

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

~~2RSS~~

$$= y^T y - \overbrace{y^T X \beta}^{\text{scalar}} - \overbrace{\beta^T X^T y}^{\text{scalar}} + \overbrace{\beta^T X^T X \beta}^{\text{scalar}}$$
~~$$= y^T y - 2X^T y \beta + X^T X \beta^T \beta$$~~

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

$$\frac{\partial RSS(\beta)}{\partial \beta} = -2X^T y + 2X^T X \hat{\beta} = 0$$

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