Gradient Descent Explanation

For any model the ultimate goal is to reduce the cost and gradient descent is doing great in reducing the cost or minimizing the cost for a model.

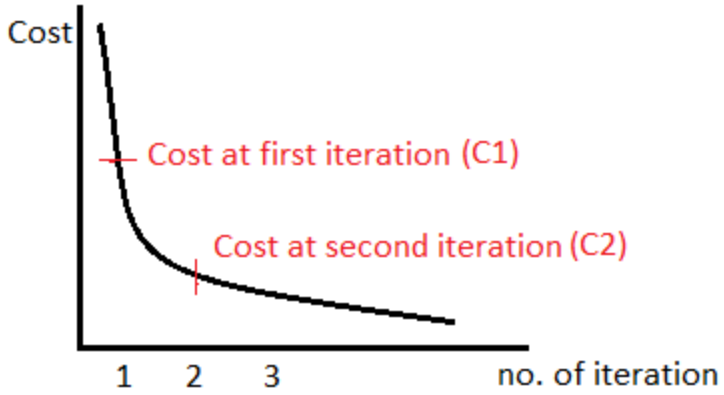
Cost of linear regression is:

$$J(\theta_0, \theta_1) = \sum_{i=0}^{m} h(\theta) \; - \; y, \; where \; h(\theta) = \theta_0 + \theta_1 x$$

where m = number of data points.

Looking at above equations it is understood that bias and variance must be tuned in such
a way that the predictions are very much close to line of fit and as a result the cost of the model will be minimum.

Start with some random values for bias and variance, run model, predict values for test data and calculate cost. If the cost is higher than expected,
adjust bias and variance, and repeat the whole process. Expectation is with each iteration the cost should reduce. If we plot cost with number of iterations, it should be downwards slope.



In the above diagram if cost needs to be reduced from C1 to C2, the bias and variance needs to be reduced.

How we calculate the difference of bias and variance at C1 and C2 is simply by taking derivatives of bias and variance with respect to cost. Once we get derivatives, reduce

the
differentiation's from the original estimated bias and variance and
those will be new values for bias and variance.

Let’s, say bias as θ_0 and variance as θ_1 for simplicity of notations.

$dJ/d\theta_0 = \sum_{i=i}^{m}(h_\theta(x^i) - y^i)(x_0^i)$

$dJ/d\theta_j = \sum_{i=1}^{m}(h_\theta(x^i) - y^i)(x_j^i)$

where j is number of features.

$new\theta_0 = \theta_0 - \lambda * dJ/d\theta_0$

$new\theta_j = \theta_j - \lambda * dJ/d\theta_j$

λ : hyperparameter to control by what amount derivative amounts should reduce. Also known as learning rate.

λ can be increased at initial iterations because there the cost reduction is maximum and slope is quite steep, and as we reach near to the bottom we can reduce \lambda value to take small steps towards minimum cost so that we should not miss the minimum cost and gets lost.

Use these new values for bias and variance and retrain model.

The iterations will keep going until either of three conditions match:

1. Execution of iterations reach to configured number of iterations.
2. Cost is reduced to expected levels.
3. Cost if not decreasing after few iterations.