Given the following regularized regression problem:

$$\hat{\beta}^{Ridge} = \arg\min_{\beta\mathbb{R}} \|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots.(1)$$

Show that the solution is

$$\hat{\beta}^{Ridge} = (X^T X + \lambda I_P)^{-1} X^T y \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(2)$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix}$$

and

$$\beta = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

Then $X\beta \in \mathbb{R}^N$ and

$$X\beta = \begin{bmatrix} \sum_{j=1}^{p} b_j x_{1j} \\ \sum_{j=1}^{p} b_j x_{2j} \\ \vdots \\ \sum_{j=1}^{p} b_j x_{Nj} \end{bmatrix} \rightarrow y - X\beta = \begin{bmatrix} y_1 - \sum_{j=1}^{p} b_j x_{1j} \\ y_2 - \sum_{j=1}^{p} b_j x_{2j} \\ \vdots \\ y_N - \sum_{j=1}^{p} b_j x_{Nj} \end{bmatrix}$$

Therefore,

$$(y - X\beta)^T (y - X\beta) = \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} b_j x_{ij} \right)^2 = \|y - X\beta\|^2$$

Now we can write (1) as

$$\hat{\beta}^{Ridge} = \arg\min_{\beta\mathbb{R}} \|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2$$

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

$$= y^T y - 2\beta^T X^T y + \beta^T X^T X\beta + \lambda\beta^T\beta$$

Differentiation $\hat{\beta}^{Ridge}$ with respect to $d\beta$

$$\frac{\partial}{\partial\beta}\hat{\beta}^{Ridge} = -2X^{T}y + 2X^{T}X\beta + 2\lambda\beta \dots\dots\dots\dots\dots\dots..(3)$$

To minimize this, set the result of (3) to 0.

$$-2X^{T}y + 2X^{T}X\beta + 2\lambda\beta = 0$$

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

$$(X^{T}y - X^{T}X\beta - \lambda\beta) = 0$$

$$X^{T}y = X^{T}X\beta + \lambda\beta$$

$$X^{T}y = (X^{T}X + \lambda I_{P})\beta$$

$$X^{T}y = (X^{T}X + \lambda I_{P})\beta$$

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

So, the closed-form ridge estimator through regularization least squares:

$$\hat{\beta}^{Ridge} = (X^{T}X + \lambda I_{P})^{-1}X^{T}y$$