Gradient Descent

Contents

- Introduction
- Formula
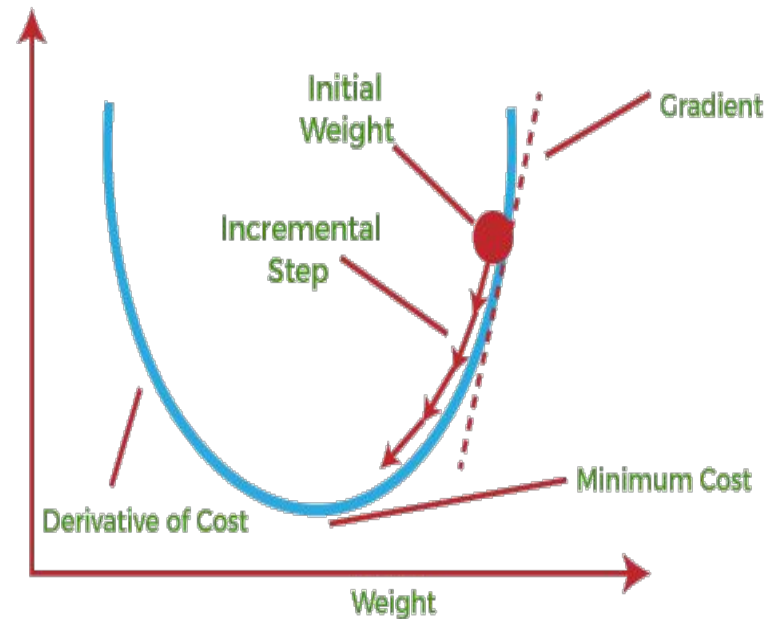- Minimization algorithm
- Types of Gradient descent

Introduction

- Gradient Descent is known as one of the most commonly used optimization algorithms to train machine learning models by means of minimizing errors between actual and expected results. Further, gradient descent is also used to train Neural Networks.
- It measures how much the output of function changes if we change the inputs a little.
- It can also be thought of as the slope of function ,where higher gradient means the sleeper slope, which ultimately makes the model learn faster. It helps in finding the local minimum of a function.

Local minimum or local maximum of a function using gradient descent :

●	If we move towards a negative gradient or away from the gradient at the current point, it will give the local minimum of that function.

●	Whenever we move towards a positive gradient or towards the gradient of the function at the current point, we will get the local maximum of that function.
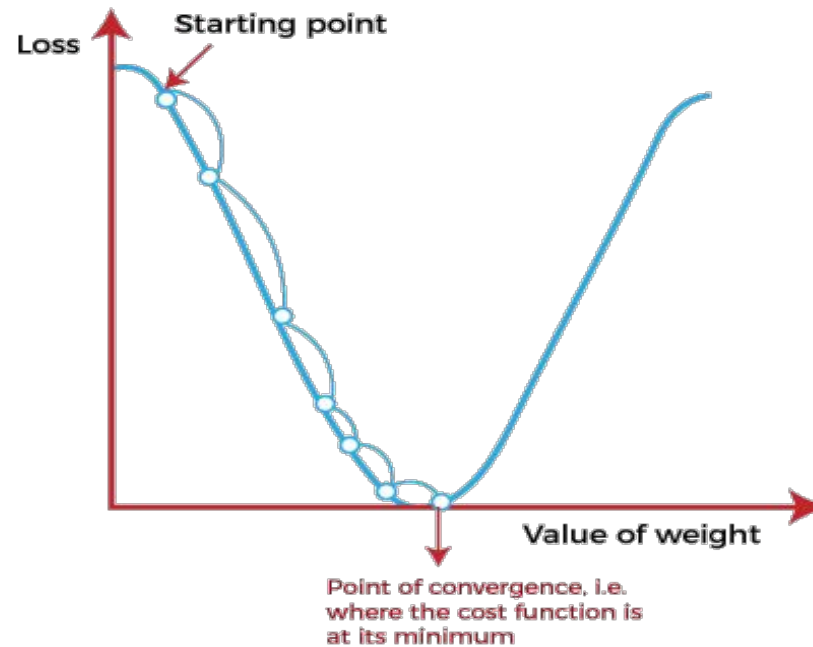
Formula:

$$\mathbf{b} = \mathbf{a} - \gamma \, \nabla \, \mathbf{f(a)}$$

- b=new value

- a=current value

- $-$ refers to minimization part of gradient descent

- here γ is learning rate

- $\nabla \, \mathbf{f(a)}$ gradient term is direction of the steepest descent

Minimization algorithm

- Working of gradient descent algorithm:

- To know the how actually gradient descent algorithm work we find out the slope of a line from linear regression
  Y =mx+C

- At starting point, we will derive the first derivative or slope and then use a tangent line to calculate the steepness of this slope. Further, this slope will inform the updates to the parameters (weights and bias).
- here we need to find the values of weights and bias with some random numbers and gradient descent start at this point .then It takes one step after the other in the steepest downside direction.
- It reaches the point where the cost function is as small as possible.which is called **a point of convergence.**

Types of gradient descent

- :**Batch gradient descent (BGD):** It is used to find the error for each point in the training set and update the model after evaluating all training examples.

- **Stochastic gradient descent (SGD)** : it processes a training epoch for each example within a dataset and updates each training example parameters one at a time.

- **Mini Batch gradient descent:** It is the combination of both batch gradient descent and stochastic gradient descent. It divides the training datasets into small batch sizes then performs the updates on those batches separately.

Thank you