \bar{y}) x_{il} = \sum_{i=1}^N y_i x_{il} - \bar{y} \underbrace{\sum_{i=1}^N x_{il}}_{=0} \\ &= \sum_{i=1}^N y_i x_{il} \\ \rightarrow &\sum_{j=1}^p \hat{\beta}_j \sum_{i=1}^N x_{ij} x_{il} + \lambda \hat{\beta}_l = \sum_{i=1}^N y_i x_{il}, \quad l = 1, \dots, p \end{aligned}$$

On vector/matrix form, we have: $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^N$

$$\begin{aligned}
 \text{V}_{\text{ridge}}(\beta_0, \beta) &= (\mathbf{y} - \beta_0 \mathbf{1} - \mathbf{X} \beta)^T (\mathbf{y} - \beta_0 \mathbf{1} - \mathbf{X} \beta) + \lambda \beta^T \beta \\
 &= (\mathbf{y} - \mathbf{X} \beta)^T (\mathbf{y} - \mathbf{X} \beta) + \lambda \beta^T \beta - \beta_0 \mathbf{1}^T (\mathbf{y} - \mathbf{X} \beta) \\
 &\quad - \underbrace{\beta_0 (\mathbf{y} - \mathbf{X} \beta)^T \mathbf{1}}_{\mathbf{1}^T (\mathbf{y} - \mathbf{X} \beta)} + \underbrace{\beta_0^2 \mathbf{1}^T \mathbf{1}}_{= N} \\
 &= \mathbf{y}^T \mathbf{y} - 2 \mathbf{y}^T \mathbf{X} \beta + \beta^T \mathbf{X}^T \mathbf{X} \beta + \lambda \beta^T \beta - 2 \beta_0 (\underbrace{\mathbf{1}^T \mathbf{y}}_{= \mathbf{1}^T \mathbf{y}} - \underbrace{\mathbf{1}^T \mathbf{X} \beta}_{= 0}) \\
 &\quad + N \beta_0^2 \\
 &= \mathbf{y}^T \mathbf{y} - 2 \mathbf{y}^T \mathbf{X} \beta + \beta^T (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}) \beta - 2 \beta_0 \underbrace{\mathbf{1}^T \mathbf{y}}_{= N \bar{y}} + N \beta_0^2 \\
 &= \mathbf{y}^T \mathbf{y} - 2 \mathbf{y}^T \mathbf{X} \beta + \beta^T (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}) \beta - N \beta_0 \bar{y} + N \beta_0^2
 \end{aligned}$$

We get:

$$\frac{\partial \text{V}_{\text{ridge}}(\beta_0, \beta)}{\partial \beta} = -2 \mathbf{X}^T \mathbf{y} + 2(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}) \beta$$

We want to solve:

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

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

$$\longrightarrow \hat{\beta} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$