Gradient Descent Algorithm

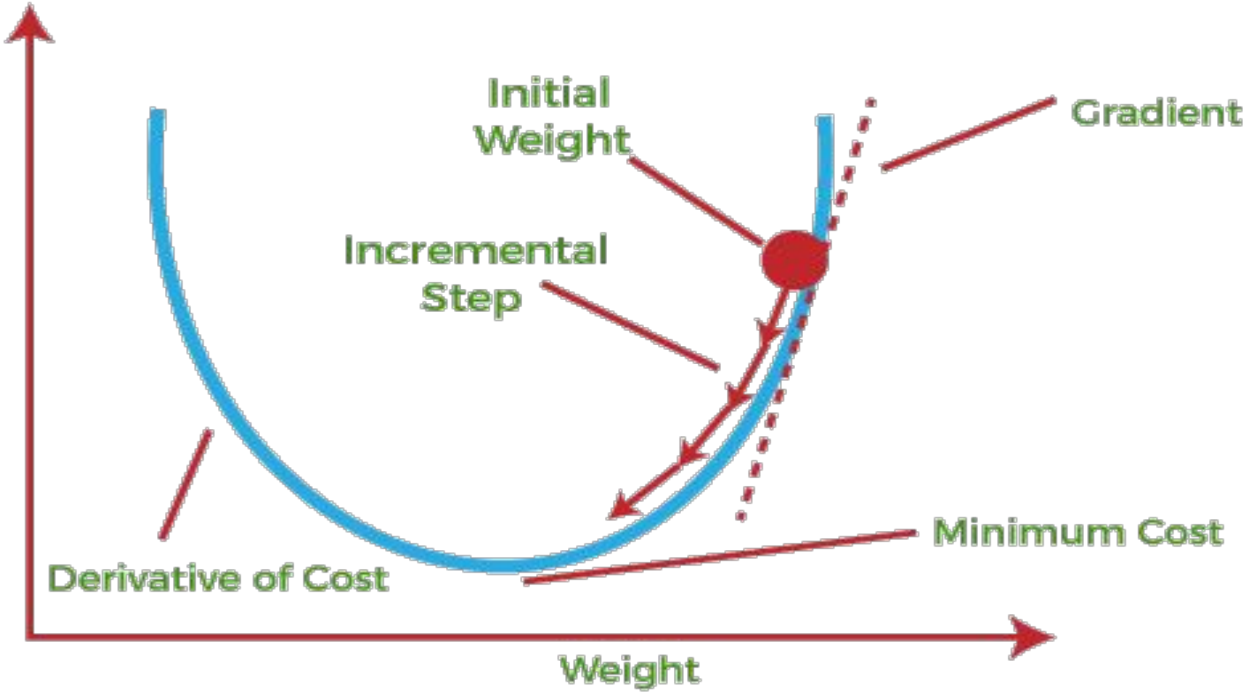
Sayeda Suaiba Anwar
Lecturer, Department of CSE
East Delta University

Definition

- 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.
- Until the function is close to or equal to zero, the model will continue to adjust its parameters to yield the smallest possible error.
- ***It helps in finding the local minimum of a function.***

