

Gradient Descent – An Introduction

Gradient Descent

Gradient Descent Gradient Descent is an optimization algorithm to find the minimum of the function.

Linear Regression Cost Function

In a linear regression, we refer to the function as the Cost Function, which is the Mean Square Error ie

$$\text{Mean Square Error} = (\sum (y - \hat{y})^2) / n$$

in other words

$$\text{since we know } \hat{y} = mx + b$$

$$\text{Mean Square Error} = (\sum (y - (mx + b))^2) / n$$

So our goal is to find the m (slope) and the b (intercept), such that the Mean Square Error is minimum.

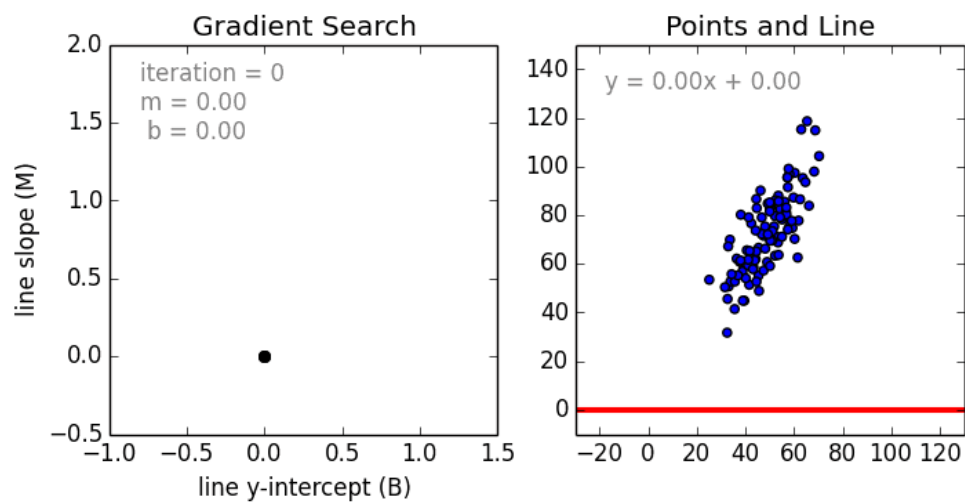
How to Find the minimum

Changing the m and b and such that to look for the minimum error is achieved by the partial derivatives of the m & b, which gives us the delta increment in slope and intercept, ie our direction.

We also need to move towards finding the minimum error, by certain speed that is our Learning Rate, ie the delta m and delta b to be incremented at which rate to find the minimum error. This learning rate needs to be in small size. Normally we consider this to be 0.01 or 0.001 or 0.0001 according to the approximation we look for.

Smaller the Learning Rate, you need to go for many steps of changing the m & b to find the minimum but you will surely hit the minimum where as if the Learning Rate is a bigger number, you are likely to miss out the minimum.

The same is explained in the following picture by *Matt Nedrich* and one picture speaks for it.



source :

https://github.com/mattnedrich/GradientDescentExample/blob/master/gradient_descent_example.gif

Please note the m & b are getting incremented in small terms for each iterations and b becomes static after some iterations, while m is still being adjusted to reach the best fitment.