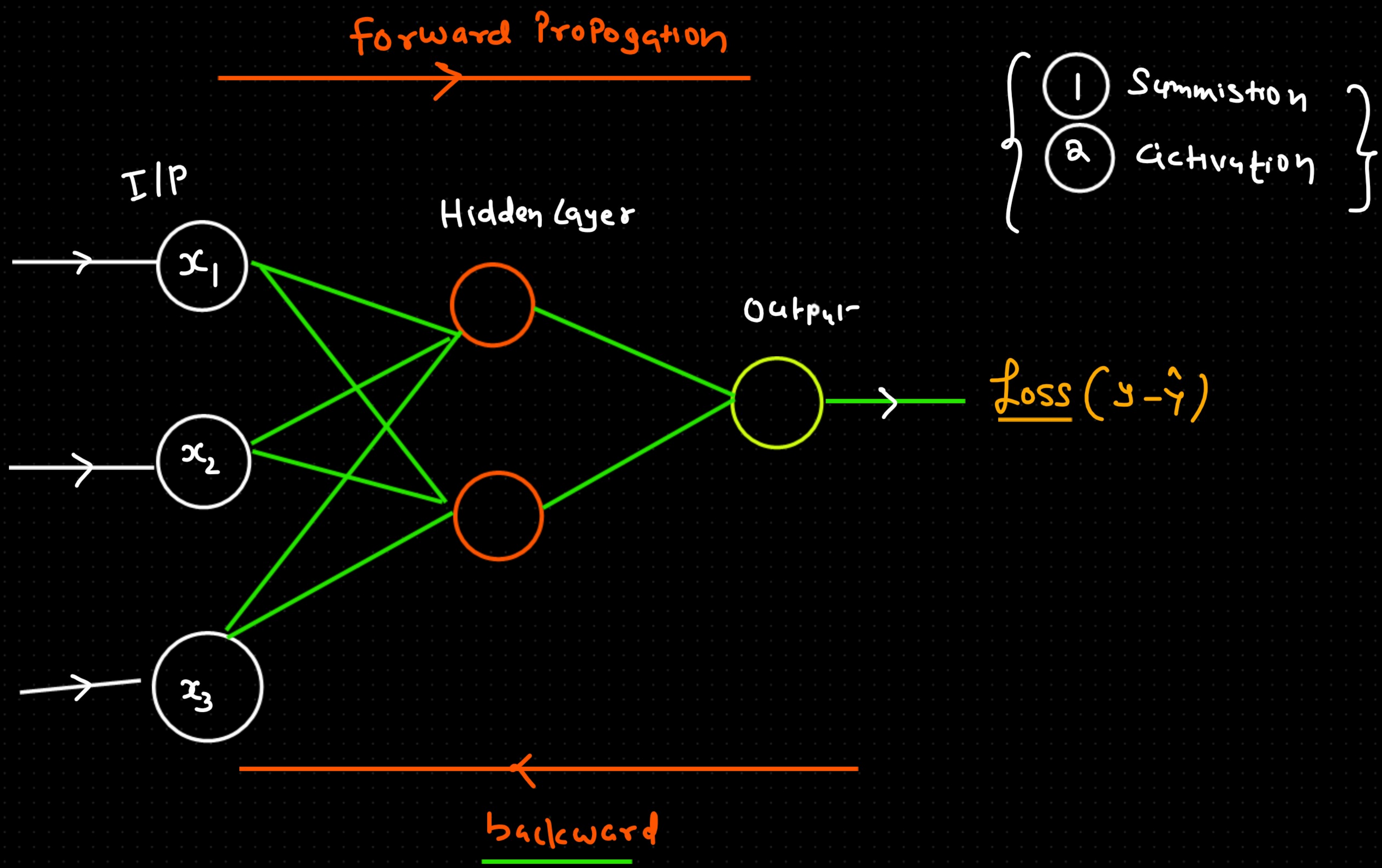


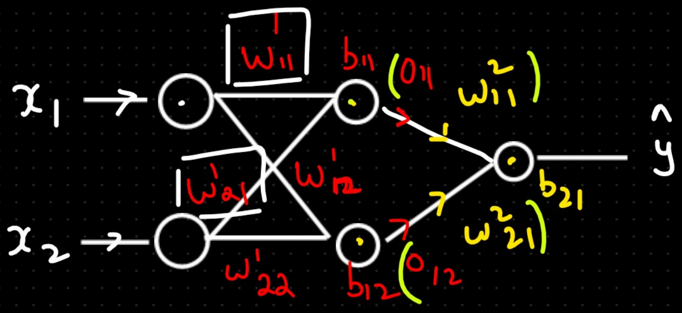
Forward Propogation
I/P
x_1
x_2
x_3
Hidden Layer
Output
Loss $(y - \hat{y})$
① Summistion
② activation
backward

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

$(y-\hat{y})$

$\uparrow$

fix



Loss $= (y-\hat{y})$

$\Downarrow$

?

Actual value

$\boxed{y = f(x)}$

$\boxed{L(\hat{y})}$

$\sigma(\hat{y})$

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

Loss $(y-\hat{y})$

$\downarrow$

fix

$L(\hat{y})$

$$\boxed{\hat{y} = w^2_{11}\left[x_1 w^1_{11} + w^1_{21} x_2 + b_{11}\right] + w^2_{21}\left[w^1_{12} x_1 + w^1_{22} x_2 + b_{12}\right] + b_{21}}$$

(Loss fn is a fn of all the given params)

Gradient :-

- Slope ⇒ ratio b/w y and x
