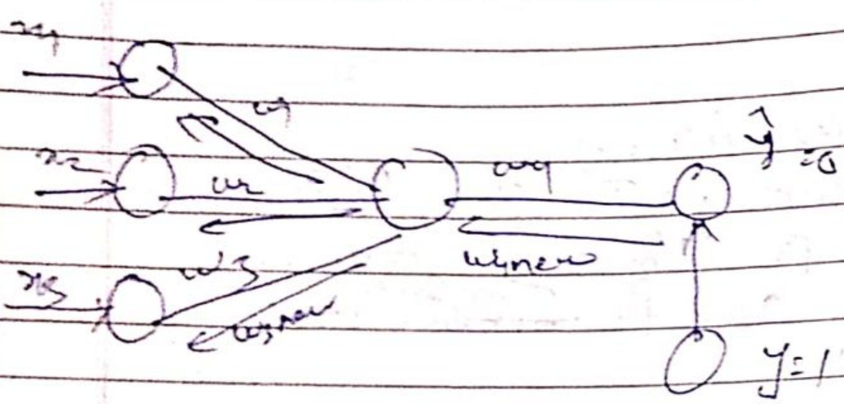
DEEP Learning

① ANN ② CNN ③ RNN

Back Propagation

	Play	Study	Sleep
	2h	4h	8h



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

$(1-0)^2 = 1$ ↓ this should be reduce

So, to reduce lost we do Back propagation.
So we need to adjust wegiht for same

minimize loss
→ optimizer

η = learning rate

$w_{4new} = w_{4old} - \eta \frac{dL}{dw_4}$

w_3, w_2 here will calculate as same way

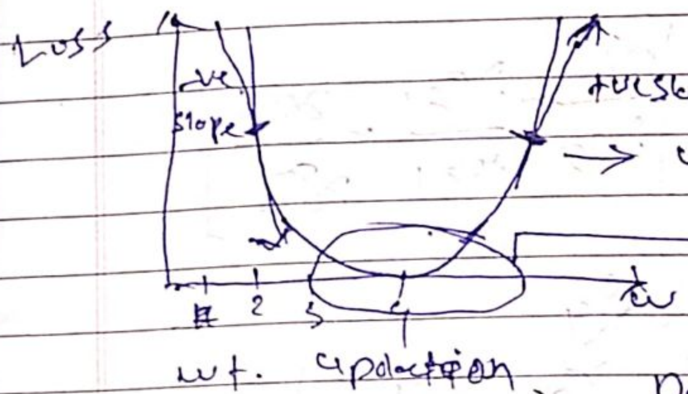
after getting all wegiths again front propagation will happen

Cost function → For multiple no. of records

$= \sum_{i=1}^{n} (y-\hat{y})^2$ as here we have only O/P 1

Optimisers

* Gradient Decent:-



→ wt. points

→ global minima, our main aim of gradient Decent to achive this area

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

η → Learning Rate. Should be take, which suite for G.D.

$\frac{\partial L}{\partial w}$ → derivitive will give slope of the wt. points.

If slope is -ve then it will add

-II- +ve then it will ~~do~~ minus

Unless an until we reach to the global minima updation of wt. will happen with Back Propogation

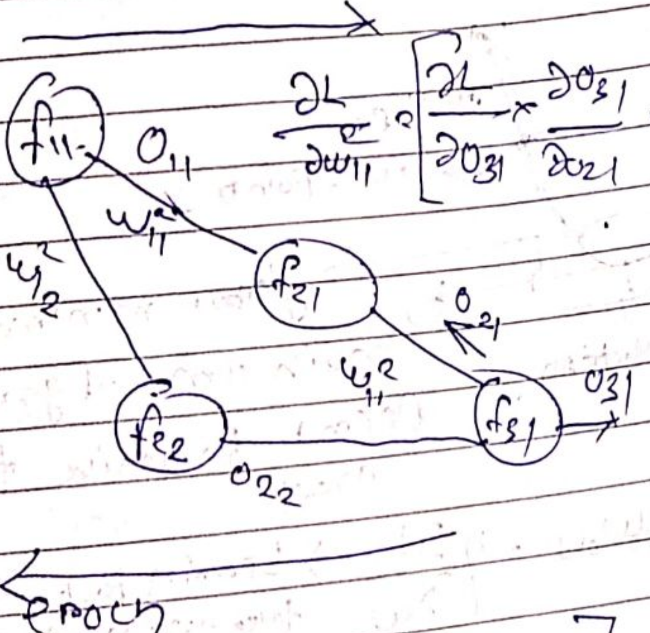
[loss $= (y-\hat{y})^2$], this value must be decreasing so we can achive global minima

Hyperparameter optimisation helps to give Learning Rate.

