$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + \varepsilon_i$$

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_i \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ik} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \cdots \\ b_i \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \cdots \\ \varepsilon_i \end{bmatrix}$$

$$y = Xb + \varepsilon$$

$$\varepsilon = y - Xb$$

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} ((\cancel{b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik}} + \varepsilon) - (\cancel{b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik}}))^2$$

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = [y - Xb]^T [y - Xb]$$

$$\varepsilon^T \varepsilon = [y - Xb]^T [y - Xb] = [y^T - b^T X^T][y - Xb] = y^T y - y^T X b - b^T X^T y + b^T X^T X b$$

$$\varepsilon^T \varepsilon = y^T y - y^T X b - (Xb)^T y + b^T X^T X b = y^T y - 2 y^T X b + b^T X^T X b$$

$$\frac{\partial}{\partial_b} = 0 - 2 y^T X + 2 X^T X b = 0 \leftrightarrow -2 y^T X + 2 X^T X b = 0$$

$$X^T X b = y^T X$$

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

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