1a.

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}_{N\times 1} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}_{(N+1)\times 1} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix}_{N\times(N+1)}$$

$$(y - X\beta) = \underset{N\times 1}{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}} - \underbrace{\underset{N\times(N+1)}{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N1} & \cdots & x_{Nn} \end{bmatrix}} \underset{(N+1)\times 1}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}}}_{N\times 1}$$

$$= \begin{pmatrix} y_1 - (\beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + \dots x_{1n}\beta_n) \\ y_2 - (\beta_0 + x_{21}\beta_1 + x_{22}\beta_2 + \dots x_{2n}\beta_n) \\ \vdots \\ y_N - (\beta_0 + x_{N1}\beta_N + x_{N2}\beta_2 + \dots x_{Nn}\beta_n) \end{pmatrix}$$

$$= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} \quad \text{where } a_i = y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij} \quad — \bigstar$$

$$(y - X\beta)^T (y - X\beta) = (a_1, a_2, \dots a_N) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}$$

$$= a_1^2 + a_2^2 + \dots a_N^2$$

$$= \sum_{i=1}^{N} a_i^2$$

$$\overset{\bigstar}{=} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij} \right)^2 = RSS(\beta) \quad //$$

1b. $RSS(\beta) = (y - X\beta)^T (y - X\beta)$

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

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

$$\frac{d\,RSS(\beta)}{d\beta} = \frac{d(y^T y - 2\beta^T X^T y + \beta^T X^T X \beta)}{d\beta}$$

$$= -\frac{d(2\beta^T X^T y)}{d\beta} + \frac{d(\beta^T X^T X \beta)}{d\beta}$$

$$= -2X^T y + 2X^T X \beta \quad //$$

($\because X^T X$ is symmetric, we can use given result $\frac{d\,X^T A X}{dX} = 2AX$)

1c. Set $\frac{d\,RSS(\beta)}{d\beta} = 0$ to find minimum

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

$$X^T y = X^T X \beta$$

$$(X^T X)^{-1} X^T y = (X^T X)^{-1} (X^T X) \beta$$

$$= I\beta$$

$$\therefore \hat{\beta} = (X^T X)^{-1} X^T y \quad //$$

2a)

$$A = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & & & 0 \\ \vdots & \vdots & 1 & & \vdots \\ \vdots & \vdots & & 1 & \vdots \\ \vdots & \vdots & & & \ddots \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

$(n+1) \times (n+1)$ $\qquad$ $(n+1) \times 1$

$$\underset{1 \times (n+1)}{\beta^T A = (\beta_0, \beta_1 \ldots \beta_n)} \underset{(n+1)\times(n+1)}{\begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & & & 0 \\ \vdots & \vdots & 1 & & \vdots \\ \vdots & \vdots & & 1 & \vdots \\ \vdots & \vdots & & & \ddots \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix}} = \underset{1 \times (n+1)}{(0, \beta_1, \beta_2, \ldots \beta_n)}$$

$$\beta^T A \beta = \underset{1 \times (n+1)}{(0, \beta_1, \beta_2, \ldots \beta_n)} \underset{(n+1) \times 1}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}}$$

$$= (\beta_1^2 + \beta_2^2 + \ldots \beta_n^2) = \sum_{i=1}^{n} \beta_i^2 \quad //$$

2b)

$$J(\beta) = \underbrace{\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij}\right)^2}_{\downarrow \text{ shown in Q1}} + \alpha \underbrace{\sum_{i=1}^{n} \beta_i^2}_{\downarrow \text{ shown in Q2a}}$$

$$= \underbrace{(y - X\beta)^T (y - X\beta)}_{\downarrow \text{ shown in Q1}} + \alpha(\beta^T A \beta)$$

$$\frac{dJ(\beta)}{d\beta} = -2X^T y + 2X^T X \beta + \underbrace{\frac{d\alpha(\beta^T A \beta)}{d\beta}}_{\downarrow \; = \alpha 2 A\beta}$$

($\because A$ is symmetric identity matrix, we can use result $\frac{d x^T A x}{dx} = 2Ax$)

Set $\frac{dJ(\beta)}{d\beta} = 0$ to find minimum

$$-2X^T y + 2X^T X\beta + \alpha 2 A \beta = 0 \Rightarrow X^T X \beta + \alpha A \beta = X^T y$$

$$(X^T X + \alpha A)\beta = X^T y \qquad \therefore \beta = (X^T X + \alpha A)^{-1} X^T y \quad //$$

3.

① $$\underset{\beta}{\operatorname{argmin}}\left\{\sum_{i=1}^{N}\big(\overbrace{y_i-\beta_0-\sum_{j=1}^{n}\beta_j x_{ij}}^{a}\big)^2+\alpha\sum_{i=1}^{n}\beta_i^2\right\}$$

② $$\underset{\beta_c}{\operatorname{argmin}}\left\{\sum_{i=1}^{N}\big(\overbrace{y_i-\beta_0^c-\underbrace{\sum_{j=1}^{n}(x_{ij}-\bar{x}_j)\beta_j^c}_{\text{Expand.}}}^{b}\big)^2+\alpha\sum_{j=1}^{n}(\beta_j^c)^2\right\}$$

$$\sum_{j=1}^{n}(x_{ij}-\bar{x}_j)\beta_j^c=\sum_{j=1}^{n}x_{ij}\beta_j^c-\sum_{j=1}^{n}\bar{x}_j\beta_j^c$$

Compare ① a with ② b expanded:

1ⓐ $y_i-\beta_0-\sum_{j=1}^{n}\beta_j x_{ij}$

2ⓑ $y_i-\beta_0^c-\left(\sum_{j=1}^{n}\beta_j^c x_{ij}-\sum_{j=1}^{n}\bar{x}_j\beta_j^c\right)$

We can see that the two equations ① and ② are equivant, with

$$\beta_0=\beta_0^c+\sum_{j=1}^{n}\bar{x}_j\beta_j^c \quad //$$

$$\beta_j=\beta_j^c \quad //$$