

# Gradient Descent For Multiple Variables

## Gradient Descent for Multiple Variables

The gradient descent equation itself is generally the same form; we just have to repeat it for our 'n' features:

```
repeat until convergence: {
 $\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x_0^{(i)}$ 
 $\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x_1^{(i)}$ 
 $\theta_2 := \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x_2^{(i)}$ 
...
}
```

In other words:

```
repeat until convergence: {
 $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$       for j := 0...n
}
```

The following image compares gradient descent with one variable to gradient descent with multiple variables:

### Gradient Descent

Previously (n=1):

Repeat {

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

$\frac{\partial}{\partial \theta_0} J(\theta)$

$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_1^{(i)}$

(simultaneously update  $\theta_0, \theta_1$ )

}

### New algorithm (n ≥ 1):

Repeat {

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

(simultaneously update  $\theta_j$  for  $j = 0, \dots, n$ )

}

---

$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_0^{(i)}$

$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_1^{(i)}$

$\theta_2 := \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_2^{(i)}$

...

*Handwritten notes in red:*

- For the first algorithm, the term  $\frac{\partial}{\partial \theta_0} J(\theta)$  is boxed, and  $x_1^{(i)}$  is circled with a note  $x_1^{(i)}$ .
- For the new algorithm, the term  $\frac{\partial}{\partial \theta_j} J(\theta)$  is written above the first equation, and  $x_0^{(i)}$  is circled with a note  $x_0^{(i)}$ .