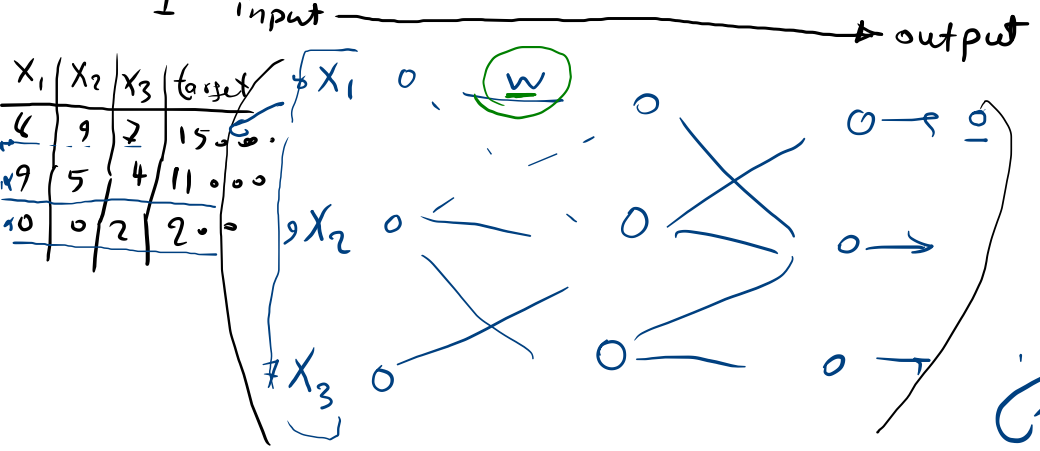
کوه را آمد بست کران سوزن

1



X_1	X_2	X_3	target
4	9	2	15.00
9	5	4	11.00
10	0	2	2.00

۴ وزن را به صورت رندوم مقداردهی می‌کنیم

وزن‌ها را تغییر دهیم تا خروجی با کمترین خطا حاصل شود!

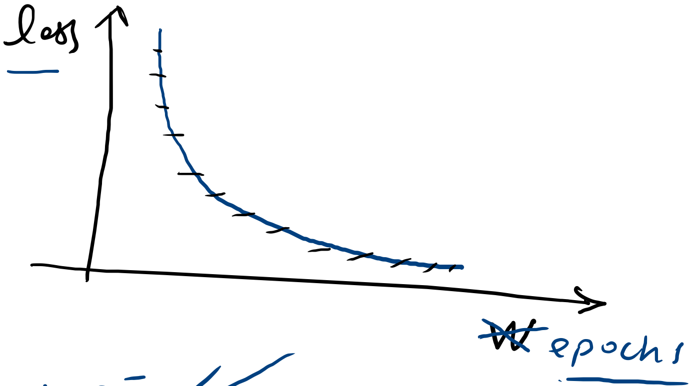
2. loss function (L) RMSE

MSE

SSE

error = $Y_{true} - Y_{predicted}$

۳- آپدیت کردن وزن‌ها در جهت کاهش loss



epoch ها: هر بار که کل دیتاست را تمامش یک بار مدل را ملاقات کرده باشد!

وزن‌ها چگونه آپدیت می‌شوند؟!

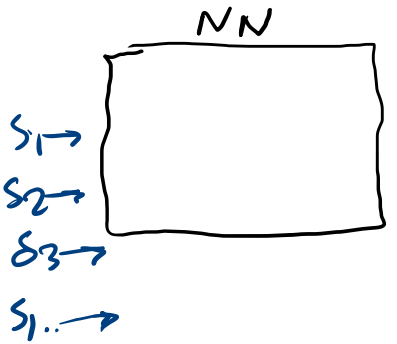
Batch Gradient Descent

زمان وزن‌ها را آپدیت کنیم؟ همه

همه‌ی داده‌ها وارد مدل می‌شوند!

و loss همه‌ی نمونه‌ها حساب می‌شود!

	X_1	X_2	X_3	target
S_1				y_1
S_2				y_2
S_3				y_3
⋮				
S_{100}				y_{100}



$S_1 \rightarrow \hat{y}_1$ / y_1

$S_2 \rightarrow \hat{y}_2$ / y_2

⋮

$S_{100} \rightarrow \hat{y}_{100}$ / y_{100}

$$\text{loss} = \frac{1}{N} \sum_{i=1}^{N} (y_{i,true} - \hat{y}_i)^2$$

$$\text{loss} = \frac{1}{100}\left[(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_{100} - \hat{y}_{100})^2 \right]$$

$$w_{new} = w_{old} - \alpha \frac{\partial L}{\partial w_{old}}$$

(L ← سر)

