3. The output of the sigmoid function directly represent the probability of the positive class, making it intuitive to interpret the results.

Backpropagation

(The What)

Backpropagation

↓

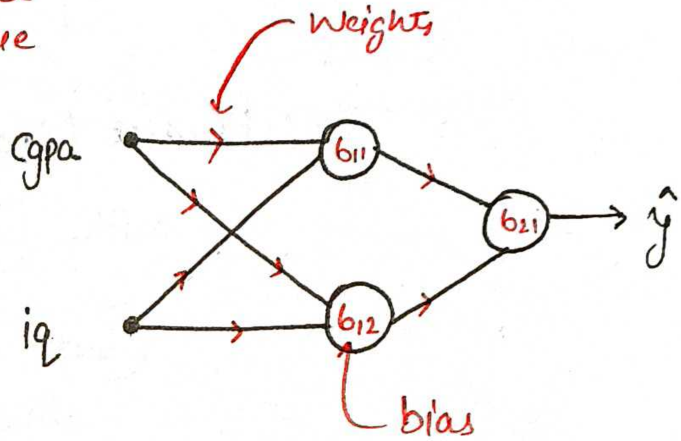
Algo → Train nn

↓

Training — weights and bias $\xrightarrow{\text{find}}$ Correct value

according to this data

cgpa	iq	lpa
8	80	8
7	70	7

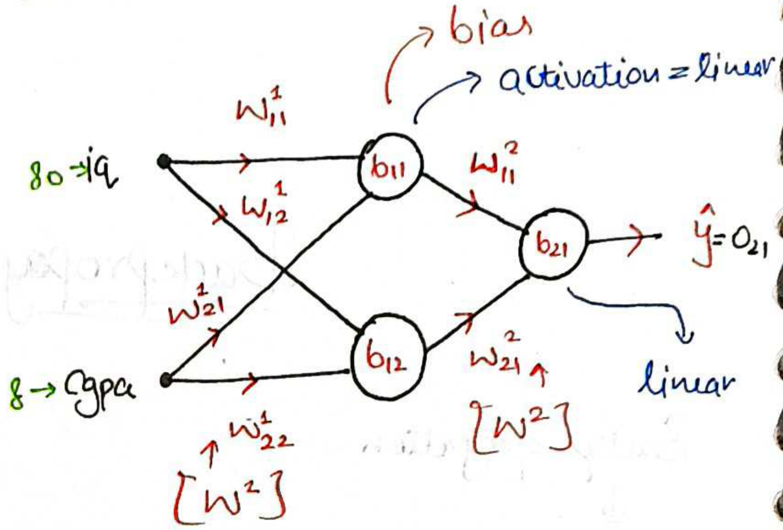


Backpropagation is an algorithm used in neural network to adjust the weights of the network by minimizing the error between the predicted and actual outputs. It does this by propagating the error backward through the network.

Backpropagation

→ 4 students

iq	cgpa	lpa
80	8	3
60	9	5
70	5	8
120	7	11



Steps →

0) initialize W, b → (W, b) Random Value
→ $\boxed{W \to 1, \; b \to 0}$ ↳ we are using this method

1) You select a point (row)
↳ student

2) Predict (lpa) → forward prop (Dot product)
with $\boxed{W \to 1, \; b \to 0}$ → ($\hat{y}$) Output : 18 Lpa
[but original y is 3 Lpa] ←

3) Choose a loss function

$MSE = (y - \hat{y})^2 \rightarrow (3-18)^2 = 225$ → error
(L)

↳ we have to reduce error → cannot change y value bcz it is true

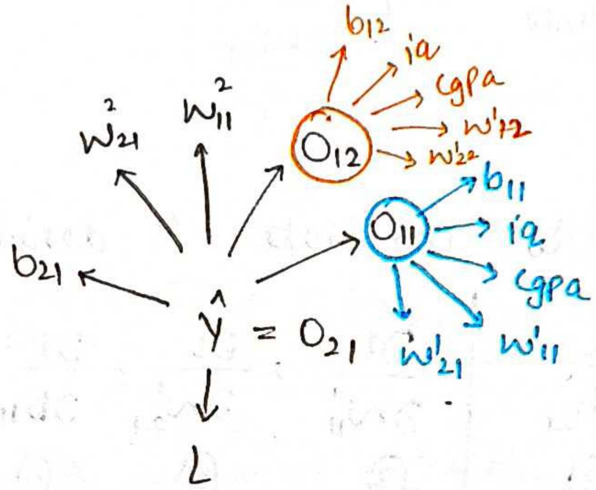
but we can change $\hat{y}$ value

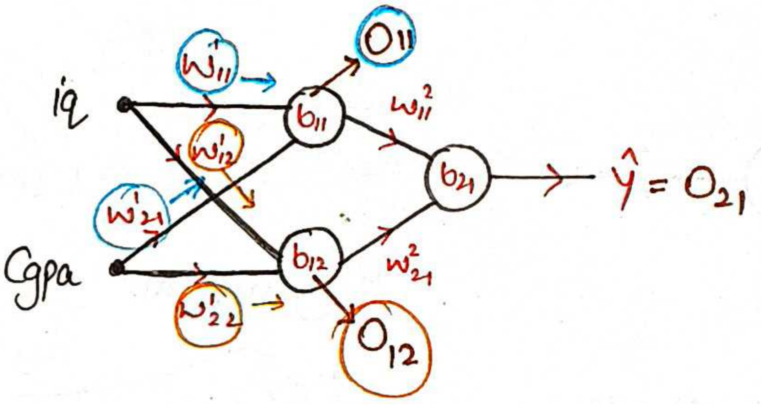
In our case, we have to decrease $\hat{y}$ value
but sometime we have to increase $\hat{y}$ value for low error.

basically, O_{21} is the final output.

