CS 534. HW 1.

1. We augment the centered matrix $\vec{X}$ with k additional rows $\sqrt{\lambda}\,\vec{I}$ and augment $\vec{y}$ with k zeros, which means training data $(x_1, y_1), \cdots (x_N, y_N)$ become $(x_1, y_1) \cdots (x_N, y_N) (x_{N+1}, 0) (x_{N+2}, 0) \cdots (x_{N+k}, 0)$, so

$$RSS(\beta) = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \cdot \beta_j\right)^2 + \sum_{i=N+1}^{N+k} \left(0 - \sum_{j=1}^{p} x_{ij} \cdot \beta_j\right)^2. \quad ①$$

we do not need to add intercept β_0 to $y_{N+1} = y_{N+2} = y_{N+3} \cdots = y_{N+k} = 0$. because we use centered inputs: each x_{ij} gets replaced by $x_{ij} - \bar{x}_j$. We estimate β_0 by $\bar{y} = \frac{1}{N}\sum_{1}^{N} y_i$, $\bar{y}' = \frac{1}{k}\sum_{N+1}^{N+k} y_i = 0$

$\because$ k additional rows is $\sqrt{\lambda}\,\vec{I}$, $\vec{I}$ is $k \times k$ identity matrix,

$\therefore$ $k = p$. and $\sum_{i=N+1}^{N+k} \left(\sum_{j=1}^{p} x_{ij} \cdot \beta_j\right)^2 = \sum_{i=N+1}^{N+p} \left(\sum_{j=1}^{p} x_{ij} \cdot \beta_j\right)^2$, where $x_{ij} = \begin{cases} \sqrt{\lambda}, & i = j+N \\ 0, & i \neq j+N \end{cases}$

$$= \sum_{j=1}^{p} (\sqrt{\lambda} \cdot \beta_j)^2$$

$$= \lambda \sum_{j=1}^{p} \beta_j^2$$

$\therefore$ $RSS(\beta) = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \cdot \beta_j\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$, $(\lambda \geq 0.)$

which is almost equal to $\hat{\beta}^{ridge}$ of original $\vec{X}$ and $\vec{y}$

$$\Rightarrow \hat{\beta}^{ridge} = \underset{\beta}{\arg\min} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \cdot \beta_j\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}. \ (\lambda \geq 0)$$