CHAIN RULE IN BACK Propagation

$$\frac{\partial L}{\partial w^3_{11}} = \frac{\partial h}{\partial_{31}} \times \frac{\partial O_{31}}{\partial w^2_{11}}$$

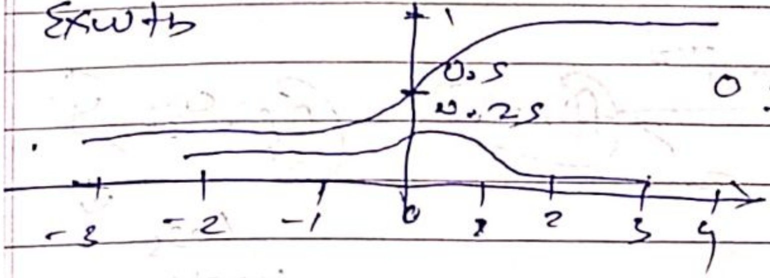
$$\frac{\partial L}{\partial w^2_{11}} = \left[\frac{\partial L}{\partial O_{31}} \times \frac{\partial O_{31}}{\partial O_{21}} \times \frac{\partial O_{21}}{\partial w^2_{11}}\right]$$



$$\frac{\partial L}{\partial w^2_{11}} = \left[\frac{\partial L}{\partial O_{31}} \times \frac{\partial O_{31}}{\partial O_{22}} \times \frac{\partial O_{22}}{\partial w^2_{12}}\right] + $$

Vanishing gradient Problem

Exwtb



$\frac{1}{1+e^{-z}}$ → Sigmoid

$0 \le \sigma(z) \le 0.25$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial O_{21}}{\partial O_{11}} \times \frac{\partial O_{11}}{\partial w_{11}}$$ Chain Rule

$= 0.20 \times 0.1 \times 0.05 \times 0.1$

$= 2.5 - 1 \times 0.04$

$= \cancel{2.45}$

$= 2.46$, 2.49999

~~After Some time~~

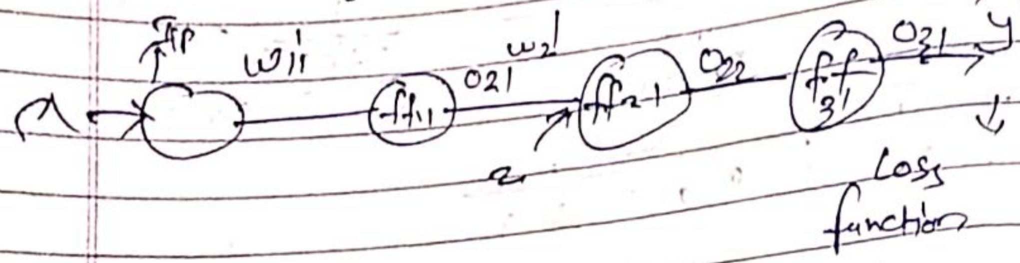
$w_{11new} \simeq w_{11old}$ → as its will Reduce learning Rate gradient Value decreasing

As gradient value will become very small as your previous wt. & new wt. will become approximately equal

As no. of layer increased derivative of wt. will become very small number.

& your previous wt. & new wt. are approximate equal number.

Exploding gradient problem



$$w^1_{11\,new} = w^1_{11\,old} - \eta \frac{\partial L}{\partial w^1_{11}}$$ → Learning Rate

$$\frac{\partial L}{\partial w^1_{11}} = \frac{\partial O_{31}}{\partial O_{21}} \boxed{\frac{\partial O_{21}}{\partial O_{11}}} \cdot \frac{\partial O_{11}}{\partial w^1_{11}}$$

$$O_{21} = \phi(z) \qquad \left[0 \le \frac{\partial \phi(z)}{\partial z} \le 0.25\right]$$

$$z = w_{21} \cdot O_{11} + b_2 \qquad [0 \le \phi(z) \le 0.25]$$

$$\frac{\partial O_{21}}{\partial O_{11}} = \frac{\partial \phi(z)}{\partial z} \cdot \frac{\partial z}{\partial O_{11}}$$

$$= 0 \le \phi(z) \le 0.25 \times \frac{\partial (w_{21} \times O_{11} + b_2)}{\partial O_{11}}$$

$$= -11- \times w_{21}$$

$$= 0.25 \times 500 = 125$$ → wt. are so high

wt. are high → it will get $\frac{\partial L}{\partial w_{11}}$ very high (consider) & fail to reach global minima