

We have:

$$D_{\text{ridge}}(\beta_0, \beta) = y^T y - 2 y^T X \beta + \beta^T (X^T X + \lambda I) \beta - N \beta_0^2$$

We get

$$\frac{\partial D_{\text{ridge}}(\beta_0, \beta)}{\partial \beta} = -2 X^T y + 2 (X^T X + \lambda I) \beta = 0$$

$$\rightarrow (X^T X + \lambda I) \beta = X^T y$$

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

$$\hat{\beta}_0 = \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} \mathbf{1}^T \mathbf{y}, \quad \hat{\beta} = (X^T X + \lambda I)^{-1} X^T \mathbf{y}$$

We have:

$$\begin{aligned} \hat{\mathbf{y}} &= \hat{\beta}_0 \mathbf{1} + X \hat{\beta} \\ &= \frac{1}{N} \mathbf{1} \mathbf{1}^T \mathbf{y} + X (X^T X + \lambda I)^{-1} X^T \mathbf{y} \\ &= \underbrace{\left(\frac{1}{N} \mathbf{1} \mathbf{1}^T + X (X^T X + \lambda I)^{-1} X^T \right)}_{H_\lambda} \mathbf{y} = H_\lambda \mathbf{y} \end{aligned}$$

We get

$$\begin{aligned} df(\lambda) &= \text{tr}(H_\lambda) = \text{tr} \left(\frac{1}{N} \mathbf{1} \mathbf{1}^T + X (X^T X + \lambda I)^{-1} X^T \right) \\ &= \frac{1}{N} \text{tr}(\mathbf{1} \mathbf{1}^T) + \text{tr} \left(X (X^T X + \lambda I)^{-1} X^T \right) \\ &= \frac{1}{N} \underbrace{\text{tr}(\mathbf{1}^T \mathbf{1})}_{= N} + \text{tr} \left((X^T X + \lambda I)^{-1} X^T X \right) \\ &\quad \underbrace{\hspace{10em}}_{= 1} \end{aligned}$$