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

1 iteration of gradient descent

$$X = \begin{bmatrix} x_0^1 & x_1^1 \\ x_0^2 & x_1^2 \\ x_0^3 & x_1^3 \\ \vdots & \vdots \\ x_0^m & x_1^m \end{bmatrix}_{m \times 2} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}_{2 \times 1} \quad y = \begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ \vdots \\ y^m \end{bmatrix}_{m \times 1}$$

$$\underbrace{\text{prediction} - y}_{\| \; X * \text{theta}} = \begin{bmatrix} h_\theta(x^{(1)}) - y^{(1)} \\ \vdots \\ h_\theta(x^{(m)}) - y^{(m)} \end{bmatrix}_{m \times 1}$$

$$X^T(\text{prediction} - y) = \begin{bmatrix} x_0^1 & x_0^2 & \cdots & x_0^m \\ x_1^1 & x_1^2 & \cdots & x_1^m \end{bmatrix}_{2 \times m} \times \begin{bmatrix} h_\theta(x^{(1)}) - y^{(1)} \\ h_\theta(x^{(2)}) - y^{(2)} \\ \vdots \\ h_\theta(x^{(m)}) - y^{(m)} \end{bmatrix}_{m \times 1}$$

$$= \begin{bmatrix} h_\theta(x^{(1)}) - y^{(1)} \cdot x_0^{(1)} + h_\theta(x^{(2)}) - y^{(2)} \cdot x_0^{(2)} + \cdots + h_\theta(x^{(m)}) - y^{(m)} \cdot x_0^{(m)} \\ h_\theta(x^{(1)}) - y^{(1)} \cdot x_1^{(1)} + \cdots + h_\theta(x^{(m)}) - y^{(m)} \cdot x_1^{(m)} \end{bmatrix}$$

$$\theta := \theta - \alpha * (X^T(\text{predictions} - y))$$