- derivative ⇒ rare of change

$$\boxed{y = f(x) = x^2 + x}$$

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

{rate of chage of y wrt x}

Partial derivation

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

gradient

$$\frac{\partial z}{\partial x} = \frac{\partial}{\partial x}(x^2 + y^2) = 2x$$

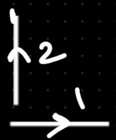
$$\frac{\partial z}{\partial y} = \frac{\partial}{\partial y}(x^2 + y^2) = 2y$$

(Loss fn depends on many weight | Biases)

Derivatives

ratio y and x

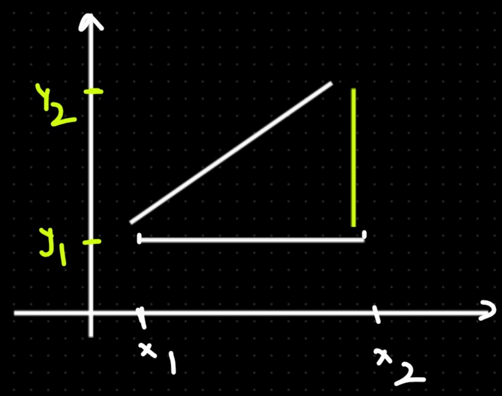
$2:1$



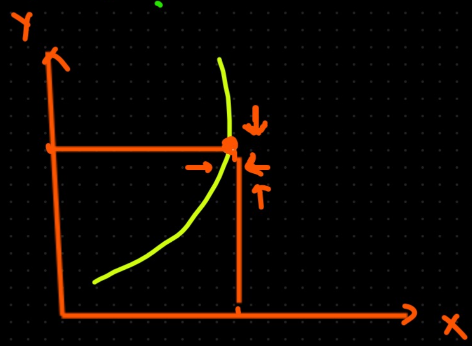
{rate of change}

$$\text{Slope} = \frac{\text{change in } y}{\text{chang in } x} = \frac{y_2 - y_1}{x_2 - x_1}$$

{avgerage slope}



How do you find the slope at a point?



