SGD with momentum - loss fn

Saturday, December 11, 2021 7:24 PM

$$W_N = W_0 - \eta \left.\frac{\partial L}{\partial W}\right|_{W=W_0} \rightarrow (I)$$ - update

$$m = \beta \cdot m + \eta \frac{\partial L}{\partial W} \rightarrow (II)$$ where β

$$W_N = W_0 - m \Rightarrow W_0 - (\beta m + \eta \frac{\partial L}{\partial}$$

Lets Step 1 $= m_0 = 0$, $\beta = 0.9$ - default

$$m_1 = \beta \cdot m_0 + \eta \left.\frac{\partial L}{\partial W}\right|_{W=W_0}$$

($m_0 \neq 0$; $\beta \rightarrow 0.9$; $m_0 \rightarrow 0$)

$$m_1 = \eta \left.\frac{\partial L}{\partial W}\right|_{W=W_0} \rightarrow (IV) =$$

$$W_1 = W_0 - m_1 \Rightarrow W_0 - \eta \left.\frac{\partial L}{\partial W}\right| W$$

ule – SGD

– momentum parameter

$W) = W_0 - \beta m - \eta \frac{dL}{dW}\Big|_{W=W_0}$

SGD = SGD with momentum

$W = W_0$ → (111)

$$m_1 = \beta \cdot m_0 + \eta \frac{\partial L}{\partial w}$$

$$\Big|_{w=w_0}$$

$$m_2 = \beta m_1 + \eta \frac{dL}{dw}\Big|_{w=w_0}$$

Step 2 → $m_1 = \eta \frac{\partial L}{\partial w}\Big|_{w=w_0}$

$$m_2 = \beta m_1 + \eta \frac{\partial L}{\partial w}\Big|_{w=w_1}$$

$$m_2 = \beta \left[\eta \frac{\partial L}{\partial w}\Big|_{w=w_0}\right] + \left[\eta \frac{\partial L}{\partial w}\Big|_{w=w_1}\right]$$

$$m_2 = \eta \left[\beta \frac{\partial L}{\partial w}\Big|_{w=w_0} + \frac{\partial L}{\partial w}\Big|_{w=w_1}\right]$$

So with momentum update $W_2 = W_1 - m_2 \Rightarrow$

Without momentum $\Rightarrow W_2 =$

$\rightarrow$ (V)

$\eta \left. \frac{\partial L}{\partial w} \right|_{w=w_1} \Big]$

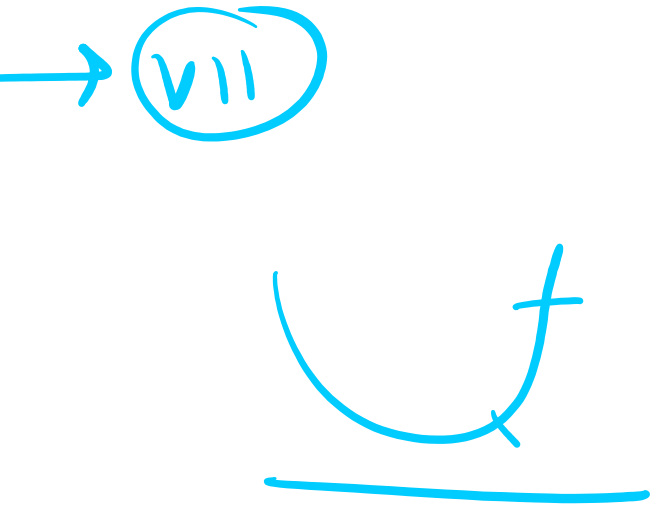
$+ \left. \frac{\partial L}{\partial w} \right|_{w=w_1} \Big]$ $\rightarrow$ (VI)

(A)

$w_1 - \eta \left[\beta \cdot \left. \frac{\partial L}{\partial w} \right|_{w=w_0} + \left. \frac{\partial L}{\partial w} \right|_{w=w_1} \right]$

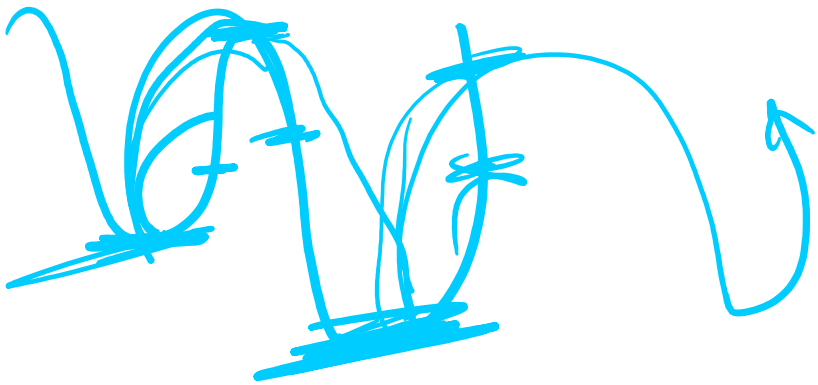
$w_1 - \boxed{\eta \left. \frac{\partial L}{\partial w} \right|_{w=w_1}}$ $= 5$ $\rightarrow$ (b)

$\beta = 0.9$


VII

$$W_2 = W_1 - \eta \left[\frac{\partial Q}{\partial W_1} \right]$$

fixes coefficient



$$W_3 = W_2 - \eta \left[\beta^2 \right]$$

$$\left.\frac{\partial L}{\partial w}\right|_{w=w_0} + \left.\frac{\partial L}{\partial w}\right|_{w=w_1} \Big] \rightarrow (VI)$$

slope at w_0 slope at w_1

$$w_3 = \beta \cdot m_2 + \eta \frac{\partial L}{\partial w}$$

$$\cdot \left.\frac{\partial L}{\partial w}\right|_{w=w_0} + \beta \cdot \left.\frac{\partial L}{\partial w}\right|_{w=w_1} + \frac{\partial}{\partial}$$

$\alpha \cdot \eta$

11

$\Big|_{w = w_2}$

$$\left[\frac{L}{w} \Big|_{w = w_2} \right] \rightarrow (1x)$$

100r-slope at 2