m = number of examples

n = number of features

$x^{(i)}$ = input of ith traning example.

$x^{(i)}_{(j)}$ = value of feature j in ith training example.

let $x_0 = 1$

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 \cdots \theta_n x_n \Rightarrow [\theta_0\ \theta_1\ \theta_2 \cdots \theta_n] \times \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$\theta^T X$ ⇐ product of param θ and feature x

let θ = $n+1$ demensional vector = param

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 = \text{cost function}$$

$$\theta_j := \theta_j - \alpha \frac{d}{d\theta} J(\theta) = \text{Gradient Descent}$$

New algorithm:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^{(i)}) x_j^{(i)}$$