derivatives is also a rate of at a particular point

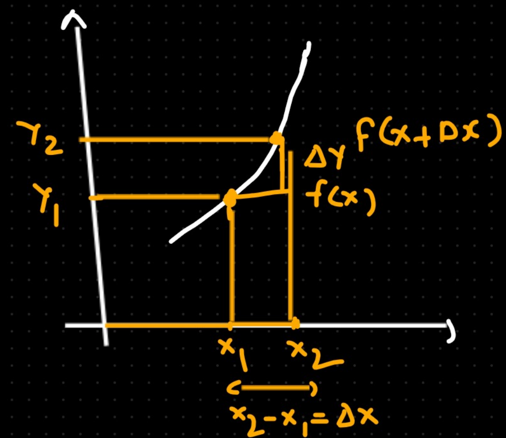
$y = f(x)$

$$\text{Slope} = \frac{\text{change in } y}{\text{change in } x} = \frac{\Delta y}{\Delta x}$$

$y = f(x)$

x changes from $\boxed{x \text{ to } \boxed{x + \Delta x}}$

y changes from $f(x)$ to $f(x+\Delta x)$



$$= \boxed{\frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x}}$$

$$\boxed{\Delta x \rightarrow 0}$$

$$\boxed{x^2 = 2x}$$

$$f(x) = x^2$$

$$f(x+\Delta x) = (x+\Delta x)^2 \Rightarrow (a+b)^2 \Rightarrow a^2 + b^2 + 2ab$$

$$= (x^2 + 2x\Delta x + (\Delta x)^2)$$

$$= \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$= \frac{\cancel{x^2} + 2x\Delta x + (\Delta x)^2 - \cancel{x^2}}{\Delta x}$$

$$= \frac{2x\Delta x + (\Delta x)^2}{\Delta x}$$

$$= \frac{\cancel{\Delta x}(2x + \Delta x)}{\cancel{\Delta x}} = \boxed{2x + \Delta x}$$

