Day 2

1. forward propogation
2. Chain Rule of Derivative.
3. Vanishing Gradient Problem
4. loss function.

* Backword Propogation.

weight updating formula.

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

η = Learning rate.
∂ = Derivative.
L = Loss
W = Weights.

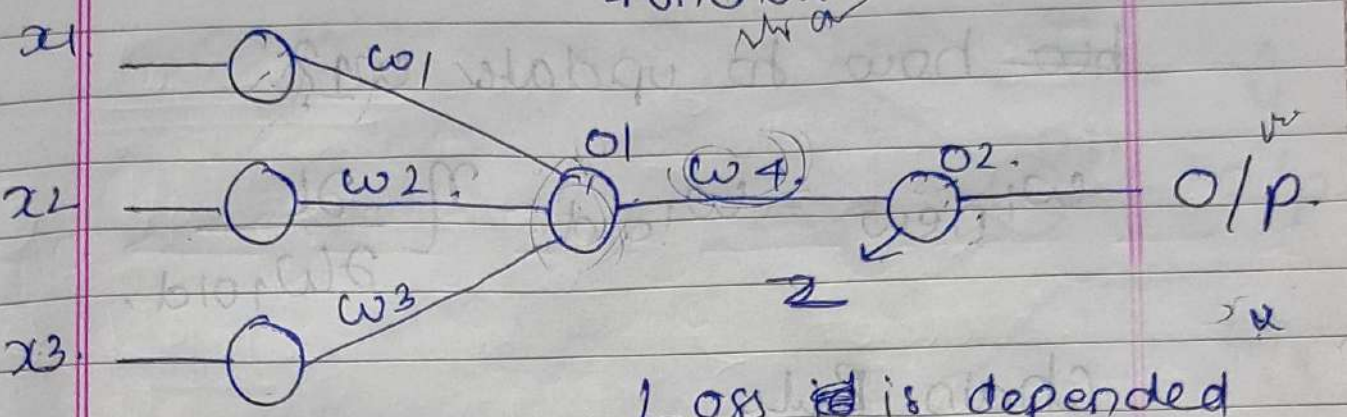
How to find $\frac{\partial L}{\partial W_{old}}$ ← To find this we use.

↓

Chain role of Derivative

$z = \sigma(o_1 w_4 + b)$.

σ = Activation function.



Loss is depended on o_2

$$\frac{\partial L}{\partial w_{4old}} = ?$$

$$w_{4new} = w_{4old} - \eta \frac{\partial L}{\partial w_{4old}}$$

Whenever we want to update w_4 { chain Rule of Derivative }

$$\frac{\partial L}{\partial w_{4old}} = \frac{\partial L}{\partial o_2} * \frac{\partial o_2}{\partial w_{4old}}$$

Bias Updation formula.

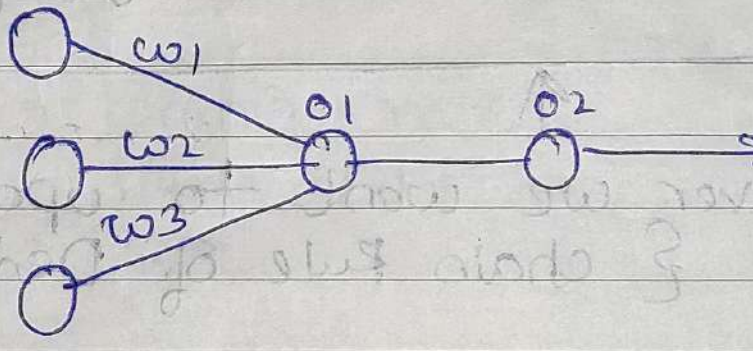
$$b_{2w} = b_{2old} - \eta \frac{\partial L}{\partial b_{2old}}$$

how to update w_1?