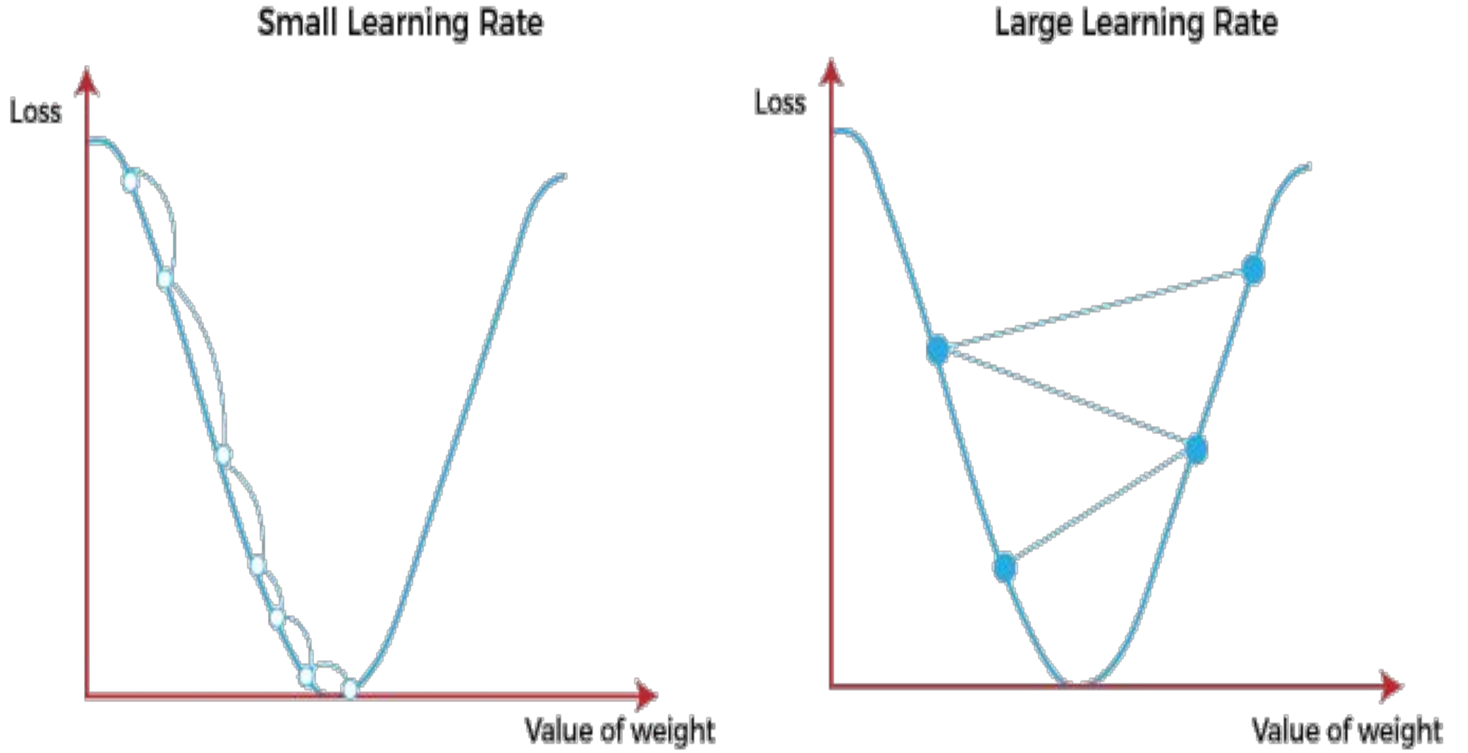
How does Gradient Descent work?

A



Learning Rate

- The size of the steps that are taken to reach the minimum.
- This is typically a small value, and it is evaluated and updated based on the behavior of the cost function. High learning rates result in larger steps but risks overshooting the minimum.
- Conversely, a low learning rate has small step sizes. While it has the advantage of more precision, the number of iterations compromises overall efficiency as this takes more time and computations to reach the minimum.


Small Learning Rate
Loss
Value of weight
Large Learning Rate
Loss
Value of weight

Cost Function

- measures the difference, or error, between actual output and predicted output at its current position.
- This improves the machine learning model's efficiency by providing feedback to the model so that it can adjust the parameters to minimize the error and find the local or global minimum.
- It continuously iterates, moving along the direction of steepest descent until the cost function is close to or at zero. At this point, the model will stop learning.

Types of Gradient Descent

- **Batch Gradient Descent :**

Batch gradient descent sums the error for each point in a training set, updating the model only after all training examples have been evaluated. This process referred to as a training epoch.

While this batching provides computation efficiency, it can still have a long processing time for large training datasets as it still needs to store all of the data into memory.

- It produces less noise in comparison to other gradient descent.
- It produces stable gradient descent convergence.

Types of Gradient Descent

- **Stochastic Gradient Descent :**

Stochastic gradient descent (SGD) runs a training epoch for each example within the dataset and it updates each training example's parameters one at a time.

It shows some computational efficiency losses in comparison to batch gradient systems as it shows frequent updates that require more detail and speed. Further, due to frequent updates, it is also treated as a noisy gradient.

- It is easier to allocate in desired memory.
- It is relatively fast to compute than batch gradient descent.
- It is more efficient for large datasets.

Types of Gradient Descent

- **Mini Batch Gradient Descent:**

Mini Batch gradient descent 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.

- It is easier to fit in allocated memory.
- It is computationally efficient.
- It produces stable gradient descent convergence.

Vanishing and Exploding Gradients

Vanishing Gradients:

- This occurs when the gradient is too small.
- During backpropagation, this gradient becomes smaller that causing the decrease in the learning rate of earlier layers than the later layer of the network.
- Once this happens, the weight parameters update until they become insignificant.

Vanishing and Exploding Gradients

Exploding Gradients:

- This happens when the gradient is too large, creating an unstable model.
- In this case, the model weights will grow too large, and they will eventually be represented as NaN.
- One solution to this issue is to leverage a dimensionality reduction technique, which can help to minimize complexity within the model.

Suggested Reading

- https://towardsdatascience.com/an-introduction-to-gradient-descent-and-backpropagation-in-machine-learning-algorithms-a14727be70e9
- https://builtin.com/data-science/gradient-descent