So, $L(\text{Min}) \xrightarrow{\text{depend}} \hat{y} \xrightarrow{\text{is}} O_{21}$ and O_{21} depend on 5 value.

$$O_{21} = w^2_{11} O_{11} + w^2_{21} O_{12} + b_{21}$$





basically, we have to change the value of w'_{11}, w'_{12}, w'_{21} and w'_{22} → that's why this algo is called backpropagation

4) Weights and bias update

↳ Gradient Descent

$$\boxed{W_{new} = W_{old} - \eta \frac{dL}{\partial W_{old}}}$$

for weight

$$\boxed{b_{new} = b_{old} - \eta \frac{dL}{\partial b_{old}}}$$

for bias

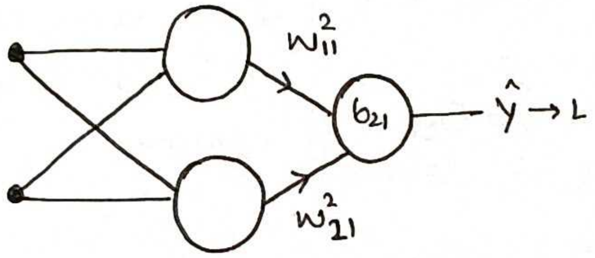
$$\boxed{W^2_{11\,new} = W^2_{11\,old} - \eta \boxed{\frac{\partial L}{\partial W^2_{11}}}}$$

$$\boxed{W^2_{21\,new} = W^2_{21\,old} - \eta \boxed{\frac{\partial L}{\partial W^2_{21}}}}$$

derivative of Loss fn w.r.t. weight $\frac{\partial L}{\partial W^2_{11}}$ and $\frac{\partial L}{\partial W^2_{21}}$

Now, we have to calculate 9 derivatives

$$\underset{①}{\frac{dL}{\partial W^2_{11}}}, \underset{②}{\frac{\partial L}{\partial W^2_{21}}}, \underset{③}{\frac{\partial L}{\partial b_{21}}} \;\Big|\; \underset{④}{\frac{\partial L}{\partial W^1_{11}}}, \underset{⑤}{\frac{\partial L}{\partial W^1_{21}}}, \underset{⑥}{\frac{\partial L}{\partial b_{11}}} \;\Big|\; \underset{⑦}{\frac{\partial L}{\partial W^1_{12}}}, \underset{⑧}{\frac{\partial L}{\partial W^1_{22}}}, \underset{⑨}{\frac{\partial L}{\partial b_{12}}}$$



change in W^2_{11}

$\hat{y}$ change

$\hat{y}$

W^2_{11}

Not directly change

Loss change

L (Min)

Why we do derivative ?

↳ means

$\left\{\frac{dY}{dx}\right\}$ when small change in X then how much change in Y.

And We cannot find directly derivative of $\frac{\partial L}{\partial W^2_{11}}$. first we have to find $\frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial W^2_{11}}$ derivative.

①

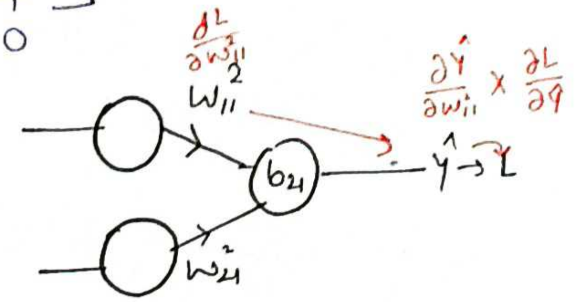
$$\boxed{\frac{\partial L}{\partial W^2_{11}} = \underbrace{\boxed{\frac{\partial L}{\partial \hat{y}}} \times \boxed{\frac{\partial \hat{y}}{\partial W^2_{11}}}}_{\text{break down for}}}$$

This method is called Chain rule of diffrentiation

$$\left[\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}} (y - \hat{y})^2 = -2 (y - \hat{y})\right]$$

$$\frac{\partial \hat{y}}{\partial w_{11}^2} = \frac{\partial}{\partial w_{11}^2}\left[\underset{\downarrow\, 1}{O_{11} w_{11}^2} + \underbrace{O_{12} w_{21}^2}_{0} + \underbrace{b_{21}}_{0}\right] = O_{11}$$

$$\boxed{\frac{\partial L}{\partial w_{11}^2} = -2(y-\hat{y})\,O_{11}} \longrightarrow$$



(2)

$$\begin{array}{c} \hat{y} \longrightarrow w_{21}^2 \\ \uparrow \\ L \end{array} \qquad \frac{\partial L}{\partial w_{21}^2} \longrightarrow \boxed{\frac{\partial L}{\partial w_{21}^2} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_{21}^2}} \Rightarrow$$ Chain rule of differentiation

$$\frac{\partial \hat{y}}{\partial w_{21}^2} = \frac{\partial}{\partial w_{21}^2}\left[O_{11} w_{11}^2 + O_{12} w_{21}^2 + b_{21}\right] = O_{12}$$

$$\boxed{\frac{\partial L}{d\hat{y}} = -2(y-\hat{y})}$$

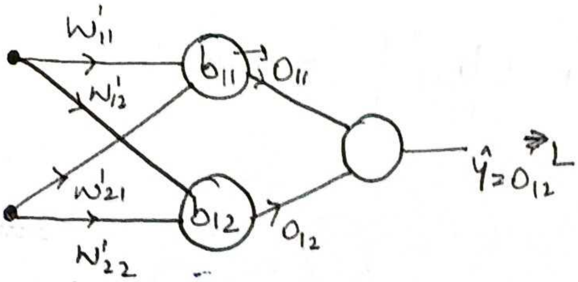
$$\boxed{\frac{\partial L}{\partial w_{21}^2} = -2(y-\hat{y})\,O_{12}}$$

(3)

$$\begin{array}{c} \hat{y} \longrightarrow b \\ \uparrow \\ L \end{array} \qquad \longrightarrow \qquad \frac{\partial L}{\partial b_{21}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial b_{21}}$$ ⇒ Chain rule of differen-

$$\frac{\partial L}{\partial \hat{y}} \to -2(y-\hat{y}) \qquad \frac{\partial \hat{y}}{\partial b_{21}} \to 1$$

$$\boxed{\frac{\partial L}{\partial b_{21}} = -2(y-\hat{y})}$$

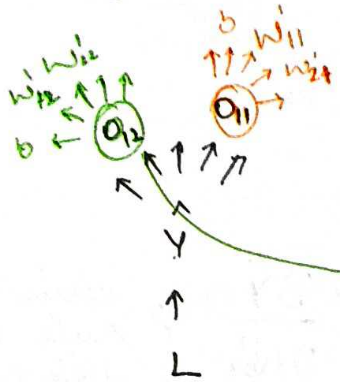


Derivative eq^n I

$$\frac{\partial L}{\partial w'_{11}} = \boxed{\frac{\partial L}{\partial \hat{y}} \; \frac{\partial \hat{y}}{\partial O_{11}}} \; \frac{\partial O_{11}}{\partial w'_{11}}$$

$$\frac{\partial L}{\partial w'_{21}} = \boxed{\frac{\partial L}{\partial \hat{y}} \; \frac{\partial \hat{y}}{\partial O_{11}}} \; \frac{\partial O_{11}}{\partial b_{11}}$$

$$\frac{\partial L}{\partial b_{11}} = \boxed{\frac{\partial L}{\partial \hat{y}} \; \frac{\partial \hat{y}}{\partial O_{11}}} \; \frac{\partial O_{11}}{\partial b_{11}}$$



$$\frac{\partial L}{\partial w'_{12}} = \boxed{\frac{\partial L}{\partial \hat{y}} \; \frac{\partial \hat{y}}{\partial O_{12}}} \; \frac{\partial O_{12}}{\partial w'_{12}}$$

$$\frac{\partial L}{\partial w'_{22}} = \boxed{\frac{\partial L}{\partial \hat{y}} \; \frac{\partial \hat{y}}{\partial O_{12}}} \; \frac{\partial O_{12}}{\partial w'_{22}}$$

