Back-propogation

Concepts:

IQ	Level	Salary
3	3	50
8	55	1000
3	8	60
4	8	65
3	2	30
9	88	5000
3	2	35
8	6	40

Total data(rows) = 8

no of Batches = 4

$\therefore$ Batch size = $\frac{\text{Total data (rows)}}{\text{no of Batches}}$

$= \frac{8}{4}$

$= 2$

$\boxed{\text{Batchsize}}$ => new no of rows in each batch.

The $\boxed{\text{no of four batches}}$ with $\boxed{\text{batch size two}}$ visualization

Batch 1

IQ	level	salary
3	3	50
8	55	1000

Batch 2

IQ	level	salary
3	8	60
4	8	65

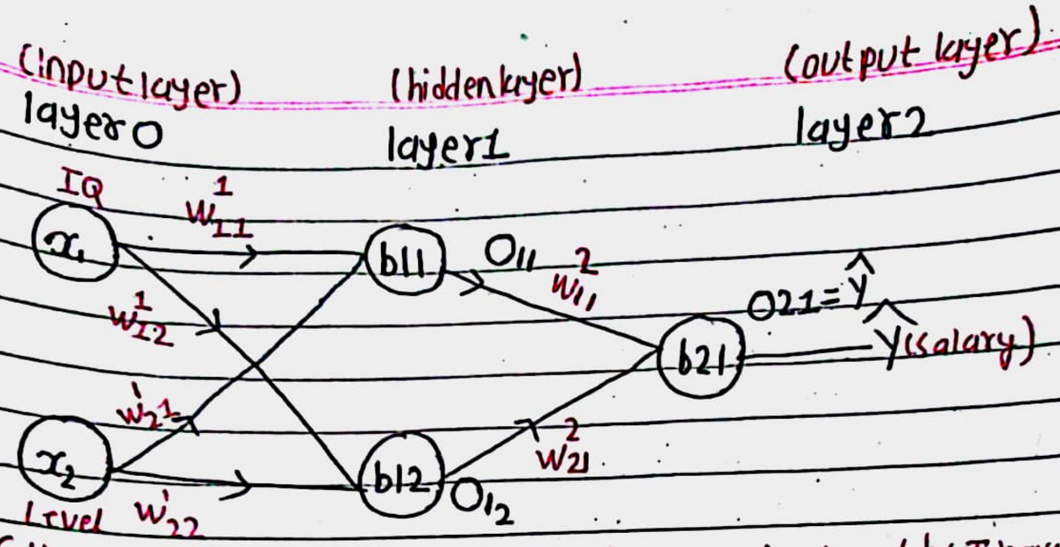
Batch3

IQ	level	salary
3	2	30
9	88	5000

Batch4

IQ	level	salary
3	2	35
8	6	40

$\boxed{\text{no of rows in each Batch} = \text{Batch size} = 2}$



(Hami mini-batch gradient descent bata bhuxam)

Yo harmo neural network ho. ABA hami harmo data set lai 4ota batch banayera 2 ota Batch size (row) feed garxam. ani suru ma parameter pani random assign garxam

At first:
Batch 1 ko data neural network ma feed hunxa

...	...	...	
x	x	+	row1
x	x	+	row2

(Batch1)

Suru ma Batch 1 ko row 1 ko data neural network ma Janxa ani $\hat{y}$ calculate hunxa tespaxi row 2 ko Janxa ani feri $\hat{y}$ calculate hunxa. (similarly batch ko sabai row feed hunxa ra respective $\hat{y}$ calculate hunxa). Aba sabai ko $\hat{y}$ calculate gare paxi loss calculate hunxa ra tyo loss minimize garna weights update hunxa gradient decent algo bata.
(we got our weights updated)

then, At second:
Aba hami ley feri Batch 2 ko data Neural network ma feed garxam tara hami sanga aba Batch 1 bata update vayeko weights and bias hunxa.

...	...	...	
x	x	+	row1
x	x	+	row2

(Batch 2)

Similarly, yes choti pani Batch 2 ko row 1 ra row 2 palai pilo feed hunxa ani $\hat{y}$ pani calculate hunxa. $\hat{y}$ calculate sabai row ko vayo aba loss feri calculate hunxa ra feri weights and bias update garinxa gradient decent algo bata.
(we got our weights and bias updated second time)

naya weights ko Neural Network ma

Similarly, aba feri batch 3 ko rows haru feed hunxa ra respective ŷ calculate garinxa. Ani feri loss calculate garinxa ra feri weights update garinxa.
(we got our weights updated 3 times)

feri Batch 4 ko rows naya aayeko weights ma rakhera Neural Network ma feed garinxa feri loss calculate garinxa ani feri weights update garinxa.
(we got our weights updated 4 times)

