Optimization Algorithms

Gradient descent :

It is Optimization algorithm used in machine learning to tweak the parameters iteratively to minimize the given cost function(differentiable function) to its local minimum.

The idea is to take repeated steps in the opposite direction of the gradient (or approximate gradient) of the function at the current point, because this is the direction of steepest descent.

Steps :

1. Start with some random values parameters.
2. Keep changing the parameters until we end up at a minimum value of cost function.

Algorithm :

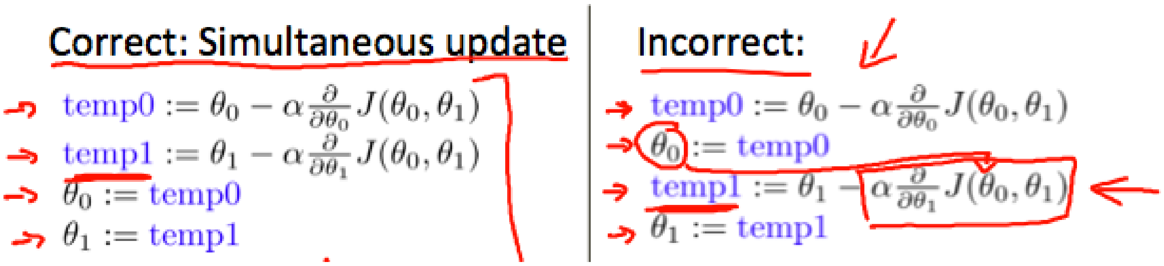
Repeat until converges {

$$\Theta \mathrm{j} = \Theta \mathrm{j} - \alpha \frac{\partial \mathrm{J}}{\partial \theta} J(\theta_0, \theta_1, \dots \dots, \theta_n)$$

}

Here " ***α*** " is the learning rate, it decides how big the step taken by algorithm to reach local minimum.

We must simultaneously update the parameter values. Suppose we have 2 parameters θ_0, θ_1 then we must update the parameters as follows :



Correct: Simultaneous update

$temp0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$

$temp1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$

$\theta_0 := temp0$

$\theta_1 := temp1$

Incorrect:

$temp0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$

$\theta_0 := temp0$

$temp1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$

$\theta_1 := temp1$

In the incorrect one the updated theta_0 will be used in updating the theta_1 which is an error, so we follow simultaneous update in gradient descent algorithm.