PadhAI: Variants of Gradient Descent

One Fourth Labs

Dissecting the update rule for momentum based gradient descent

Can we dissect the equations in more detail?

1. Let us further dissect the momentum based Gradient Descent
2. $\upsilon_t = \gamma * \upsilon_{t-1} + \eta \nabla \omega_t$ this variable is called the history.
3. $\omega_{t+1} = \omega_t - \upsilon_t$ this variable represents the current movement to be made
4. Consider every instance in time denoted by the subscript, ranging from 0 to t
5. $\upsilon_0 = 0$
6. $\upsilon_1 = \gamma * \upsilon_0 + \eta \nabla \omega_1 = \eta \nabla \omega_1$
7. $\upsilon_2 = \gamma * \upsilon_1 + \eta \nabla \omega_2 = \gamma.\eta \nabla \omega_1 + \eta \nabla \omega_2$
8. $\upsilon_3 = \gamma * \upsilon_2 + \eta \nabla \omega_3 = \gamma(\gamma.\eta \nabla \omega_1 + \eta \nabla \omega_2) + \eta \nabla \omega_3$
   a. $\upsilon_3 = \gamma^2.\eta \nabla \omega_1 + \gamma.\eta \nabla \omega_2 + \eta \nabla \omega_3$
9. $\upsilon_4 = \gamma * \upsilon_3 + \eta \nabla \omega_4 = \gamma^3.\eta \nabla \omega_1 + \gamma^2.\eta \nabla \omega_2 + \gamma^1.\eta \nabla \omega_3 + \eta \nabla \omega_4$

   .
   .
   .
10. $\upsilon_t = \gamma * \upsilon_{t-1} + \eta \nabla \omega_t = \gamma^{t-1}.\eta \nabla \omega_1 + \gamma^{t-2}.\eta \nabla \omega_2 + \dots + \eta \nabla \omega_t$ t
11. Here, we take an Exponentially Decaying Weighted Sum, whereby as we move further and further into the series, the weight decays more.
12. The intuition behind this is as we progress further and further down a series/direction, we can place lesser and lesser importance to the later gradients as we move along the same direction.