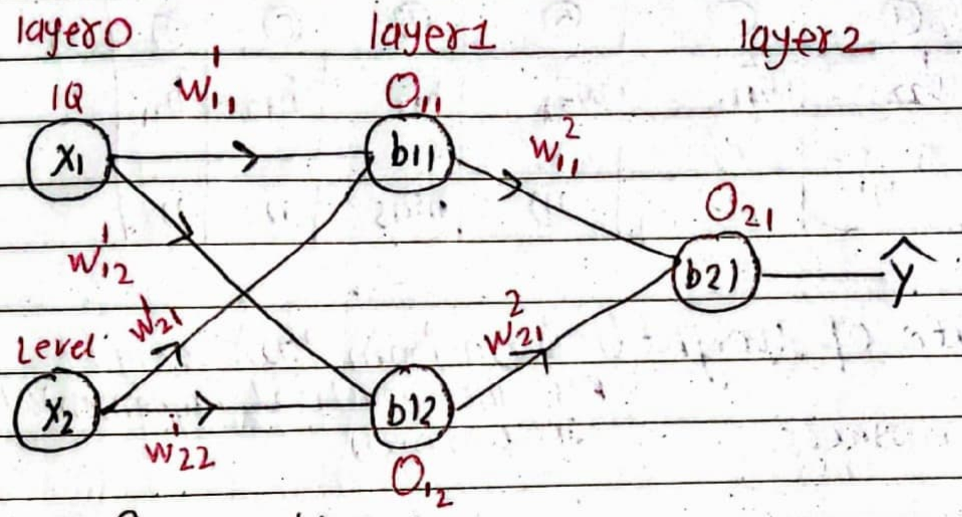
Harmo 4 otai Batch ko sabai data Neural Network ma feed garem our (1 epoch) completed.

Aba feri (2 epoch...... n epoch) samma feri sabai Batch ko data palai pilo feed garxa naya updated weights ma until we get the best weights and bais having very less loss.

2 epoch ma ne same agi ko step repet hunxa
Jastai:-
agi hamiley Batch 4 bata New updated weights and bias pako thimem aba yo choti feri fei weights rakhera Batch 1 train garinxa, loss calculate hunxa, weights update garinxa
feri Batch 2 ko turn aauxa feri yesma ne Bhakar ko new weights ma train garinxa loss calculate hunxa, weights update garinxa
This goes on... naya weights ma train garai rakhinxa unless convergence

Now let's see how back propogation works

Suppose harmo 1^{st} Batch ko sabai row feed vayera loss calculate vayo aba hami parameters update garram let's see down how does parameter update take places:-



weights = ?　　bias = ?

layer 1 (kolagi)
weights = 2x2 = 4
bias = 2
Total = 6

layer 2 (kolagi)
weights = 2x1 = 2
bias = 1
Total = 3

∴ We have to update total (6+3) = 9 parameters
let's see how we do it in back propagation.

Our $\hat{y}$ is given by :-

$$\boxed{O_{21} = \hat{y} = O_{11} \cdot w^2_{11} + O_{12} + w^2_{21} + b_{21}}$$

let's suppose we use linear activation to all the nodes.

The loss is given by :- $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

In our batch we have 2 row. we pass two row in Neural Network, calculate $\hat{y}$ the loss so, for our case loss is given by :-

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

1	2	3	4	5	6	7	8	9
w_{11}^1	w_{12}^1	w_{21}^1	w_{22}^1	w_{11}^2	w_{21}^2	b_{11}	b_{12}	b_{21}
weight	"	"	"	"	"	bias	"	"

let's see first update of weight (let's go from last weight as like practical implementation in libraries)

suppose α be '1' for easyness

6 w_{21}^2

$$= w^2_{21\,new} = w^2_{21\,old} - \frac{d\,loss\,f}{d w_{21}^2}$$

If we change w_{21}^2 It effects (O_{21}) $\hat{y}$ which further effects loss.

$\therefore\ w_{21}^2 \rightarrow O_{21}(\hat{y}) \rightarrow loss$

$\Rightarrow w^2_{21\,new} = w^2_{21\,old} - A.$ (let) — ean(i)

$$\Rightarrow A = \frac{d\,loss\,f}{d w_{21}^2}$$

$$w_{21}^2 \longrightarrow O_{21} \longrightarrow loss$$

$$\Rightarrow \frac{d\,loss}{dw_{21}^2} = \frac{d\,loss}{dO_{21}} \times \frac{dO_{21}}{dw_{21}^2}$$

$$= \frac{d\,loss}{dO_{21}} \times \frac{dO_{21}}{dw_{21}^2}$$

$$= \sum_{i=1}^{2} \frac{d(y_i - \hat{y}_i)^2}{dO_{21}} \times \frac{dO_{21}}{dw_{21}^2}$$

