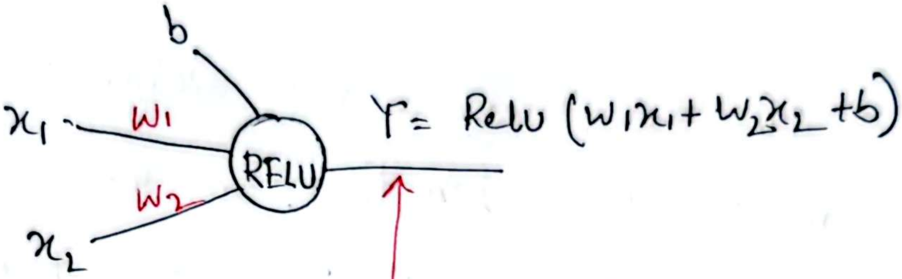
Parameter update with

RELU function

Let's say we have a simple neural networ (perceptron) like below →



Predicted target

The loss function,

$$\text{Loss} = \frac{1}{2}(\text{Ob} - Y)^2$$

actual target

To train this neural network model, we have to continuously updater the parameters weights and biases. for this networ it is w_1, w_2 and b.

We know updates take place like →

~~(old) w_1 →~~

$$(\text{new})\, w_1 = (\text{old})\, w_1 - \eta \frac{d\,\text{Loss}}{d\,w_1}$$

$$(\text{new})\, w_2 = (\text{old})\, w_2 - \eta \frac{d\,\text{Loss}}{d\,w_2}$$

$$(\text{new})\, b = (\text{old})\, b - \eta \frac{d\,\text{Loss}}{d\,w_2}$$

we need to find this differentiations.

so, we are differentiating Loss function $Loss = \frac{(Ob-Y)^2}{2}$ interms of w_1 and w_2, $\frac{dLoss}{dw_1}$ and $\frac{dLoss}{dw_2}$.

we can see there is no weight (w_1 or w_2) or bias (b) term inside the loss function. so, we cannot directly differentiate it. But the Loss function has Y inside it, ~~which~~ is $Y = Relu(x_1w_1 + x_2w_2 + b)$, which has w_1, w_2 and b inside it. So, we can differentiate Y interms of weights and biases. so, we get a chain rule for $\frac{dLoss}{dw_1}$ ~~and~~, $\frac{dLoss}{dw_2}$ and $\frac{dLoss}{db}$

$$\frac{dLoss}{dw_1} = \frac{dLoss}{dY} \times \frac{dY}{dw_1}$$

→ differentiate Y interms of w_1

↓ differentiate Loss function interm of Y

and $$\frac{dLoss}{dw_2} = \frac{dLoss}{dY} \times \frac{dY}{dw_2}$$

and $$\frac{dLoss}{db} = \frac{dLoss}{dY} \times \frac{dY}{db}$$

Now, $\dfrac{d\,Loss}{dY} = \dfrac{d}{dY}\left\{\frac{1}{2}(Ob-Y)^2\right\}$

$= \frac{1}{\cancel{2}} \times \cancel{2} \times (Ob-Y) \times (-1)$

$= -(Ob-Y)$

$= -Err$

Now, $\dfrac{dY}{dw_1} = \dfrac{d}{dw_1}\left[Relu\,(w_1x_1+w_2x_2+b)\right]$

$= Relu'(w_1x_1+w_2x_2+b) \times x_1$

differentiation of Relu function

$\therefore \boxed{\dfrac{dLoss}{dw_1}} = \dfrac{dLoss}{dY} \times \dfrac{dY}{dw_1}$

$= \underbrace{-Err \times Relu'(w_1x_1 w_2x_2+b)}_{\text{Let us consider it as } \delta}\,x_1$

