Gradient Descent

Raktim Gautam Goswami and Abhishek Bairagi

December 9, 2018

Abstract

This is a write-up on gradient descent method used in linear regression model in Machine Learning

Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point.

Use of gradient descent in machine learning : While training the linear regression algorithm, initially we give random values to the weights (W) and biases (b). We find the output using these. The predicted output is then compared to the actual output (also called the label). We find the MSE (mean square error) as

$$Err = \sum_i (Y_{i,predicted} - Y_{i,actual})^2$$

Now, we take the derivative of RMSE w.r.t. W and call this delta lossW ($\Delta lossW$). Therefore,

$$\Delta lossW = \partial Err / \partial W$$

Using this we find the change in W (ΔW) as

$$\Delta W = \Delta lossW \times lr$$

.Now, we take the derivative of RMSE w.r.t. b and call this delta lossb ($\Delta lossb$). Therefore,

$$\Delta lossb = \partial Err / \partial b$$

Using this we find the change in b (Δb) as

$$\Delta b = \Delta lossb \times lr$$

, where lr is the learning rate, which defines the magnitude of the step to be taken towards minima of the error function. The new W and b are written as

$$W_{new} = W - \Delta W$$

$$b_{new} = b - \Delta b$$

The above process repeats for a number of times till we get the least loss.

Further references : For better understanding of gradient descent, one can take reference from this video on a 3D animation of gradient descent.
Also, one can look at this video for more details and a better understanding on gradient descent.