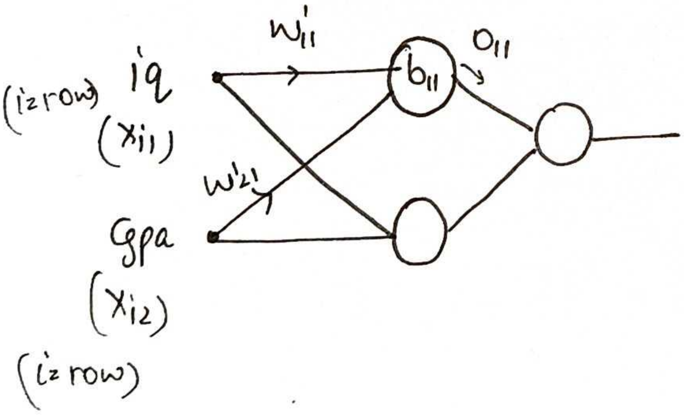
$$\frac{\partial L}{\partial b_{12}} = \boxed{\frac{\partial L}{\partial \hat{y}} \; \frac{\partial \hat{y}}{\partial O_{12}}} \; \frac{\partial O_{12}}{\partial b_{12}}$$

Finding Value of eq^n

$$\boxed{\frac{\partial L}{\partial \hat{y}} = -2(y-\hat{y})}$$

$$\boxed{\frac{\partial \hat{y}}{\partial O_{11}} = \frac{\partial}{\partial O_{11}}\left[w^2_{11} O_{11} + w^2_{21} O_{21} + b_{21}\right] = w^2_{11}}$$

$$\boxed{\frac{\partial \hat{y}}{\partial O_{12}} = \frac{\partial}{\partial O_{12}}\left[w^2_{11} O_{11} + w^2_{21} O_{11} + b_{21}\right] = w^2_{21}}$$



$$\frac{\partial O_{11}}{\partial w'_{11}} = \frac{\partial}{\partial w'_{11}}\left[iq\, w'_{11} + gpa\, w'_{21} + b_{11}\right]$$

$$= iq \; (x_{i1})$$

$$\frac{\partial O_{11}}{\partial w'_{21}} = Cgpa \rightarrow x_{i2}$$

$$\frac{\partial O_{12}}{\partial w'_{12}} = \frac{\partial}{\partial w'_{12}}\left[iq\, w'_{12} + Cgpa\, w'_{22} + b_{12}\right] = iq \rightarrow x_{i1}$$

$$\frac{\partial O_{11}}{\partial b_{11}} = 1$$

$$\frac{\partial O_{12}}{\partial w'_{22}} = x_{i2} \rightarrow Cgpa$$

$$\frac{\partial O_{2}}{\partial b_{12}} = 1$$

Assign value in eqⁿ (Answer)

$$\frac{\partial L}{\partial w'_{11}} = -2(y-\hat{y})\, w^2_{11}\, x_{i1} \rightarrow iq$$

$$\frac{\partial L}{\partial w'_{21}} = -2(y-\hat{y})\, w^2_{11}\, x_{i2} \rightarrow Cgpa$$

$$\frac{\partial L}{\partial b_{11}} = \frac{\partial L}{\partial \hat{y}}\,\frac{\partial \hat{y}}{\partial O_{11}}\,\frac{\partial O_{11}}{\partial b_{11}} = -2(y-\hat{y})\, w^2_{11}$$

$$\frac{\partial L}{\partial w'_{12}} = -2(y-\hat{y})\, w^2_{21}\, x_{i1} \rightarrow iq$$

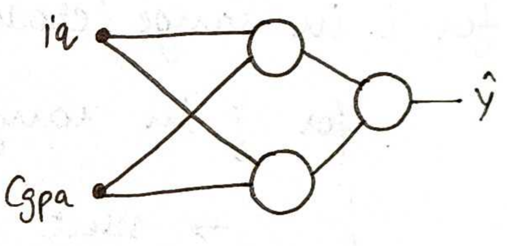
$$\frac{\partial L}{\partial w'_{22}} = -2(y-\hat{y})\, w^2_{21}\, x_{i2} \rightarrow Cgpa$$

$$\frac{\partial L}{\partial b_{12}} = -2(y-\hat{y})\, w^2_{21}$$

Steps (Once again)

4 Student data

Cgpa	iq	lpa
8	80	8
6	60	6
7	70	7
9	90	9



0) weight/bias → initialize
- random
- w=1, b=0

for loop in range(epochs): → Convergence (Loss Min)

1) for i in range(4): → 4 student

  1a) 1 student → forward propa → predict (lpa)

  1b) Loss calculate (MSE)

  1c) Adjust all weights and bias → Gradient Descent

Next epochs 1-1000

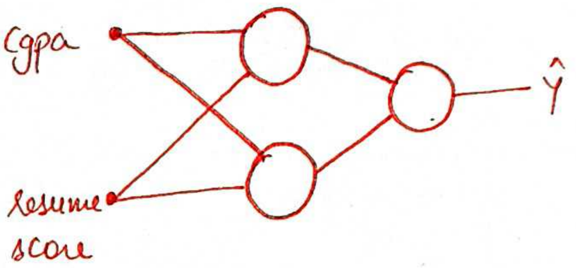
for next student row=2, 3, 4

Gradient Descent ← $\boxed{W_{new} = W_{old} - \eta \frac{\partial L}{\partial W_{old}}}$ → 9 times

Backpropagation → The How

Regression Problem

cgpa (X_{i1})	resume score (X_{i2})	package (lpa)
8	8	4
7	9	5
6	10	6
5	12	7



Backpropagation Algorithm

epochs = 5

```
for i in range (epochs):
    for j in range (X.shape[0]):
        → select 1 row (random)
        → Predict (using forward prop)
        → Calculate loss (using loss function → mse)
        → Update weights and bias using GD
```

$$W_n = W_0 - \eta \frac{\partial L}{\partial w}$$

↳ Calculate avg loss for the epoch

loss student1

loss of student 2

$$\frac{L_1 \; L_2 \; L_3 \; L_4}{4} = \text{Avg loss of single epoch}$$

$$\frac{\partial L}{\partial w^2_{11}} = -2(y-\hat{y})\, O_{11} \qquad \frac{\partial L}{\partial w^1_{11}} = -2(y-\hat{y})\, w^2_{11}\, x_{i1}$$