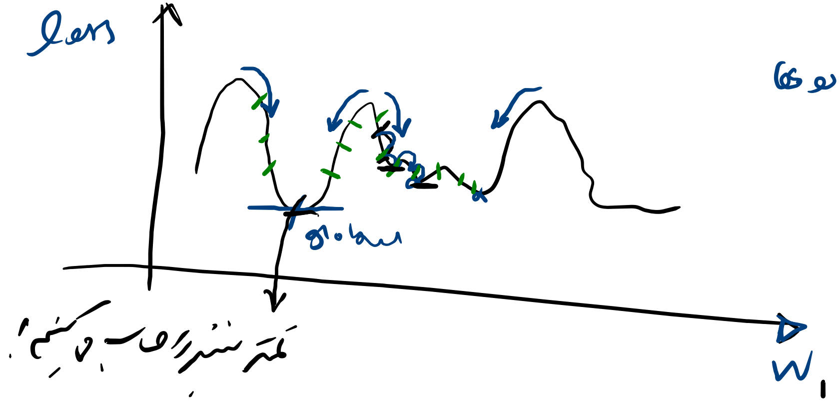
منتظر می‌مونیم loss روی همه‌ی نمونه‌ها حساب بشند! ← Batch Gradient Descent

↓

در هر epoch یک بار وزن‌ها آپدیت می‌شوند!

↓

مشکل: در local min



همه داده‌ها ← یک بار ∇ را بدست بیاوریم!

راه حل

به جای اینکه منتظر باشیم تا همه داده‌ها درون شبکه عصبی وارد شوند

و سپس ∇ را بدست بیاوریم ← به ازای هر sample ∇ را بدست

کار کنیم!

$$\rightarrow \text{loss} = \frac{1}{N} \sum_{i=1}^{N} (y_{i,true} - \hat{y}_i)^2$$

Stochastic gradient descent ⇝ SGD

local minima

one sample $\longrightarrow$ $\text{loss} = (y_{true} - \hat{y})^2$

$$w_{new} = w_{old} - \alpha \frac{\partial L}{\partial w_{old}}$$

Batch Gradient Descent (BGD)

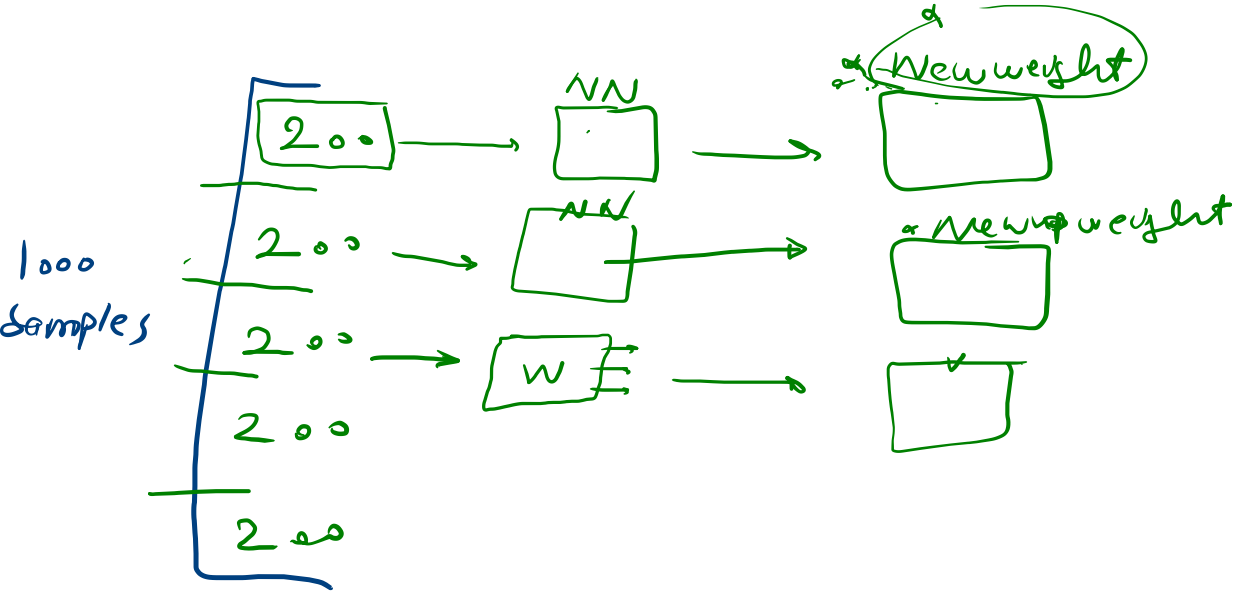
→ وزن‌ها را از روی تمام داده‌ها یک‌جا آپدیت می‌کند

Stochastic Gradient Descent → (SGD)

