A.1

$$SSE(\beta) = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \|X\beta - y\|^2$$

$$MSE(\beta) = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \frac{1}{n} \|X\beta - y\|^2$$

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

$$SSE(\beta) = (y - X\beta)^T (y - X\beta)$$

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

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

A.2.

Scaling the loss by a positive constant only changes its magnitude, not its shape. the minimizer stays.

A.3

$$\frac{\partial L}{\partial \beta} = \frac{1}{n} X^T (X\beta - y) \qquad \beta \leftarrow \beta - \alpha \frac{1}{n} X^T (X\beta - y)$$

A.4.

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

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