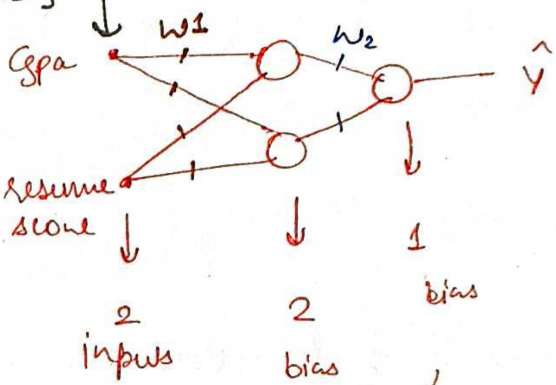
$$\frac{\partial L}{\partial w^2_{21}} = -2(y-\hat{y})\, O_{12} \qquad \frac{\partial L}{\partial b_{11}} = -2(y-\hat{y})\, w^2_{11}$$

$$\frac{\partial L}{\partial b_{21}} = -2(y-\hat{y}) \qquad \frac{\partial L}{\partial w^1_{12}} = -2(y-\hat{y})\, w^2_{21}\, x_{i1}$$

$$\frac{\partial L}{\partial w^1_{22}} = -2(y-\hat{y})\, w^2_{21}\, x_{i2} \qquad \frac{\partial L}{\partial b_{12}} = -2(y-\hat{y})\, w^2_{21}$$

Code: doubt → [layer – dims]

Architecture input →

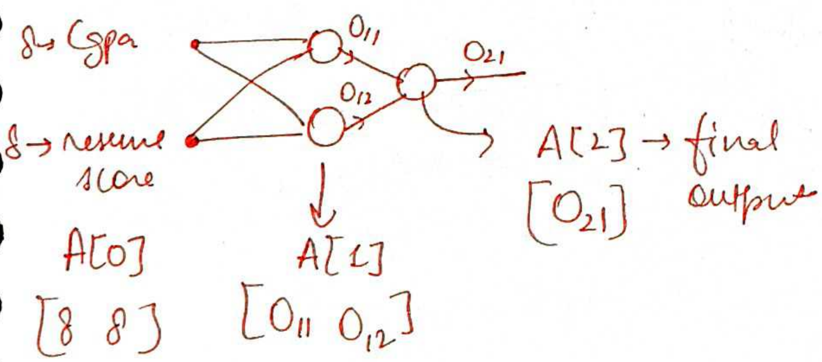


Architecture input is

[2 2 1]

{So it create total 9 weights and biases}

forward algo output in code



initialize parameter → code part → function

layer_dims = [2, 2, 1]

for l in range (1, L): → 2 times

$$\text{parameter['w'+ str(l)]} = \text{np.ones}(\underbrace{(\text{layer\_dims}[l-1], \text{layer\_dims}[l])}_{\text{2D array}}) * 0.1$$

first loop → W1 covered → $w^1_{11}, w^1_{12}, w^1_{21}, w^1_{22}$

Second loop → W2 covered → w^2_{21}, w^2_{22}

[assign value in W1 and W2]

parameter ['b'+ str(l)] = np.zeroes ((layer_dims[l-1], 1))

output:

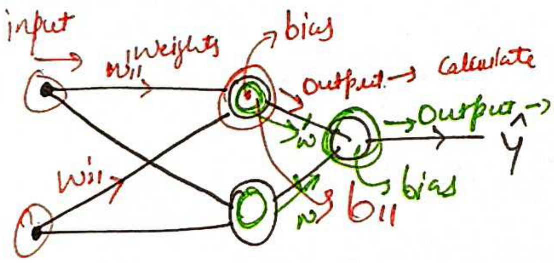
$$W1: \text{array}([[\overset{w^1_{11}}{0.1}, \overset{w^1_{12}}{0.1}], [\overset{w^1_{21}}{0.1}, \overset{w^1_{22}}{0.1}]]),$$

$$b1: \text{array}([[\overset{b_{11}}{0.}], [\overset{b_{12}}{0.}]]),$$

$$W2: \text{array}([[\overset{w^2_{11}}{0.1}], [\overset{w^2_{12}}{0.1}]]),$$

$$b2: \text{array}([\overset{b_{21}}{0.}])$$

linear forward → function



Calculate output of any given neuron

neuron depend on all coming inputs, weights and bias.

$b_{11} \xrightarrow{\text{depend}}$ input (2)
- → weights (w^1_{11}, w^1_{21}) (2)
- → bias (1)
- → Total → 5

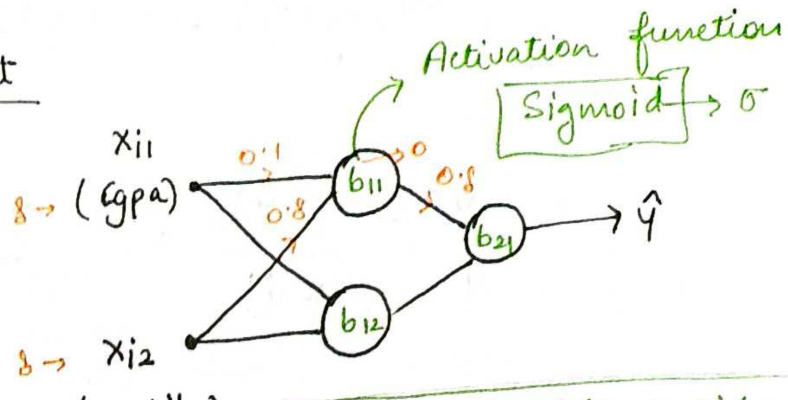
b_{21} → input → (b_{11} and b_{12} → output)
- → w^2_{11} and w^2_{12}
- → b_{21} → bias

← - - - - - -

- - - - - - →

Classification Example

cgpa	profile score	placement
8	8	1
7	9	1
6	10	0
5	5	0



$$\text{Loss} = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

→ binary cross entropy

Backpropagation Algorithm

epochs = 5.

for i in range (epochs):

for j in range (x.shape[0]):

→ Select 1 row (random)

→ Predict (using forward prop)

→ Calculate loss (using loss function → mse)

