Exercise [redacted] B

Part I is identical to what we did in the slides. Let

$$f(x_i;\alpha,\beta) = \beta_0 + x_i^T\beta = \alpha + (x_i - \bar{x})^T\beta$$

$$\hat{\alpha} = \underset{\substack{\alpha\in\mathbb{R}\\ \beta\in\mathbb{R}^p}}{\operatorname{arg\,min}} \sum_{i=1}^{m}\left(y_i - \alpha - (x_i-\bar{x})^T\beta\right)^2 + \lambda\sum_{j=1}^{p}\beta_0^2.$$

$$\Longrightarrow \quad \hat{\alpha} = \bar{y}, \text{ irrespectively of } \beta. \qquad ; \quad \beta_0 = \alpha - x_i^T\beta.$$

Then,

$$\hat{\beta}_{Bridge} = \underset{\beta}{\operatorname{arg\,min}} \sum_{i=1}\left\{(y_i-\bar{y}) - (x_i-\bar{x})^T\beta\right\}^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

Part II

$$(\hat{\beta}_0, \hat{\beta}_{scaled\text{-}ridge}) = \underset{\beta_0,\beta}{\operatorname{arg\,min}} \sum_{i=1}^{m}\left(y_i - \beta_0 - x_i^T\beta\right)^2 + \lambda\sum_{j=1}^{p} s_j^2\beta_j^2$$

Following the same steps of part I, we get:

$$\hat{\beta}_{scaled\text{-}ridge} = \underset{\beta}{\operatorname{arg\,min}} \sum_{i=1}^{m}\left\{(y_i-\bar{y}) - (x_i-\bar{x})^T\beta\right\}^2 + \lambda\sum_{j=1}^{p} s_j^2\beta_j^2$$

$$\Longrightarrow \quad \sum_{i=1}^{m}\left\{(y_i-\bar{y}) - (x_i-\bar{x})^T\beta\right\}^2 + \lambda\sum_{j=1}^{p} s_j^2\beta_j^2 =$$

$$= \sum_{i=1}^{m}\left\{(y_i-\bar{y}) - \sum_{j=1}^{p}(x_{ij}-\bar{x}_j)\frac{s_j}{s_j}\beta_j\right\}^2 + \lambda\sum_{j=1}^{p}(s_j\beta_j)^2$$

$$\tilde{\beta}_j = s_j\beta_j \longrightarrow \beta_j = \frac{\tilde{\beta}_j}{s_j}$$

$$= \sum_{i=1}^{n} \left\{ (y_i - \bar{y}) - \sum_{j=1}^{p} \left(\frac{x_{ij} - \bar{x}_j}{s_j} \right) \tilde{\beta}_j \right\}^2 + \lambda \sum_{j=1}^{p} \tilde{\beta}_j^2$$

$$\hat{\beta}_{ridge} = \underset{\tilde{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ (y_i - \bar{y}) - \sum_{j=1}^{p} \left(\frac{x_{ij} - \bar{x}_j}{s_j} \right) \tilde{\beta}_j \right\}^2 + \lambda \sum_{j=1}^{p} \tilde{\beta}_j$$

and $\hat{\beta}_{scaled\text{-}ridge} = \operatorname{diag}(1/s_1, \ldots, 1/s_p) \hat{\beta}_{ridge}$.

$\hat{\beta}_0 = \bar{y} - \bar{x}^T \hat{\beta}_{scaled\text{-}ridge}$ (as a consequence of part I).