m examples $(x^{(1)}, y^{(1)}), \cdots, (x^{(m)}, y^{(m)})$

, n features

$$x^{(i)} = \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$X = \begin{bmatrix} \text{---} (x^{(1)})^T \text{---} \\ \text{---} (x^{(2)})^T \text{---} \\ \vdots \\ \text{---} (x^{(m)})^T \text{---} \end{bmatrix}_{m \times (n+1)}$$

(design matrix)

Eg) If $x^{(i)} = \begin{bmatrix} 1 \\ x_1^{(i)} \end{bmatrix}$ here, $n = 1$

$$X = \begin{bmatrix} 1 & x_1^{(1)} \\ 1 & x_1^{(2)} \\ \vdots & \vdots \\ 1 & x_1^{(m)} \end{bmatrix}_{m \times 2} \qquad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}_{m}$$

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

Set $A = X^T X$, $A^{-1} = (X^T X)^{-1}$

Octave: pinv(X' * X) * X' * y

not need for feature scaling

$\theta = (X^T X)^{-1} X^T y \Rightarrow$ s.t. $\min_\theta J(\theta)$