$= -\delta x_1$

$\therefore w_1 = w_1 - \eta \dfrac{dLoss}{dw_1}$

$= w_1 - \eta(-\delta x_1)$

$= w_1 + \eta\,\delta x_1$

Now for $w_2 \rightarrow$

$$\frac{d\,Loss}{dw_2} = \frac{d\,Loss}{dY} \times \frac{dY}{dw_2}$$

we already derived this and value is $(-Err)$

$$\therefore \frac{dY}{dw_2} = \frac{d}{dw_2} Relu(w_1x_1 + w_2x_2 + b)$$

$$= Relu'(w_1x_1 + w_2x_2 + b)\, x_2$$

$$\therefore \frac{d\,Loss}{dw_2} = -\underbrace{Err \times Relu'(w_1x_1 + w_2x_2 + b)}_{\text{we assumed it } \delta}\, x_2$$

$$= - \, x_2 \cdot \delta$$

$$\therefore w_2 = w_2 - \eta \frac{d\,Loss}{dw_2}$$

$$= w_2 - \eta \, (-x_2 \delta)$$

$$= w_2 + \eta \, x_2 \delta$$

Similarly for b,

$$\frac{d\,Loss}{db} = \frac{d\,Loss}{dY} \times \frac{dY}{db}$$

we already derived this, $(-Err)$

$$\therefore \frac{dY}{db} = \frac{d}{db} \mathrm{Relu}(w_1x_1 + w_2x_2 + b)$$

$$= \mathrm{Relu}'(w_1x_1 + w_2x_2 + b) \times 1$$

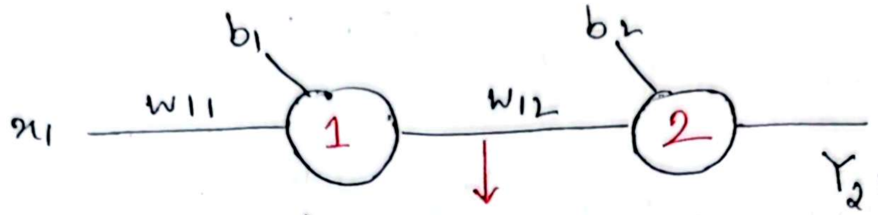
$$\therefore \frac{dLoss}{db} = \underbrace{-Err \times \mathrm{Relu}'(w_1x_1 + w_2x_2 + b)}_{\text{this is } \delta} = -\delta$$

$$\therefore b = b - \eta \frac{dLoss}{db}$$

$$= b - \eta(-\delta)$$

$$= b + \eta\delta$$

Now, Let us have another neural network →



$Y_2 = (w_{12} Y_1 + b_2)$

$Y_1 = \text{Relu}(w_{11} x_1 + b_1)$

here, weights are w_{11} and w_{12} and biases are b_1 and b_2. Here $\text{Loss} = \frac{1}{2}(Ob_2 - Y_2)^2$

Let us update w_{12} and b_2 first.

$$w_{12} = w_{12} - \eta \frac{d\,\text{Loss}}{d w_{12}}$$

$$\frac{d\,\text{Loss}}{d w_{12}} = \frac{d\,\text{Loss}}{d Y_2} \times \frac{d Y_2}{d w_{12}}$$

Loss has Y_2 Y_2 has w_{12}

$$\frac{d\,\text{Loss}}{d Y_2} = \frac{d}{d Y_2}\left\{\frac{1}{2}(Ob_2 - Y_2)^2\right\}$$

$$= \frac{1}{2} \times 2(Ob_2 - Y_2) \times (-1)$$

$$= -(Ob_2 - Y_2)$$

$$= -\text{Err}_2$$

and $\frac{dY_2}{dw_{12}} = \frac{d}{dw_{12}} \text{Relu}(w_{12}Y_1 + b_2)$

$= \text{Relu}'(w_{12}Y_1 + b_2) \times Y_1$

$\therefore w_{12} = w_{12} - \eta \frac{d\,Loss}{dw_{12}}$

$= w_{12} - \eta \left(-Err_2 \times Y_1 \underbrace{\text{Relu}'(w_{12}Y_1 + b_1)}_{\delta_2}\right)$