$\boxed{2x + \Delta x}$ $\quad \Delta x \to 0$

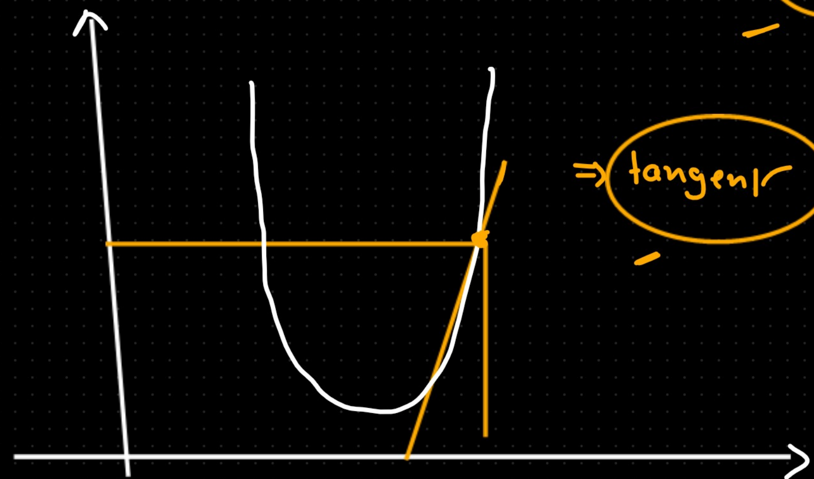
$\Downarrow$

$$y = f(x) = \frac{dx^2}{dx} = 2x$$

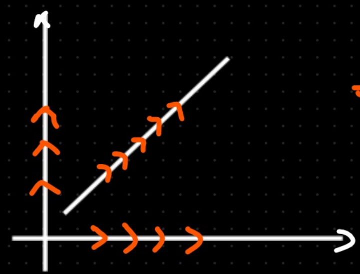
$$= 2 \times 1 = 2$$

$x = 1$

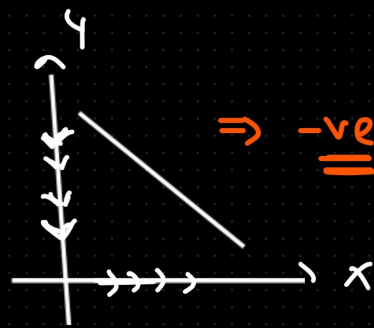
x^2



$\Rightarrow$ tangent



$\Rightarrow$ +ve

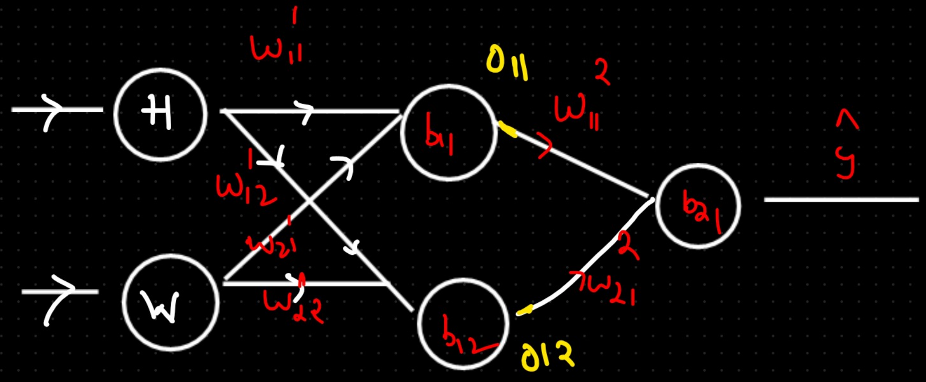


$\Rightarrow$ -ve

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

$$\left(\frac{dy}{dx}\right)_{x=5} = ?$$

Height	Weight	BMI
160	60	21
~~170~~	~~65~~	~~22~~
180	70	20
190	71	18



1. init weight and bias $\left\{ \begin{array}{l} (w, b) \rightarrow \text{Random value} \\ w \rightarrow 1, \; b \rightarrow 0 \end{array} \right]$

   formula w & b

2. Point (row)

3. FP [Mat multiplication, Dot Product]

4. Loss fn $\Rightarrow$ $L = \boxed{(y - \hat{y})^2}$ $L(\hat{y})$

   fn

5. update w & b

   $\hookrightarrow$ gradient decent

$(21 - 24) = (-3)^2$

$\Rightarrow$ g $\leftarrow$ error $\downarrow$

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

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

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

$O_{11} \rightarrow$

$w^2_{21} \downarrow$

$O_{21} = \hat{y} \leftarrow w^2_{11}$

$O_{12} \uparrow$

$\nwarrow b_{21}$

Here we have a complex hirarchy

$$\hat{y} = w^2_{11}\left[x_1 w^1_{11} + w^1_{21} x_2 + b_{11}\right] + w^2_{21}\left[w^1_{12} x_1 + w^1_{22} x_2 + b_{12}\right] + b_{21}$$

GD

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

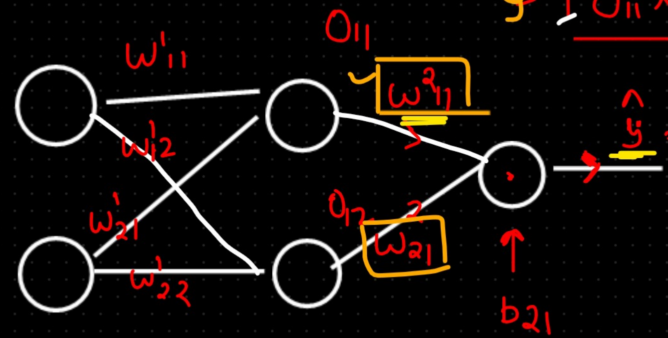
← Partial derivative

if you want to minimise the loss then you will have to adjust all the weight bars.