→ Update weights and bias using GD

$$W_n = W_o - \eta \frac{\partial L}{\partial W}$$

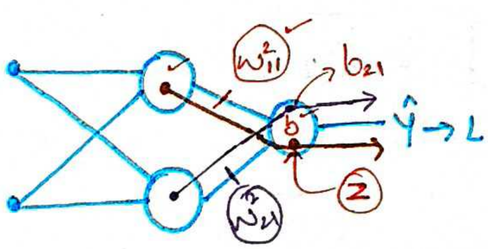
→ Calculate avg loss for the epoch

How to find output?

$$z = 0.1 \times 8 + 0.1 \times 8 + 0$$

$$\sigma(z) = O_{11} \rightarrow \text{output}$$

derivative w.r.t L



$$z_{final} = W^2_{11} O_{11} + W^2_{21} O_{12} + b_{21}$$

$$\hat{y} \rightarrow z_{final} \rightarrow W^2_{11},\ W^2_{21}$$

$$L \rightarrow \hat{y}$$

$$L = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

① ② common

$$\frac{\partial L}{\partial W^2_{11}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial W^2_{11}}$$

$$\frac{\partial L}{\partial W^2_{21}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial W^2_{21}}$$

$$\frac{\partial L}{\partial b_{21}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial b_{21}}$$

① $\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}\left[-y \log(\hat{y}) - (1-y)\log(1-\hat{y})\right]$

$= \frac{-y}{\hat{y}} - \frac{(1-y)}{(1-\hat{y})} = \frac{-y(1-\hat{y}) + \hat{y}(1-y)}{\hat{y}(1-\hat{y})} = \frac{-y + \cancel{y\hat{y}} + \hat{y} - \cancel{y\hat{y}}}{\hat{y}(1-\hat{y})}$

$$\boxed{\frac{\partial L}{\partial \hat{y}} = \frac{-(y-\hat{y})}{\hat{y}(1-\hat{y})}}$$

② $\frac{\partial \hat{y}}{\partial z} = \frac{\partial(\sigma(z))}{\partial z} = \sigma(z)[1-\sigma(z)]$ $\quad (\because \hat{y} = \sigma(z))$

$$\boxed{\frac{\partial \hat{y}}{\partial z} = \hat{y}[1-\hat{y}]}$$

① × ②

$$\boxed{\frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} = \frac{-(y-\hat{y})}{\cancel{\hat{y}(1-\hat{y})}} \times \cancel{\hat{y}(1-\hat{y})} = -(y-\hat{y})}$$

$$\left[\frac{\partial z}{\partial w_{11}^2} = \frac{\partial}{\partial w_{11}^2}\left[\underset{\downarrow\, 1}{w_{11}^2 O_{11}} + \underset{\downarrow\, 0}{w_{21}^2 O_{12}} + \underset{\downarrow\, 0}{b_{21}}\right] = O_{11}\right]$$

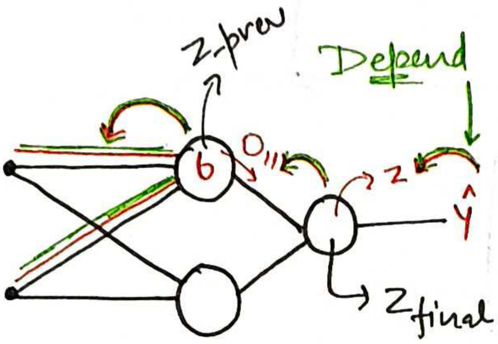
$$\left[\frac{\partial z}{\partial w_{21}^2} = \frac{\partial}{\partial w_{21}^2}\left[\underset{\downarrow\, 0}{w_{11}^2 O_{11}} + \underset{\downarrow\, 1}{w_{21}^2 O_{12}} + \underset{\downarrow\, 0}{b_{21}}\right] = O_{12}\right]$$

$$\left[\frac{\partial z}{\partial b_{21}} = \frac{\partial}{\partial b_{21}}\left[\underset{\downarrow\, 0}{w_{11}^2 O_{11}} + \underset{\downarrow\, 0}{w_{21}^2 O_{12}} + \underset{\downarrow\, 1}{b_{21}}\right] = 1\right]$$

$$\frac{\partial L}{\partial w_{11}^2} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial w_{11}^2} = -(y-\hat{y})\, O_{11}$$

$$\frac{\partial L}{\partial w_{21}^2} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial w_{21}^2} = -(y-\hat{y})\, O_{12}$$

$$\frac{\partial L}{\partial b_{21}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z} \times \frac{\partial z}{\partial b_{21}} = -(y-\hat{y})$$



$$\underset{①}{\frac{\partial L}{\partial w_{11}^1}},\quad \underset{②}{\frac{\partial L}{\partial w_{21}^1}},\quad \underset{③}{\frac{\partial L}{\partial b_{11}}}$$

$-(y-\hat{y})$ already solved

① $$\frac{\partial L}{\partial w_{11}^1} = \boxed{\frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_f}} \times \underbrace{\frac{\partial z_f}{\partial O_{11}}}_{w_{11}^2} \times \underbrace{\frac{\partial O_{11}}{\partial z_{prew}}}_{O_{11}(1-O_{11})} \times \underbrace{\frac{\partial z_{prew}}{\partial w_{11}^1}}_{x_{11}}$$

$$L \longrightarrow \hat{y} \longrightarrow Z_{final} \longrightarrow O_{11} \longrightarrow Z_{prew} \longrightarrow w_{11}$$

$$z_f = w_{11}^2 \underbrace{O_{11}}_{1} + \underbrace{w_{21}^2 O_{12}}_{0} + \underbrace{b_{21}}_{0}$$

$$\boxed{z_f = w_{11}^2}$$

$$O_{11} = \sigma(z\text{-}prew)$$
$$= \sigma(z\text{-}prew)\,[1-\sigma(z_{prew})]$$
$$= O_{11}(1-O_{11})$$

$$\frac{\partial z_{prew}}{\partial w_{11}^1} = \underbrace{w_{11}^1}_{1} x_{11} + \underbrace{w_{21}^1 x_{12}}_{0} + \underbrace{b_{11}}_{0}$$
$$= x_{11}$$

$$\boxed{\frac{\partial L}{\partial w_{11}^1} = -(y-\hat{y})\, w_{11}^2\, O_{11}(1-O_{11})\, x_{11}}$$

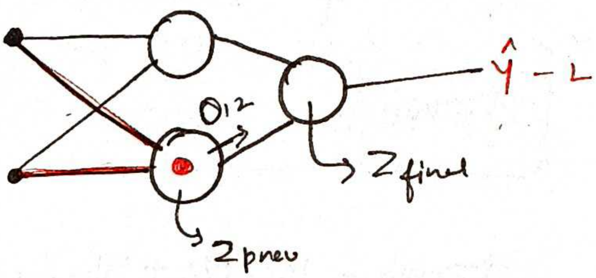
(2) $$\frac{\partial L}{\partial w'_{21}} = \left[\frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial Z_f} \times \frac{\partial Z_f}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial Z_{prev}}\right] \times \frac{\partial Z_{prev}}{\partial w'_{21}}$$

