Optimization algorithms (part 2)

RProp - Resilient Propagation

$\underline{\underline{1.5}}$ $\quad \underline{\underline{l \to 0}}$

We have $l(w) = 4w^2 - 12w + 9$

$\nabla l(w) = 8w - 12$ $\qquad g = \nabla l(w)$

$\Rightarrow \boxed{g_t = 8w_t - 12}$

Case 1:- Let us assume $w_{t-1} = 1, \quad w_t = 2$

$\Rightarrow g_{t-1} = 8w_{t-1} - 12 = 8(1) - 12 = -4$

$\Rightarrow \boxed{g_{t-1} = -4}$

Also $g_t = 8w_t - 12 = 8(2) - 12 = 4$

$\Rightarrow \boxed{g_t = 4}$

We see $g_t \cdot g_{t-1} = -16 < 0$

We have $\beta = 0.1$ such that

$t-1$

t

$t+1 \to ? \; \underline{\underline{1.1}}$

$\Rightarrow w_{t+1} = w_{t-1} + \beta \Delta w_t$

$\Delta w_t = w_t - w_{t-1}$

$= w_{t-1} + \beta (w_t - w_{t-1})$

$= 1 + 0.1(2-1)$

$\boxed{w_{t+1} = 1.1}$

Case 2:- We assume $w_{t-1} = \frac{1}{2} \Rightarrow g_{t-1} = 8\left(\frac{1}{2}\right) - 12 = -8$ $\qquad 0.5$

$w_t = \frac{3}{4} \Rightarrow g_t = 8\left(\frac{3}{4}\right) - 12 = -6$ $\qquad 0.75$

We get $g_t \cdot g_{t-1} > 0$ We will have η such that

$0.5 \to 0.75 \to 1.5$

$w_{t+1} = w_t + \eta \Delta w_t$

$= \frac{3}{4} + 1.5(w_t - w_{t-1})$ $\qquad$ [we assume $\eta = 1.5$]

$$\underline{\underline{w_{t+1}}} = \underline{w_t} + \eta \underbrace{\Delta w_t}$$

$$= \frac{3}{4} + 1.5(w_t - w_{t-1}) \quad \left[\text{we assume } \eta = \underline{\underline{1.5}}\right]$$

$$= \frac{3}{4} + \underline{1.5}\left(\frac{3}{4} - \frac{1}{4}\right)$$

$$= \underline{\underline{1.5}}$$

$g_t > 0$, $g_{t-1} < 0$

updation rule Rprop -

We have $\underline{w_t}$, $\underline{\underline{w_{t-1}}}$ ⇒ Calculate $\underline{\underline{g_t}}$ & $\underline{\underline{g_{t-1}}}$

If $\underline{g_t} \cdot \underline{g_{t-1}} > 0$

$$w_{t+1} = \underline{\underline{w_t}} + \underline{\eta} \underbrace{\Delta w_t}_{w_t - w_{t-1}} \quad [\eta > 1]$$

else

$$w_{t+1} = \underline{\underline{w_{t-1}}} + \underline{\beta} \underbrace{\Delta w_t} \quad [\beta < 1]$$

$\underline{\underline{w_t}} \rightarrow$ ✗

Gradient Descent with Momentum

$$w_{t+1} = w_t - \eta \underbrace{g_t}_{\underline{\nabla l(w_t)}}$$

Momentum update rule -

$$w_{t+1} = \underline{\underline{w_t}} + \underline{\underline{\beta \Delta w_t}} - \underline{\underline{\eta g_t}}$$

Here $\Delta w_t = w_t - w_{t-1}$

$$w_{t+1} = \underline{\underline{w^{G}_{t+1}}} + \boxed{\beta \underbrace{\underline{\Delta} w_t}_{\underline{\underline{w_t - w_{t-1}}}}}$$

Now $\underline{w_1} = \underline{w_0} + \underline{\beta}\Delta w_0 - \underline{\eta g_0} = \overbrace{w_0 - \eta g_0}^{GD} \quad [\because \underline{\Delta} w_0 = 0]$

$\underline{w_2} = \underline{w_1} + \beta \underline{\Delta} w_1 - \underline{\eta g_1} = \underline{w_1} + \beta(\underline{w_1} - \underline{w_0}) - \eta \underline{g_1}$

$w_3 = w_2 + \beta \Delta w_2 - \eta g_2 = w_2 + \beta(w_2 - w_1) - \eta g_2$

$$w_3 = w_2 + \beta \Delta w_2 - \eta g_2 = w_2 + \beta \underbrace{(w_2 - w_1)} - \eta g_2$$

$\vdots$

We have $l(w) = 4w^2 - 12w + 9 \Rightarrow l'(w) = 8w - 12$

$\Rightarrow \boxed{g_t = 8w_t - 12.}$ We assume $w_0 = 5$, $\eta = 0.2$, $\beta = 0.1$

Now

$$w_1 = w_0 - \eta g_0 = 5 - 0.2(8(5) - 12) = -0.6$$

$$w_2 = w_1 + 0.1(-0.6 - 5) - 0.2[8(-0.6) - 12] = 2.2$$

$$w_3 = 2.2 + 0.1(2.2 - (-0.6)) - 0.2[8(2.2) - 12] = 1.36$$

$$w_4 = 1.36 + 0.1(1.36 - 2.2) - 0.2[8(1.36) - 12] = 1.5$$

The solution to $l(w)$ is 1.5 only.

Nesterov Accelerating Gradient

$\Delta w_t = w_t - w_{t-1}$

Nestrov's Accelerated Gradient update

$$\boxed{w_{t+1} = [w_t + \beta \Delta w_t] - \eta \nabla(\ell(w_t + \beta \Delta w_t))}$$

$\nabla \ell(w_t + \beta \Delta w_t)$

We consider $\ell(w) = 4w^2 - 12w + 9$

$$\nabla \ell(w) = 8w - 12 \Rightarrow \nabla \ell(w_t + \beta \Delta w_t) = 8(w_t + \beta \Delta w_t) + 12$$

$$\Rightarrow \nabla \ell(w_t + \beta(w_t - w_{t-1})) = 8w_t + 8\beta(w_t - w_{t-1}) + 12$$

$$\Rightarrow \boxed{\nabla \ell(w_t + \beta(w_t - w_{t-1})) = (8 + 8\beta) w_t - 8\beta w_{t-1} + 12}$$

We consider $w_0 = 5$, $\beta = 0.1$ & $\eta = 0.2$

Now, $w_1 = w_0 + \beta \Delta w_0 - \eta \nabla \ell(w_0 + \beta \Delta w_0) = w_0 - \eta \nabla \ell(w_0)$

$$w_2 = w_1 + \beta \Delta w_1 - \eta \nabla \ell(w_1 + \beta \Delta w_1)$$

$$= w_1 + \beta(w_1 - w_0) - \eta[(8 + 8\beta) w_1 - 8\beta w_0 + 12]$$

In our case, $w_1 = 5 - 0.2[8(5) - 12] = -0.6$

$$w_2 = -0.6 + 0.1(-0.6 - 5) - 0.2[[8 + 8(0.1)](-0.6) - 8(0.1)5 + 12]$$

$$= -1.704$$