$= w_{12} + \eta \; Y_1 \delta_2$

Similarly for b_2,

$\frac{d\,LOSS}{db_2} = \frac{d\,Loss}{dY_2} \times \frac{dY_2}{db_2}$

We have this value, $-Err_2$

$= -Err_2 \times \frac{d}{db_2} \text{Relu}(w_{12}Y_1 + b_2)$

$= -Err_2 \times \text{Relu}'(w_{12}Y_1 + b_2)$

$= -\delta_2$

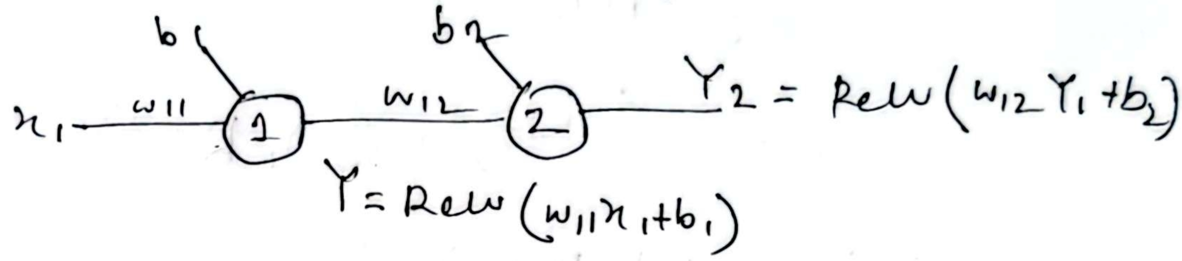
For b there will be no input multiplied like x or Y

$$\therefore b_2 = b_2 - \eta \frac{dLoss}{db_2}$$

$$= b_2 - \eta \cdot (-\delta)$$

$$= b_2 + \eta\delta$$

Now for the neuron 1 and weight w_{11} and bias b_1



So, $w_{11} = w_{11} - \eta \frac{dLoss}{dw_{11}}$.

To find $\frac{dLoss}{dw_{11}}$, we can see, Loss function $\frac{1}{2}(Ob_2 - Y_2)^2$ has Y_2. So, we can differentiate it with respect to Y_2. Then, Y_2 has Y_1. So, we can differentiate Y_2 in terms of Y_1. Then finally Y_1 has w_{11}. So we can differentiate Y_1 with respect to w_{11}.

$$\text{So, } \frac{dLoss}{dw_{11}} = \frac{dLoss}{dY_2} \cdot \frac{dY_2}{dw_{11}}$$

$$= \frac{dLoss}{dY_2} \cdot \boxed{\frac{dY_2}{dY_1} \cdot \frac{dY_1}{dw_{11}}}$$

We already know,

$$\frac{dLoss}{dY_2} = -Error_2$$

Now, $$\frac{dY_2}{dY_1} = \frac{d}{dY_1} ReLU(W_{12}Y_1 + b_2)$$
$$= ReLU'(W_{12}Y_1 + b_2)\, W_{12}$$

and $$\frac{dY_1}{dW_{11}} = \frac{d}{dW_{11}} ReLU(w_{11}x_1 + b_1)$$
$$= ReLU'(w_{11}x_1 + b_1)\, x_1$$

$$\therefore w_{11} = w_{11} - \eta\Big(\underbrace{-Error_2 \times ReLU'(W_{12}Y_1 + b_2)}_{\delta_2} W_{12}\Big) \times ReLU'(w_{11}x_1 + b_1)x_1$$
$$= w_{11} + \eta\, \delta_2\, W_{12}\, ReLU'(w_{11}x_1 + b_1)x_1$$

Similarly we will get,

$$b_1 = b_1 - \eta \frac{dLoss}{db_1}$$
$$= b_1 - \eta\left(\frac{dLoss}{db_1} \times \frac{dY_2}{dY_1} \times \frac{dY_1}{db_1}\right)$$
$$= b_1 - \eta\Big(\underbrace{-Error_2 \times ReLU'(w_{12}Y_1 + b_2)}_{\delta_2} w_{12}\Big) \times ReLU'(w_{11}x_1 + b_1)$$
$$= b_1 + \eta\, \delta_2\, W_{12}\, ReLU'(w_{11}x_1 + b_1)$$