→ solved in (1)

$$\frac{\partial Z_{prev}}{\partial w'_{21}} = \underbrace{w'_{11} x_{i1}}_{0} + \underbrace{w'_{21} x_{i2}}_{1} + \underbrace{b_{11}}_{0}$$

$$\boxed{\frac{\partial L}{\partial w'_{21}} = -(y-\hat{y})\, w^2_{11}\, O_{11}\,(1-O_{11})\, x_{i2}}$$

(3) $$\boxed{\frac{\partial L}{\partial b_{11}} = -(y-\hat{y})\, w^2_{11}\, O_{11}\,(1-O_{11})}$$



(1) $\frac{\partial L}{\partial w'_{12}}$, (2) $\frac{\partial L}{\partial w'_{22}}$, (3) $\frac{\partial L}{\partial b_{12}}$

$L \rightarrow \hat{y} \rightarrow Z_{final} \rightarrow O_{12} \rightarrow Z_{prev} \rightarrow w'_{12}$

already calculated

1) $$\frac{\partial L}{\partial w'_{12}} = \boxed{\frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial Z_{final}}} \times \boxed{\frac{\partial Z_{final}}{\partial O_{12}}} \times \frac{\partial O_{12}}{\partial Z_{prev}} \times \frac{\partial Z_{prev}}{\partial w'_{12}}$$

$$\frac{\partial Z_{fin}}{\partial O_{12}} = \underbrace{w^2_{11} O_{11}}_{0} + \underbrace{w^2_{21} O_{12}}_{1} + \underbrace{b_{21}}_{0}$$

w^2_{21}

$(O_{12}(1-O_{12}))$

x_{i1}

$$\boxed{\frac{\partial L}{\partial w'_{12}} = -(y-\hat{y})\, w^2_{21}\, O_{12}\,(1-O_{12})\, x_{i2}}$$

2) $$\boxed{\frac{\partial L}{\partial w^1_{22}} = -(y-\hat{y})\, w^2_{21}\, O_{12}\,(1-O_{12})\, x_{i2}}$$

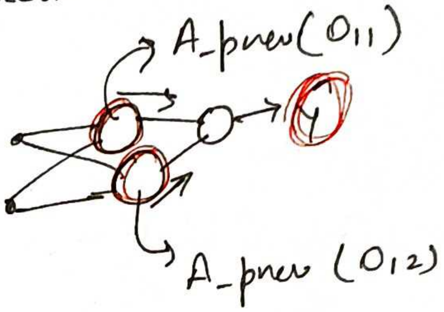
3) $$\boxed{\frac{\partial L}{\partial b_{12}} = -(y-\hat{y})\, w^2_{21}\, O_{12}\,(1-O_{12})}$$

→ Now, code ← [code_problem]

L_layer_forward → function

return A, A_prev

y_hat



Bakpropagation → The Why?

The intuition behind the algorithm

Bakpropagation Algorithm

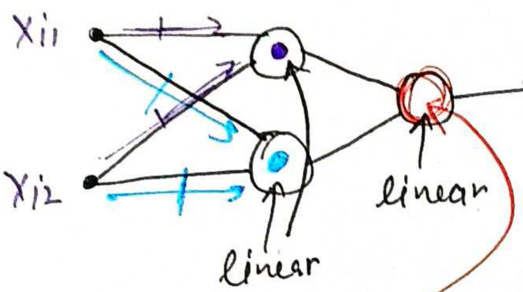
```
epochs = 5
for i in range (epochs):
    for j in range (X.shape[0]):
        → Select 1 row (random)
        → Predict (using forward propagation)
        → Calculate loss (using loss function → mse)
        → Update weights and bias using GD
              Wn = Wo - η ∂L/∂w
    → Calculate avg loss for the epoch
```

→ Loss function is a function of all [Trainable parameters]

$\boxed{L=(C-\hat{y})^2}$

cgpa | Profile-resume | Lpa

data → target
↳ $\boxed{\text{Constant}}$



$\hat{y} \rightarrow L$ → $mse = (y-\hat{y})^2$

$L = (y-\hat{y})^2$ → Mathematical func.

$L(\hat{y})$

Loss is function of $\hat{y}$

Ex: $y = f(x)$

y is a function x

which means, when we change x then how much y will be change.

how to find $\hat{y}$ of this Node?

$\hat{y} = w_{11}^2 O_{11} + w_{21}^2 O_{12} + b_{21}$

what is the $\hat{y}$ of this?

$$\hat{y} = \underset{①}{w_{11}^2}\left[\underset{②}{w_{11}^1} x_{i1} + \underset{③}{w_{21}^1} x_{i2} + \underset{④}{b_{11}}\right] + \underset{⑤}{w_{21}^2}\left[\underset{⑥}{w_{12}^1} x_{i1} + \underset{⑦}{w_{22}^1} x_{i2} + \underset{⑧}{b_{12}}\right] + \underset{⑨}{b_{12}}$$

Constant

Variable

↳ Trainable Parameter

$L(w, b)$

↳ 9 things

Concept of Gradient

↳ fancy word of derivative

Eg: $y = f(x) = x^2 + x$ → Same

Let assume, derivate of y w.r.t x

$$\frac{dy}{dx} = \frac{d}{dx}(f(x)) = \frac{d}{dx}(x^2+x) = 2x+1$$

$y \rightarrow x$ derivative $\frac{d}{dx}$ → derivative

For example:

$$Z = f(x,y) = \underbrace{x^2 + y^2}$$ → different parameter

derivative of z w.r.t x

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

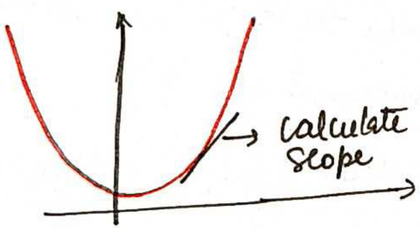
∂ → gradient (x^2+y^2) → different

d → derivative (x^2+x) → same

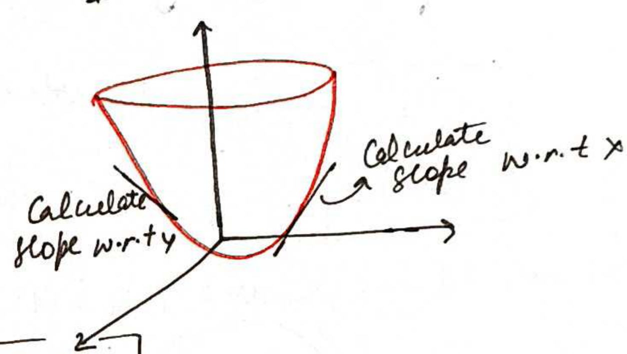
* Our loss function is complex $L(w^1_{11}, w^1_{12} \dots w^2_{41}, b_{21}, b_{41})$