BackPropogation → to adjust all the all Params (w & b) we have to go back in our NN. that's y it BP

Back Adjustment

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



$$\hat{y} = O_{11} \times W^2_{11} + O_{12} \times W^2_{21} + b_{21}$$

$$\hat{y} = O_{21} \Rightarrow \sigma(O_{21})$$

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

$$= L(\hat{y})$$

$$\hat{y} \rightarrow L(y-\hat{y})$$

→ rate of chang in L wrt w^2_{11}

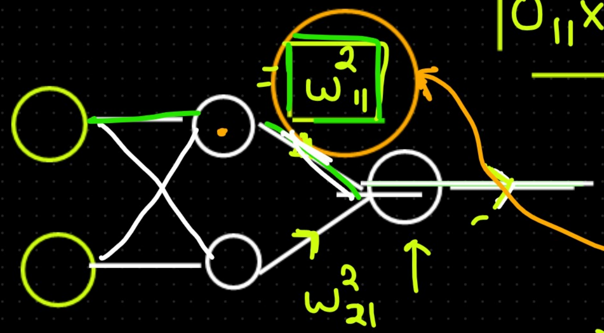
$$\frac{\partial L}{\partial w^2_{11}} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w^2_{11}} \leftarrow$$

$$(y-\hat{y}) \leftarrow \hat{y} \leftarrow O_{21} \leftarrow w^2_{11}$$

↑

chain rule of Differentiation

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



$\hat{y} = O_{21}$ Loss $\Rightarrow (y - \hat{y})$

$$\frac{\partial L}{\partial w_{11}^2} = \frac{\partial L}{\partial \hat{y}} \cdot$$

$$= \boxed{\frac{\partial L}{\partial O_{21}} \times \frac{O_{21}}{\partial w_{11}^2}}$$

Chain rule

$$\boxed{\text{Bp} = \text{chain rule of differentiation}}$$

$$\frac{\partial L}{\partial w_{11}^2} = \boxed{\frac{\partial L}{\partial \hat{y}}} \times \frac{\partial \hat{y}}{\partial w_{11}^2}$$

$$= \frac{\partial L}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial w_{11}^2}$$

$\Rightarrow$ both are same

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

$$\frac{\partial (y-\hat{y})^2}{\partial \hat{y}}$$

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

$$\frac{\partial y}{\partial \hat{y}} - \frac{\partial \hat{y}}{\partial \hat{y}}$$

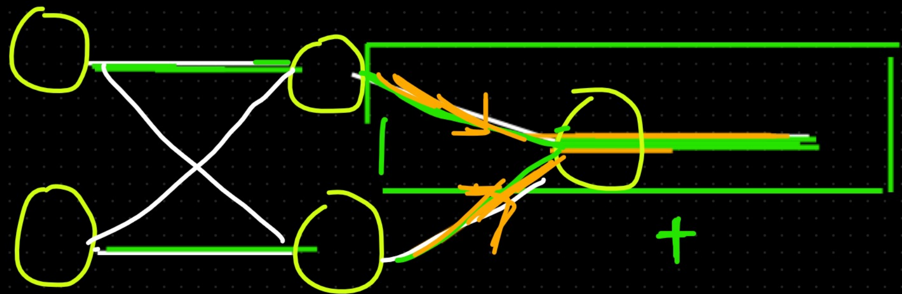
$$0 \quad (-1)$$

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

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

$$\frac{\partial \hat{y}}{\partial w_{11}^2} = O_{11}$$

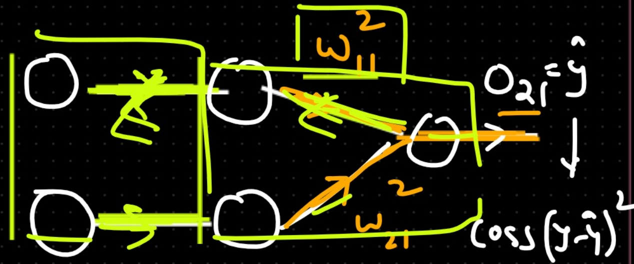
$$\boxed{\frac{\partial L}{\partial w_{11}^2} = -2(y-\hat{y})O_{11}}$$



