**Day 19**

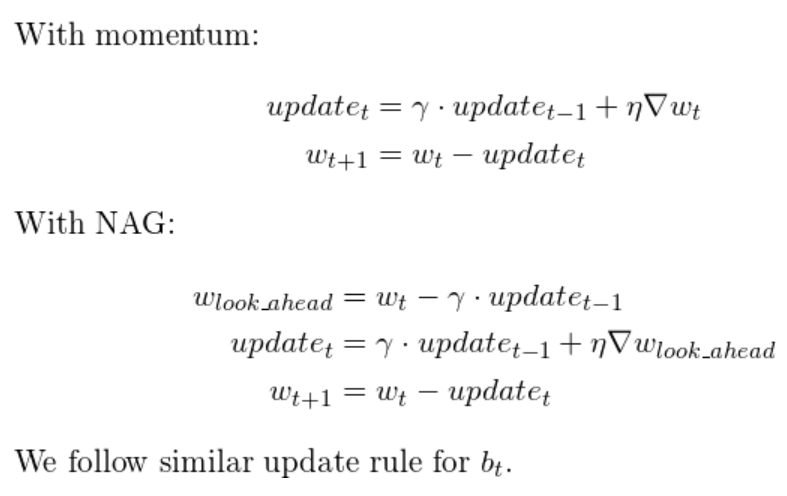
**What to do?**

Learn about Nesterov accelerated gradient.

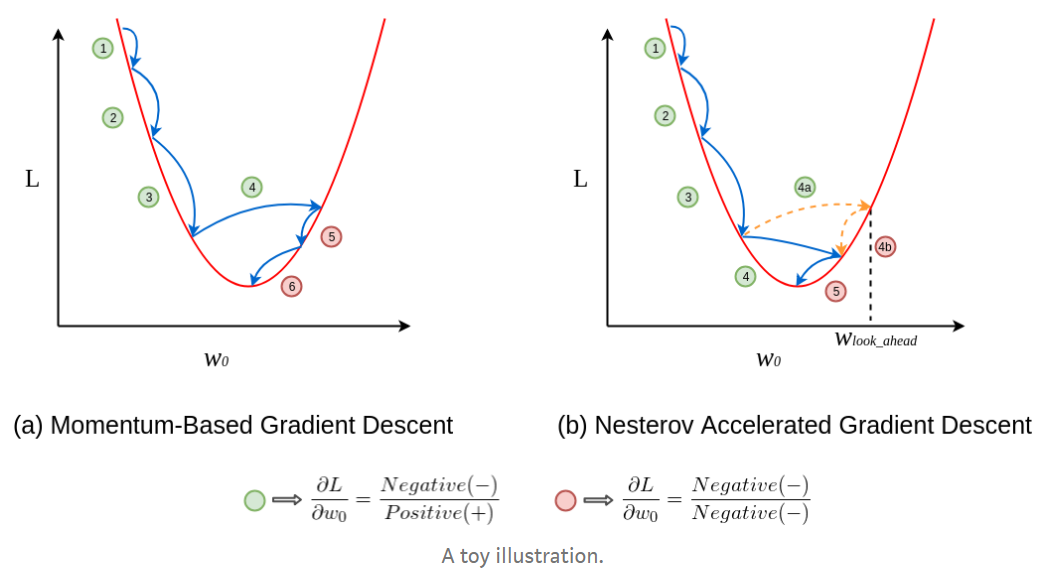
**Nesterov accelerated gradient:**

In single phrase, NAG means to “look-ahead”. Usually in momentum gradient descent, the next step of the gradient is updated on the basis of the current step (update\_t = beta \* update\_t-1 + (1-beta)\*dW). NAG idea comes up by asking the question of “why not look ahead before update.

**NAG update rule:**

****

To understand what is happening within the equations, let us take a look at the graphs of momentum and NAG.



As seen in the Graph (a), it is seen that there are positive updates from step 1 to 3, with increase in each update. When it comes to update 4, overshoots to other side of the graph and then decrease to reach local optima.

In Graph (b), updates 1 to 3 are like Graph (a) however, when it comes to update 4, the algorithm overlooked the overshoot of the gradient and skips the step by calculating the w\_lookahead. The w\_lookahead value ensure that the gradients reach local optima with limiting unnecessary noise/overshoot.