DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

Nesterov Accelerated Gradient (NAG)

Nesterov Accelerated Gradient, also known as Nesterov momentum or Nesterov's accelerated gradient descent, is an optimization technique that improves upon the standard momentum method. It was introduced by Yurii Nesterov in 1983 and has gained significant attention in recent years due to its superior convergence properties.

The key idea behind NAG is to take into account the momentum term in the calculation of the gradient, by considering the future position of the parameter. Unlike standard momentum, which updates the parameters based on the current position, NAG incorporates an estimation of the future position by looking ahead. By doing so, NAG achieves faster convergence and better handling of oscillations in the loss landscape.

First you calculate the lookahead

$$W_{la} = W_t - \beta V_{t-1}$$

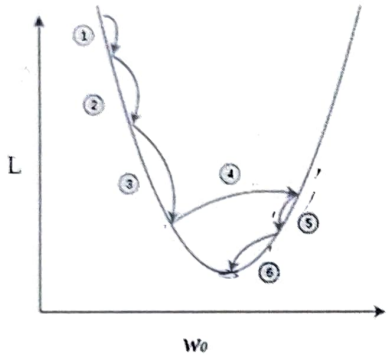
For this, you use the momentum.

Then calculate the slope at that lookahead

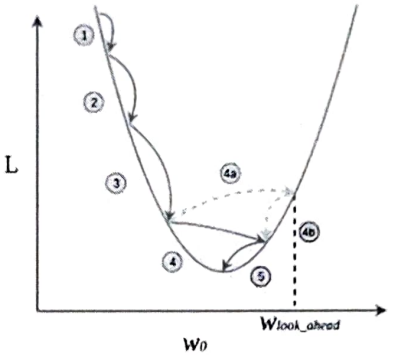
$$\therefore V_t = \beta V_{t-1} + \eta \nabla W_{la}$$

Now update final weight:

$$W_{t+1} = W_t - V_t$$



(a) Momentum-Based Gradient Descent



(b) Nesterov Accelerated Gradient Descent