

Gradient Descent

Gradient Descent is a way to minimize loss function

Learning Rate is Size of Step we take to reach minimum.

[Gradient Descent in Machine Learning: Python Examples \(vitalflux.com\)](https://vitalflux.com/gradient-descent-explained-simply-with-examples/#google_vignette)

https://vitalflux.com/gradient-descent-explained-simply-with-examples/#google_vignette

[What Is Gradient Descent? | Built In](https://builtin.com/data-science/gradient-descent)

<https://builtin.com/data-science/gradient-descent>

What is Gradient Descent?

The gradient descent algorithm is an **optimization technique** used to **minimize a function**, commonly referred to as the **objective function** in machine learning. This objective function, often termed as the **cost or loss function**, is a measure of the error, typically the **squared difference between actual values and predictions**. In machine learning, gradient descent aims to find the most optimal values of parameters or weights that minimize this loss function.

Thus, “gradient descent” is aptly named as it describes the process of descending (minimizing) a function by moving against (opposite to) its gradient. The following plot can be used to understand the gradient descent algorithm.

Gradient Descent Learning Rate

How big the steps gradient descent takes into the direction of the local minimum are determined by the learning rate, which figures out how fast or slow we will move towards the optimal weights.

For the gradient descent algorithm to reach the local minimum we must set the learning rate to an appropriate value, which is neither too low nor too high. This is important because if the steps it takes are too big, it may not reach the local minimum because it bounces back and forth between

the convex function of gradient descent (see left image below). If we set the learning rate to a very small value, gradient descent will eventually reach the local minimum but that may take a while

Learning rate must be chosen wisely as:

1. if it is too small, then the model will take some time to learn.
2. if it is too large, model will converge as our pointer will shoot and we'll not be able to get to minima.

