Derivation of gradient descent.

$$\begin{cases} x^i = x^{(i)} \\ y^i = y^{(i)} \end{cases}$$

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

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_0} \left(\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2\right)$$
$$= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_0} (h_\theta(x^i) - y^i)^2$$
$$= \frac{1}{m} \sum_{i=1}^{m} 2(h_\theta(x^i) - y^i) \frac{\partial}{\partial \theta_0} (h_\theta(x^i) - y^i)$$
$$= \frac{2}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i) \frac{\partial}{\partial \theta_0} (\theta_0 + \theta_1 x^i - y^i)$$
$$= \frac{2}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)$$

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

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x) - y)^2$$
$$= \frac{1}{2} \cdot 2 (h_\theta(x) - y) \cdot \frac{\partial}{\partial \theta_j} (h_\theta(x) - y)$$
$$= (h_\theta(x) - y) \frac{\partial}{\partial \theta_j} \left(\sum_{i=0}^{m} \theta_i x_i - y\right)$$
$$= (h_\theta(x) - y) x_j$$
$$\Rightarrow \theta_j := \theta_j + \alpha (y^i - h_\theta(x^i)) x_j^i$$

reference: Stanford CS229 Lecture Notes.