$\frac{\partial L}{\partial w^1_{11}}, \frac{\partial L}{\partial b_{11}}$ → we use gradient for 9 parameter

eg: In 2-D



In-3D



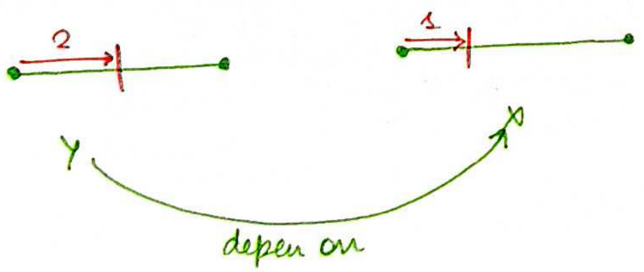
Loss function → $\boxed{\text{Complex 9-D function}}$ → $\boxed{\text{calculate 9 different slops}}$ w.r.t each dimension

Calculate gradient means

→ Concept of Derivative

$\frac{dy}{dx}$ y func of x where x is also a quantity

→ in physics derivative mean → "rate of change"



X move → 1

then Y move → 2

that's mean Y depend on X.

for eg: $\frac{dy}{dx} = -2(-ve) \rightarrow$ if x is change in 1 unit than y is changes in -2.

Magnitude → how much quantity increase or decrease

Sign → +ve means increase or decr in (+ve), −ve mean decrease or increase in (−ve)

Derivative at a point

$Y = x^2 + 2x$ find $\boxed{x=5}$

$$\frac{dy}{dx} = 2x+1 \Rightarrow 2(5)+1 = 11$$

→ $\frac{dy}{dx}$ at $x=5$ is 11 or → slope of this at $x=5$ is 11.

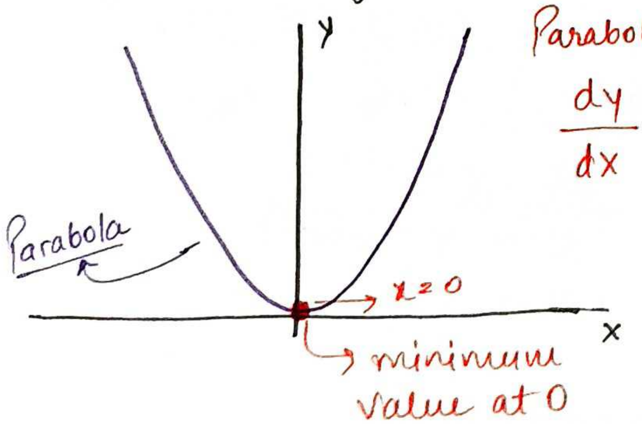
→ rate of change of w.r.t x at $x=5$ is 11.

$\boxed{\frac{\partial L}{\partial w_{11}}}$ → find when change w_{11} → 1 unit then loss → magnitude ? sign ?

$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial w}$$

with the help of derivative we are reducing W_{old}

Concept of Minima



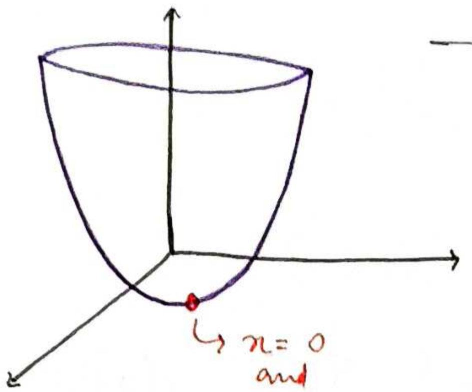
Parabola eqn

$$\frac{dy}{dx} = 2x$$

for finding minimum val

$2x = 0$

$\boxed{x=0}$ → x at 0 is minimum value



$\rightarrow$ $z = x^2 + y^2$

$(x, y) \rightarrow$ minimum value

$\frac{\partial z}{\partial x} = x^2 + y^2 = 2x \rightarrow$ for min val
$2x = 0$
$\boxed{x = 0}$

$\frac{\partial z}{\partial y} = x^2 + y^2 = 2y \rightarrow$ for min value
$2y = 0$
$\boxed{y = 0}$

* Minimum loss function (z value or) at $x = 0$ and $y = 0$

Step first $\rightarrow$ find derivative of ~~of~~ 9 loss function

second : Put all loss function is equal to 0.

$$\frac{\partial L}{\partial w_{11}} \; \cdots \; \frac{\partial b}{\partial b_{12}} = 0$$

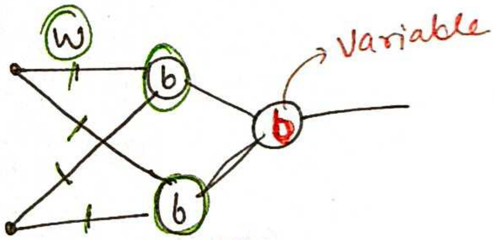
Backpropagation Intuition

$$\boxed{w_{new} = w_{old} - \eta \frac{\partial L}{\partial w}}$$

$\rightarrow$ $\eta = 1$ (assuming) $\rightarrow$ 1st assumption

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

$\rightarrow$ 2nd assumption



$\{w, b \rightarrow$ $\boxed{\text{Constant}}$

$L = L(w, b)$

9 Parameter $\rightarrow$ 8 parameter are constant

* These Assumptions only for easy calculation

$\rightarrow$ After discussion complete we have to assume that all constant parameter are not constant any more.

Now,

Loss $= L(b_{21})$

$L = (y - \hat{y})^2 \rightarrow \hat{y} = (b_{21})$

$L(b_{21})$

$\hat{y}$ depend on b_{21}. So, L also depend on b_{21}

Now,

$b_{21} = b_{21} - \boxed{\frac{\partial L}{\partial b_{21}}}$ → We have to think → why subtract this to b_{21}

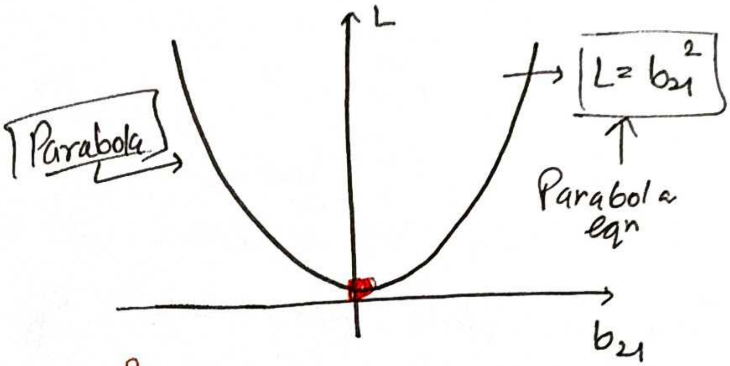
Also why negative sign is here

$\boxed{\frac{\partial L}{\partial b_{21}}}$ → L derivative w.r.t b_{21}

change b_{21} → 1 unit

L → ? change

sign ? ← ↳ magnitude?