$$= \sum_{i=1}^{2} \left(\frac{d(y_i - \hat{y}_i)^2}{d(y_i - \hat{y}_i)} \times \frac{d(y_i - \hat{y}_i)}{dO_{21}} \right) \times \frac{dO_{21}}{dw_{21}^2}$$

$$\boxed{\hat{y} = O_{21}}$$

$$= \sum_{i=1}^{2} \left(2(y_i - \hat{y}_i) \times \left(\frac{dy_i}{dO_{21}} - \frac{dO_{21}}{dO_{21}} \right) \right) \times \frac{dO_{21}}{dw_{21}^2}$$

$$= \sum_{i=1}^{2} (2(y_i - \hat{y}_i) \times (0 - 1) \times \frac{dO_{21}}{dw_{21}^2}$$

$$= \sum_{i=1}^{2} (-2(y_i - \hat{y}_i)) \times \frac{dO_{21}}{dw_{21}^2}$$

$$= \sum_{i=1}^{2} (2(\hat{y}_i - y_i)) \times \frac{d(O_{11} \cdot w_{11}^2 + O_{12} \cdot w_{21}^2 + b_{21})}{d\ dw_{21}^2}$$

$$= \sum_{i=1}^{2} (2(\hat{y}_i - y_i) \times (0 + O_{12} + 0)$$

$\therefore$ $A = \sum_{i=1}^{2} (2(\hat{y}_i - y_i)) \times O_{12}$ (putting in ean i)

$$\therefore\ \boxed{\begin{aligned} w_{21\,new}^2 &= w_{21\,old}^2 - \left\{ 2\sum_{i=1}^{2} ((\hat{y}_i - y_i)) \times O_{12} \right\} \\ &= w_{21\,old}^2 - \left\{ 2((\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)) \times O_{12} \right\} \end{aligned}}$$

Note from above derivation

$$\boxed{\frac{d\,loss}{dO_{21}} = 2\sum_{i=1}^{2}(\hat{y}_i - y_i)} \text{ or } 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - \hat{y}_2)\}^2$$

Now, let's do for w^2_{11}

⑤ w^2_{11}

$$\Rightarrow w^2_{11new} = w^2_{11old} - \frac{d\,loss}{dw^2_{11}}$$

$$\Rightarrow w^2_{11\,new} = w^2_{11}old - A$$

$[w^2_{11} \rightarrow O_{21} \rightarrow loss]$

$$\therefore A = \frac{d\,loss}{dw^2_{11}} = \frac{d\,loss}{dO_{21}} \times \frac{dO_{21}}{dw^2_{11}}$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - \hat{y}_2)\} \times \frac{[O_{11} \cdot w^2_{11} + O_{12} \cdot w^2_{21} + b_{21}]}{dw^2_{11}}$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - \hat{y}_2)\} \times [O_{11} + 0 + 0]$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - \hat{y}_2)\} \times O_{11}$$

$$\boxed{\therefore w^2_{11new} = w^2_{11}old - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times O_{11}\right]}$$

Now, let's do for w^1_{22} ...

(4) w^1_{22}

$$\Rightarrow w^1_{22new} = w^1_{22old} - \frac{d\,loss}{dw^1_{22}}$$

$$\Rightarrow w^1_{22new} = w^1_{22old} - A$$

$$[w^1_{22} \rightarrow O_{12} \rightarrow O_{21} \rightarrow LOSS]$$

$$\therefore A = \frac{d\,loss}{dw^1_{22}} = \frac{d\,loss}{dO_{21}} \times \frac{dO_{21}}{dO_{12}} \times \frac{dO_{12}}{dw^1_{22}}$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times \frac{d[O_{11} \cdot w^2_{11} + O_{12} \cdot w^2_{21} + b_{21}]}{dO_{12}} \times \frac{[x_1 \cdot w^1_{12} + x_2 \cdot w^1_{22} + b_{12}]}{dw^1_{22}}$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times [0 + w^2_{21} + 0] \times [0 + x_2 + 0]$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{21} \times x_2$$

$$\boxed{\therefore w^1_{22new} = w^1_{22old} - [2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{21} \times x_2]}$$

Now, similary for other weights (w^1_{21}, w^1_{12}, w^1_{11})

$$\boxed{\therefore w^1_{21_{new}} = w^1_{21_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{11} \times x_2\right]}$$

$$\boxed{\therefore w^1_{12_{new}} = w^1_{12_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{21} \times x_1\right]}$$

$$\boxed{\therefore w^1_{11_{new}} = w^1_{11_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{11} \times x_1\right]}$$

Now, let's see the update of bias.

