Derivation of Normal Eq^n

Hypothesis func. of Linear Regression :

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

where θ_i is coeff., n is no. of features

Cost func.

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

where m is no. of samples.

Writing in matrix notation :

$$h_\theta(x) = \theta^T x \text{ , where } \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$\therefore \quad J(\theta) = \frac{1}{2m} (X\theta - y)^T (X\theta - y)$$

$$= \frac{1}{2m} \left(\theta^T X^T X \theta - 2(X\theta)^T y + y^T y \right)$$

$$\frac{\partial J}{\partial \theta} = \frac{1}{2m} \left(2X^T X\theta - 2X^T y \right) = 0$$

$$\Rightarrow X^T X \theta = X^T y$$

Assuming $X^T X$ is invertible,

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

Normal Eq^n