وزن‌ها را از روی هر sample آپدیت می‌کند

Mini batch Gradient Descent

وزن‌ها را از روی دسته‌های mini batch آپدیت می‌کند


1000 samples
200
200
200
200
200
NN
NN
WE
New weight
New weight

MBGD

$$\text{grad } f(x,y) = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y}$$

گرسین :

$$f(x,y) = x^2 + y^2 \longrightarrow \nabla f = 2x\,\hat{x} + 2y\,\hat{y}$$

مشتق جزئی:

$$f(x,y) = x^2 + y^3$$

$$\frac{\partial f}{\partial x} = 2x$$

$$\frac{\partial f}{\partial y} = 3y^2$$

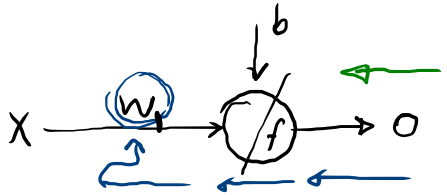
$y = 8(x^5+9)^3 + (x^5+9)^2 + 5$

$\frac{\partial y}{\partial x} = ?$ ✓

$y = 8z^3 + z^2 + 5$

$z = x^5 + 9$

$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial x}$



1. Feed forward process

$net = w_1 X + b$

$O = f(net)$

$\hat{y} = O$

2. $Loss = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$

$-2(y_i - \hat{y})$

$w_{1new} = w_{1old}$

back propagation!

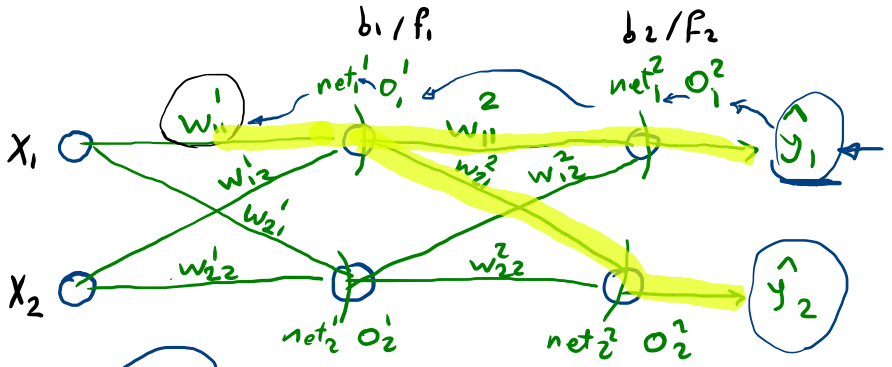
3. backward process

$w_{1new} = w_1 - \alpha \frac{\partial L}{\partial w_1}$

$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O} \frac{\partial O}{\partial net} \frac{\partial net}{\partial w_1}$

$w_{1new} = w_1 - \alpha [\quad]$

$-2(y_i - \hat{y}) \times (1) \left(\frac{\partial O}{\partial net}\right) X$



(1)

$$net_1^1 = w_{11}^1 X_1 + w_{12}^1 X_2 + b_1$$

$$net_2^1 = w_{21}^1 X_1 + w_{22}^1 X_2 + b_1$$

$$O_1^1 = f_1(net_1^1)$$

$$O_2^1 = f_1(net_2^1)$$

$$net_1^2 = w_{11}^2 O_1^1 + w_{12}^2 O_2^1$$

$$net_2^2 = w_{21}^2 O_1^1 + w_{22}^2 O_2^1$$

$$\hat{y}_1 = O_1^2 = f_2(net_1^2)$$

$$\hat{y}_2 = O_2^2 = f_2(net_2^2)$$

② Loss

$$loss = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

(3) $w_{11}^{1}{}_{new} = ?$

$$w_{11}^{1}{}_{new} = w_{11}^{1} - \alpha \frac{\partial L}{\partial w_{11}^{1}}$$

$$\frac{\partial L}{\partial w_{11}^{1}} = \left[\frac{\partial L}{\partial \hat{y}_1} \frac{\partial \hat{y}_1}{\partial o_1^2} \frac{\partial o_1^2}{\partial net_1^2} \frac{\partial net_1^2}{\partial o_1^1} \frac{\partial o_1^1}{\partial net_1^1} \frac{\partial net_1^1}{\partial w_{11}^1} \right] \quad (\hat{y}_1)$$

$$+ \left[\frac{\partial L}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial o_2^1} \frac{\partial o_2^2}{\partial net_2^2} \frac{\partial net_2^2}{\partial o_1^1} \frac{\partial o_1^1}{\partial net_1^1} \frac{\partial net_1^1}{\partial w_{11}^1} \right]$$

$$= \left[\frac{\partial L}{\partial \hat{y}_1} \frac{\partial \hat{y}_1}{\partial o_1^2} \frac{\partial o_1^2}{\partial net_1^2} \frac{\partial net_1^2}{\partial o_1^1} + \frac{\partial L}{\partial \hat{y}_2} \frac{\partial \hat{y}_2}{\partial o_2^2} \frac{\partial o_2^2}{\partial net_2^2} \frac{\partial net_2^2}{\partial o_1^1} \right] \left[\frac{\partial o_1^1}{\partial net_1^1} \frac{\partial net_1^1}{\partial w_{11}^1} \right]$$


END