⑨ b_{21}

$$\Rightarrow b_{21_{new}} = b_{21_{old}} - \frac{d\,loss}{d\,b_{21}}$$

$$\Rightarrow b_{21_{new}} = b_{21_{old}} - A \text{ (let)}$$

$$\therefore A = \frac{d\,loss}{d\,b_{21}}$$

$$= [b_{21} \rightarrow O_{21} \rightarrow loss]$$

$$\Rightarrow \frac{d\,loss}{d\,b_{21}} = \frac{d\,loss}{d\,O_{21}} \times \frac{d\,O_{21}}{d\,b_{21}}$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times \frac{d[O_{11} \cdot w^2_{11} + O_{12} \cdot w^2_{21} + b_{21}]}{d\,b_{21}}$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times [0 + 0 + 1]$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\}$$

$$\boxed{\therefore\ b_{21_{new}} = b_{21_{old}} - [2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\}]}$$

Now, let's see the update of another bias

⑧ b_{12}

$$\Rightarrow b_{12_{new}} = b_{12_{old}} - \frac{d\,loss}{d\,b_{12}}$$

$$\Rightarrow b_{12_{new}} = b_{12_{old}} - A \text{ (Let)}$$

$$[b_{12} \rightarrow O_{12} \rightarrow O_{21} \rightarrow loss]$$

$$\therefore A = \frac{d\,loss}{d\,b_{12}} = \frac{d\,loss}{d\,O_{21}} \times \frac{d\,O_{21}}{d\,O_{12}} \times \frac{d\,O_{12}}{d\,b_{12}}$$

$$= 2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times \frac{d[O_{11} \cdot w^2_{11} + O_{12} \cdot w^2_{21} + b_{21}]}{d\,O_{12}} \times$$

$$\frac{d[x_1 \cdot w^1_{12} + x_2 \cdot w^1_{22} + b_{12}]}{d\,b_{12}}$$

$$= 2\{(\hat{Y}_1 - Y_1) + (\hat{Y}_2 - Y_2)\} \times w^2_{21} \times 1$$

$$\boxed{\therefore\ b_{12_{new}} = b_{12_{old}} - \left[2\{(\hat{Y}_1 - Y_1) + (\hat{Y}_2 - Y_2)\} \times w^2_{21}\right]}$$

Now, let's do update of bias b_{11}

(7) b_{11}

$$\Rightarrow b_{11_{new}} = b_{11_{old}} - \frac{d\,loss}{d\,b_{11}}$$

$$\Rightarrow b_{11_{new}} = b_{11_{old}} - A$$

$$[b_{11} \rightarrow O_{11} \rightarrow O_{21} \rightarrow loss]$$

$$\therefore A = \frac{d\,loss}{d\,b_{11}}$$

$$\Rightarrow \frac{d\,loss}{d\,b_{11}} = \frac{d\,loss}{d\,O_{21}} \times \frac{d\,O_{21}}{d\,O_{11}} \times \frac{d\,O_{11}}{d\,b_{11}}$$

$$= 2\{(\hat{Y}_1 - Y_1) + (\hat{Y}_2 - Y_2)\} \times \frac{d[O_{11} \cdot w^2_{11} + O_{12} \cdot w^2_{21} + b_{21}]}{d\,O_{11}} \times$$

$$\frac{d[X_1 \cdot w^1_{11} + X_2 \cdot w^1_{21} + b_{11}]}{d\,b_{11}}$$

$$= 2\{(\hat{Y}_1 - Y_1) + (\hat{Y}_2 - Y_2)\} \times w^2_{11} \times 1$$

$$\boxed{\therefore\ b_{11_{new}} = b_{11_{old}} - \left[2\{(\hat{Y}_1 - Y_1) + (\hat{Y}_2 - Y_2)\} \times w^2_{11}\right]}$$

Conclusion:

weights:

$$\boxed{w^1_{11_{new}} = w^1_{11_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{11} \times x_1\right]}$$

$$\boxed{w^1_{12_{new}} = w^1_{12_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{21} \times x_1\right]}$$

$$\boxed{w^1_{21_{new}} = w^1_{21_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{11} \times x_2\right]}$$

$$\boxed{w^1_{22_{new}} = w^1_{22_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{21} \times x_2\right]}$$

$$\boxed{w^2_{11_{new}} = w^2_{11_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times O_{11}\right]}$$

$$\boxed{w^2_{21_{new}} = w^2_{21_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times O_{12}\right]}$$

bias:

$$\boxed{b_{11_{new}} = b_{11_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{11}\right]}$$

$$\boxed{b_{12_{new}} = b_{12_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\} \times w^2_{21}\right]}$$

$$\boxed{b_{21_{new}} = b_{21_{old}} - \left[2\{(\hat{y}_1 - y_1) + (\hat{y}_2 - y_2)\}\right]}$$