1. Summation form:

$$SSE(\beta) = \sum_{i=1}^{n} (x_i^T \beta - y_i)^2$$

$$MSE(\beta) = \frac{1}{n} \sum_{i=1}^{n} (x_i^T \beta - y_i)^2$$

$$L(\beta) = \frac{1}{2n} \sum_{i=1}^{n} (x_i^T \beta - y_i)^2$$

Matrix form:

$$SSE(\beta) = \|X\beta - y\|^2$$

$$MSE(\beta) = \frac{1}{n} \|X\beta - y\|^2$$

$$L(\beta) = \frac{1}{2n} \|X\beta - y\|^2$$

2. $C > 0$

$$L'(\beta) = C \cdot L(\beta)$$

$$\nabla_\beta L'(\beta) = C \cdot \nabla_\beta L(\beta)$$

when $\nabla_\beta L(\beta) = 0$, $\nabla_\beta L'(\beta) = 0$.

Multiplying by a constant doesn't change the location of zeros, $\Rightarrow$ the minimizer remains the same.

3. It tells us how much the loss would increase or decrease if we adjust the parameter.

4. $\frac{1}{n} X^T (X\beta - y) = 0 \Rightarrow X^T X\beta - X^T y = 0$

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