

# Stochastic gradient descent

VO Van Nghia

PHAM Tuan Kiet

## Abstract

Stochastic gradient descent is an iterative method for optimizing an objective function regarded as a stochastic approximation of gradient descent optimization. Since it replaces the actual gradient (calculated from the entire data set) by an estimate thereof (calculated from a randomly selected subset of the data). Especially in high-dimensional optimization problems this reduces the very high computational burden, achieving faster iterations in trade for a lower convergence rate.

## 1 Gradient descent

An approximate and iterative technique for mathematical optimization is the gradient descent algorithm. It can be used to approach any differentiable function's minimum. For a better imagination about the algorithm, the process is somehow the same as a dropping drip of water following the slope.

The main problem of this algorithm is that sometimes the solution falls into a local minimum point or a saddle point instead of the desired global minimum. It is frequently used internally by data science and machine learning to optimize model parameters. For instance, gradient descent is used by neural networks to find weights and biases.

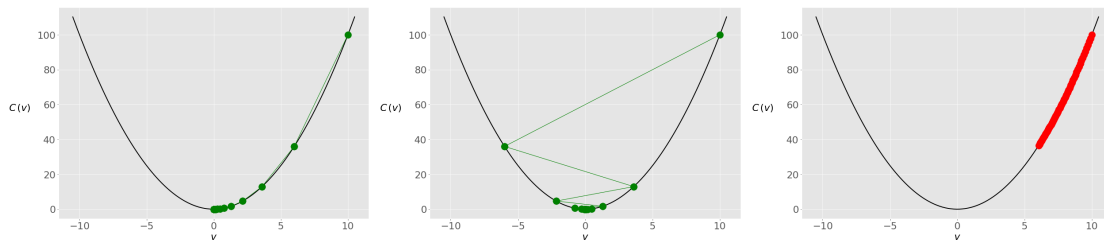


Figure 1: Impact of learning rate

Learning rate is a very important parameter that affects the performance of the algorithm. Right here is an example showing how this value influences a process of gradient descent. With a too small rate, the algorithm seems to be far away from the desired result even after a large amount of iteration. However, a high learning rate doesn't give a better performance as it cannot approach the local minimum.

## 2 Stochastic gradient descent

Stochastic gradient descent begins with an initial vector of decision variables and updates it over a number of iterations, much like in the case of the ordinary gradient descent. What takes place during the iterations differs between the two:

- Stochastic gradient descent randomly divides the set of observations into minibatches
- The gradient is calculated and the vector is moved for each minibatch.
- You can declare that an iteration, or epoch, is complete when all minibatches have been used and move on to the next one.