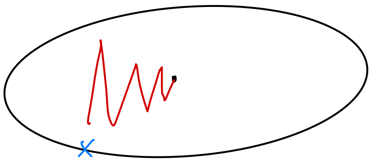
GD

$$x_{t+1} = x_t - \alpha \nabla f(x_t)$$



불필요한 왕복운동

+ Momentum

$$v_{t+1} = \rho v_t - \alpha \nabla f(x_t)$$

$$x_{t+1} = x_t + v_{t+1}$$

- oscillation ⇓ (진동 ⇓)
- fast convergence (가속 ⇑)
- escape local min (관성)

Nesterov Momentum

$$v_{t+1} = \rho v_t - \alpha \nabla f(x_t + \rho v_t)$$

$$x_{t+1} = x_t + v_{t+1}$$

smooth.

⇒ 방향 컨트롤.

AdaGrad

$$x = x - \alpha \cdot \frac{\nabla f(x_t)}{\sqrt{\sum (\nabla f(x_i))^2} + \varepsilon}$$

$\| \nabla f(x) \| \Uparrow$ "가파르다"

$$\begin{cases} \text{가파르면} \rightarrow \text{느리게} \\ \text{완만하면} \rightarrow \text{빠르게} \end{cases}$$

⚠ $\sqrt{\Sigma (\nabla f(x_i))^2} \longrightarrow \infty$

RMS Prop

$$\begin{cases} E[g^2]_t = \beta E[g^2]_{t-1} + (1-\beta) g_t^2 . \\ x = x - \dfrac{\eta \, g_t .}{\sqrt{E[g^2]_t} + \varepsilon} \end{cases}$$

먼 과거의 정도를 약하게!

⇒ 속도 컨트롤.

Adam

$$\boxed{x_{t+1} = x_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}}$$

(방향) $m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$

(속도) $v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$

(편향 보정) $\hat{m}_t = \dfrac{m_t}{1-\beta_1^{\,t}}$

$m_0 = 0$

$v_0 = 0$ $\qquad \hat{v}_t = \dfrac{v_t}{1-\beta_2^{\,t}}$

pf. let $\nabla f = g$ (const)

$$\begin{cases} m_k = (1-\beta_1^{\,k}) g \longrightarrow \hat{m}_k = g \\ v_k = (1-\beta_2^{\,k}) g^2 \longrightarrow \hat{v}_k = g^2 . \end{cases}$$