Estimating β in Multivariable Linear Regression

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \; RSS(\beta)$$
$$= \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2$$
$$= \underset{\beta}{\operatorname{argmin}} \; (y - X\beta)^T (y - X\beta) \quad \text{in matrix form}$$
$$= \underset{\beta}{\operatorname{argmin}} \; y^T y - 2\beta^T X^T y + \beta^T X^T X \beta$$

<u>Multivariable Calculus</u>: set the gradient equal to 0. The gradient is the analog of the derivative.

Gradient $\nabla_\beta f(\beta) = \nabla_\beta f(\beta_0, \dots, \beta_K) = \left(\frac{\partial f}{\partial \beta_0}, \dots, \frac{\partial f}{\partial \beta_K} \right)$ is the vector of partial derivatives.

$$0 = \nabla_\beta RSS(\beta) = \nabla_\beta \left(y^T y - 2\beta^T X^T y + \beta^T X^T X \beta \right)$$
$$= -2X^T y + 2(X^T X)\beta \implies X^T X \beta = X^T y$$

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

This is the matrix form of multivariable linear regression!