Hypothesis:

$h_\theta(x) = \theta_0 + \theta_1 x$

Cost function:

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

Gradient Descent algorithm:

$\text{repeat until convergence:}\{$
$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$
$\}$

Let us first expand the term partial derivative term, by subbing in our cost function define above.

We expand the $\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ term:

$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 \tag{4}$$

Subbing in for $h_\theta(x^{(i)})$:

$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2 \tag{5}$$

We will now take the partial derivatives of θ_0 and θ_1.

The equation for the partial derivative of the cost function with respect to θ_0 is shown below:

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2 \tag{6}$$

Applying the Chain Rule to (6) gives the following:

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = 2 \cdot \frac{1}{2m} \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^{2-1} \left[\frac{\partial}{\partial \theta_0} \left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)\right] \tag{7}$$

Looking at the terms in the square brackets, we take the partial derivative of the term with respect to θ_0. All other terms are treated as constants, and the derivative of a constant is zero. We can therefore interpret (7) as:

$$\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)\left[\frac{\partial}{\partial\theta_0}(\theta_0 + \text{constant} - \text{constant})\right]$$

The derivative of θ_0 with respect to θ_0 is 1, so we have:

$$\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)\cdot 1 \tag{8}$$

$$\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right) \tag{9}$$

Similarly, we take the partial derivative with respect to θ_1:

$$\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_1}\frac{1}{2m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^2 \tag{10}$$

We apply the Chain Rule to (10) to arrive at the equation below:

$$\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = 2\cdot\frac{1}{2m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)^{2-1}\left[\frac{\partial}{\partial\theta_1}\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)\right] \tag{11}$$

Since we wish to take the derivative with respect to θ_1 we treat other terms in the square braces as constants:

$$\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)\left[\frac{\partial}{\partial\theta_0}\left(\text{constant} + \theta_1 x^{(i)} - \text{constant}\right)\right]$$

The derivate $\theta_1 x^{(i)}$ with respect to θ_1 is simply $x^{(i)}$, and so we have:

$$\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x^{(i)} - y^{(i)}\right)\cdot x^{(i)} \tag{12}$$

$$\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)\cdot x^{(i)} \tag{13}$$

Finally, we are left with the following algorithms for Gradient Descent:

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \tag{14}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \cdot x^{(i)} \tag{15}$$