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

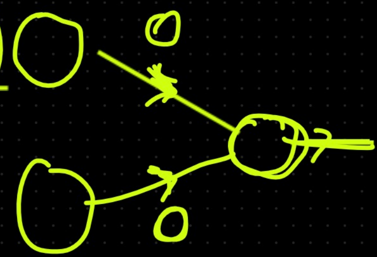
$$= O_{11}$$

$$\Rightarrow \frac{\partial L}{\partial w_{21}^2} \Rightarrow \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_{21}^2}$$



$$\Rightarrow -2(y-\hat{y})\, O_{12} \qquad \frac{\partial\left[\overset{0}{O_{11} w_{11}^2} + \overset{1}{O_{12} w_{21}^2} + \overset{0}{b_{21}}\right]}{\partial w_{21}^2}$$

$$\left(-2(y-\hat{y})O_{11} + (-2)(y-\hat{y})O_{12}\right)$$

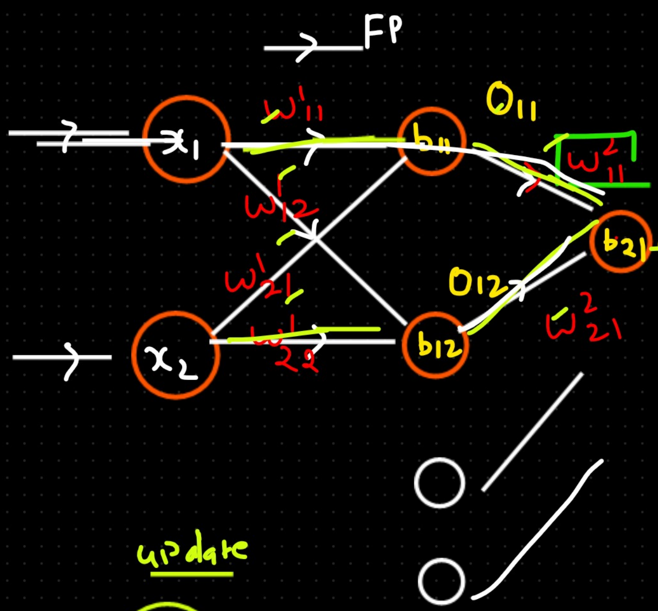


Assignment

Python → BP

NN from Scratch

Sunny.savita@ineuron.ai



$\rightarrow \boxed{O_{21}} \rightarrow \boxed{\hat{y}} \Rightarrow$ loss $(y-\hat{y})$

$\downarrow$

$L(\hat{y})$

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

$\downarrow$

$\left[x_1 w_{11}^1 + x_2 w_{21}^1 + b_{11}\right] w_{11}^2 + \left[x_1 w_{12}^1 + x_2 w_{22}^1 + b_{12}\right] w_{21}^2 + b_{21}$

update

BP $\Rightarrow$ train the Parameter

$\downarrow$

(w, b)

0.01

0-1

G.D. = $w_{new} = \boxed{w_{old}} - \boxed{\eta} \frac{\partial L}{\partial w} \leftarrow$

$b_{new} = b_{old} - \eta \frac{\partial L}{\partial b}$

$\frac{\partial L}{\partial w} \Rightarrow$ Partial Derivate

$\downarrow$

gradient

$\frac{\partial L}{\partial w} \Rightarrow \times$ in between so many thing

$$\boxed{\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_{11}^2}}$$

$\downarrow$

Chain rule

$\frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_{11}^2}$

$\Downarrow$

$\frac{\partial L}{\partial O_{21}}$

$\frac{\partial L}{\partial w_{11}^2}, \frac{\partial L}{\partial w_{21}^2}, \frac{\partial L}{\partial w_{11}^1}, \frac{\partial L}{\partial w_{21}^1}, \frac{\partial L}{\partial w_{12}^1}, \frac{\partial L}{\partial w_{22}^1}$

$\frac{\partial O_{21}}{\partial w_{11}^2} \Rightarrow$

ANN

2

2

→ Python → BP → from Scracth — reverse

→ act, Loss, (optimizer)

near weekend

Problem and Performance

2 Practical → tensorflow & Pytorch

4-5 2 → classes

6 class

3 week

any data ⇒ NN ⇒ FP + BP ← in Python from Scracth

Dummy

Keras