[Find the value of b_{21} when L is minimum]

* target is to reduce L ↓↓

① $\frac{\partial L}{\partial b_{21}} \xrightarrow{\text{answer}}$ +Ve ↳ which means b_{21}↑ increase and L↑ also increase.

→ So, we have to reduce ↓ → b_{21} for L↓ bcz when we increase b_{21}↑ value and L↑ also increase and answer ~~is~~ positive (+ve) that's why b_{21}↑.

$$\boxed{b_{21} = b_{21} - \frac{\partial L}{\partial b_{21}}}$$

let assume $b_{21} = 5$

reduce b_{21}

~~add~~ -ve sign

$\boxed{b_{21} = 5 - \left(+\frac{\partial L}{\partial b_{21}}\right)}$

② $\frac{\partial L}{\partial b_{21}} \xrightarrow{\text{answer}}$ -ve ↳ which means b_{21}↑ L↓

→ So, here we have to increase b_{21} for L decrease Add in b_{21} value to increase value. b_{21}↑

~~bzzz~~

$$\boxed{b_{21} = b_{21} - \left(-\frac{\partial L}{\partial b_{21}}\right)}$$ negative

$$\boxed{b_{21} = b_{21} + \frac{\partial L}{\partial b_{21}}}$$

→ increase b_{21}↑. So, L↓

$$\boxed{b_{21} = b_{21} - \frac{\partial L}{\partial b_{21}}}$$

This negative is very smart (How)?

$$\boxed{b_{21} = b_{21} - \boxed{\frac{\partial L}{\partial b_{21}}}}$$

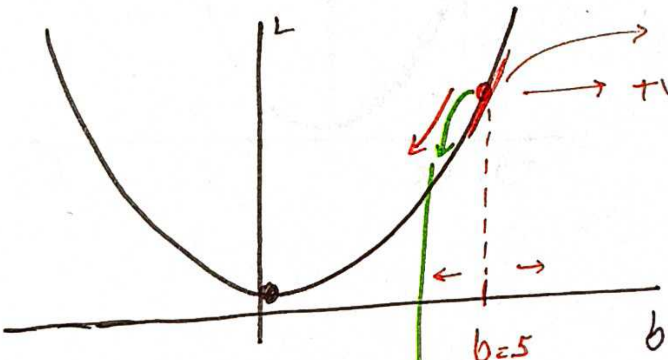
When derivative output is (-ve). So, it add into b_{21} and increase b_{21} value for decrease Loss.

$$\boxed{b_{21} = b_{21} - \boxed{\frac{\partial L}{\partial b_{21}}}}$$

When this derivative is (+ve) output. So, derivative output subtract from b_{21} value and reduced b_{21} value.

★ All game is depend on sign of derivative

Graphically Understanding



Calculate slope (derivative)

$\frac{\partial L}{\partial b}$ = slope

→ +ve

↳ reduce b value

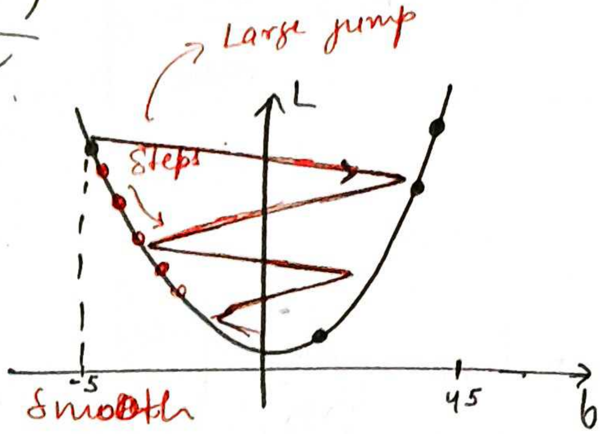
how much step go down?. So, thats why we use $b_{21} = b_{21} - \frac{\partial L}{\partial b}$ → slop value dist move down

move $\frac{\partial L}{\partial b}$ distance

to reach L minimum value.

Effect of Learning Rate (η)

$$w_n = w_0 - \eta \frac{\partial L}{\partial w}$$



Learning Rate (η) $\xrightarrow{\text{helps}}$ step → smooth

let, $\frac{\partial L}{\partial b} = -50$

$$b = -5 - (-50) = 45$$

$$\boxed{b = 45}$$

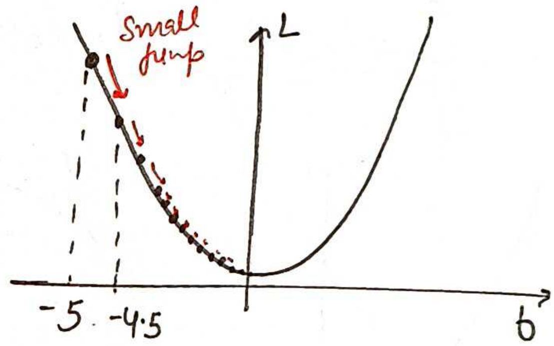
again $\frac{\partial L}{\partial b} = 50$

$$b = 45 - 50 = -5$$

$$\boxed{b = -5}$$

Now, using learning rate

$b = -5$, $\frac{\partial L}{\partial b} = -50$, $\eta = 0.01$



$$b = -5 + (50 \times 0.01)$$

$$\boxed{b = -4.5}$$

* If we take very small learning Rate then our model take large time to reach minimum loss. That's why Learning Rate parameter is very crucial
* Use google tool to visualize learning Rate

→ What is convergence?

[Loop kitni baar chalana hai $\xrightarrow{\text{depend on}}$ convergence hua ki nhi.]

$$\boxed{W_n = W_o - \eta \frac{\partial L}{\partial W}}$$

$\boxed{W_n \simeq W_o} \rightarrow$ convergence

but practically we write
"epoch = 100"
"epoch 100 bar chalega to convergence ho hi jayega"

★ Google Developer tool for backpropagation for visual understanding