$$w_{1new} = w_{1old} - \eta \frac{\partial L}{\partial w_{1old}}.$$

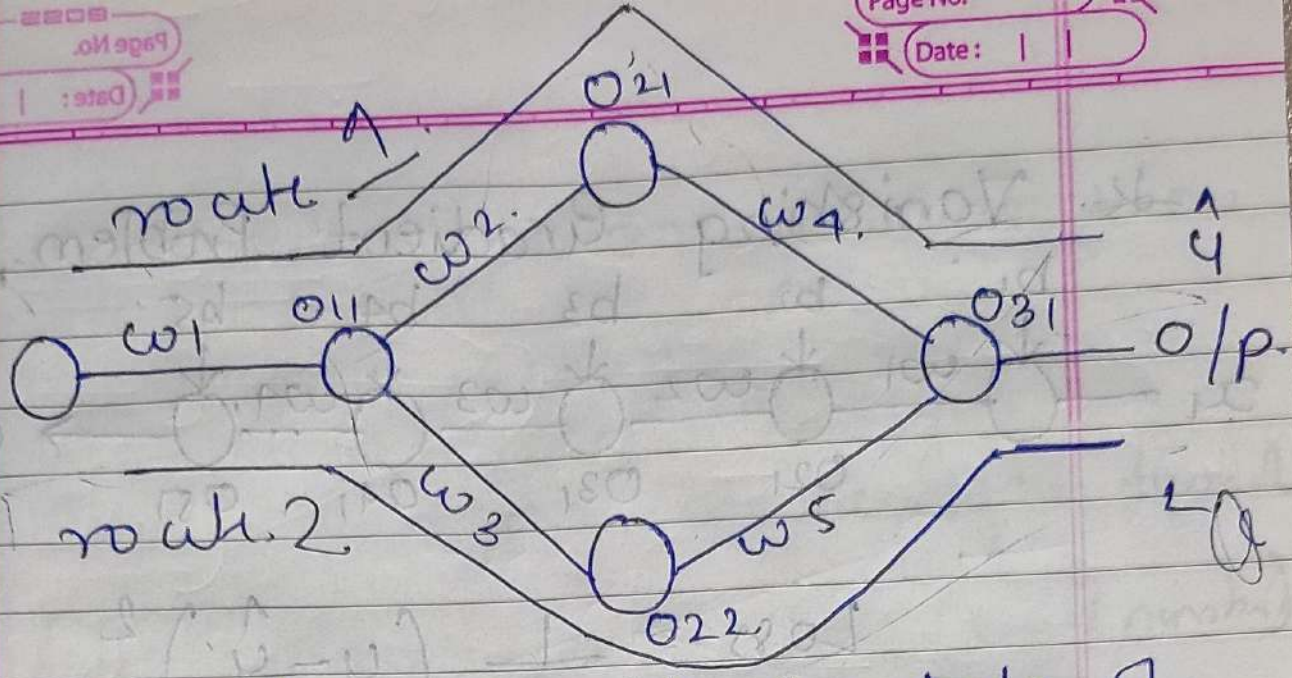
Chain Rule

loss is dependable on O_2. and O_2 is dependable on O_1 and O_1 is dependable on w_1.



$$\frac{\partial L}{\partial w_{1old}} = \frac{\partial L}{\partial O_2} \times \frac{\partial O_2}{\partial O_1} \times \frac{\partial O_1}{\partial w_{1old}}$$

This is the new chain Rule of differentiation to update w_1.



how to find w1 with the help of chain rule.

→ There are 2 two routes.

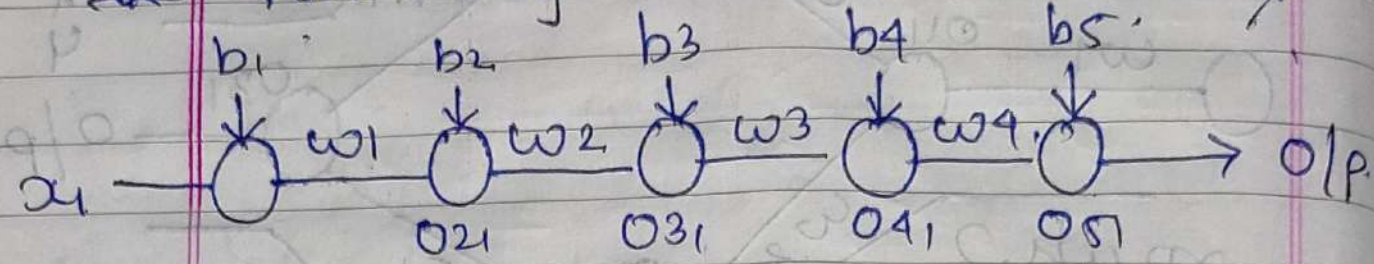
$$w1_{new} = w_{1old} - \eta \frac{\partial L}{\partial w_{1old}}$$

$$\frac{\partial L}{\partial w_{1new}} = \left[\frac{\partial L}{\partial o_{31}} * \frac{\partial o_{31}}{\partial o_{21}} * \frac{\partial o_{21}}{\partial o_{11}} * \frac{\partial o_{11}}{\partial w_{1old}}\right]$$

$$+$$

$$\left[\frac{\partial L}{\partial o_{31}} * \frac{\partial o_{31}}{\partial o_{22}} * \frac{\partial o_{22}}{\partial o_{11}} * \frac{\partial o_{11}}{\partial w_{1old}}\right]$$

* Vanishing Gradient Problem.



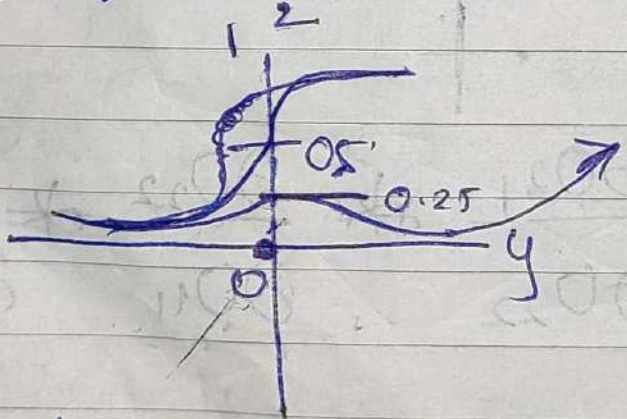
$$Loss = \frac{1}{2}(y - \hat{y})^2$$

↳ MSE

$$w_{1new} = w_{1old} - \eta \frac{\partial L}{\partial w_{1new}}$$

$$\frac{\partial L}{\partial w_{1new}} = \frac{\partial L}{\partial o_{51}} * \frac{\partial o_{51}}{\partial o_{41}} * \frac{\partial o_{41}}{\partial o_{31}} * \frac{\partial o_{31}}{\partial o_{21}} * \frac{\partial o_{21}}{\partial o_{w1}}$$

① Sigmoid Act function.



derivative is always range in between 0 to 0.25.

$0.25 * 0.15 * 0.10 * 0.05 * 0.2$,

$\downarrow$

= small value

Then well update the weight formula

$w_{new} = w_{old} - \eta \frac{\partial t}{\partial}$ (small number)

↳ small number.

$w_{new} \approx w_{old}$

Vanishing Gradient Problem

How to solve this?

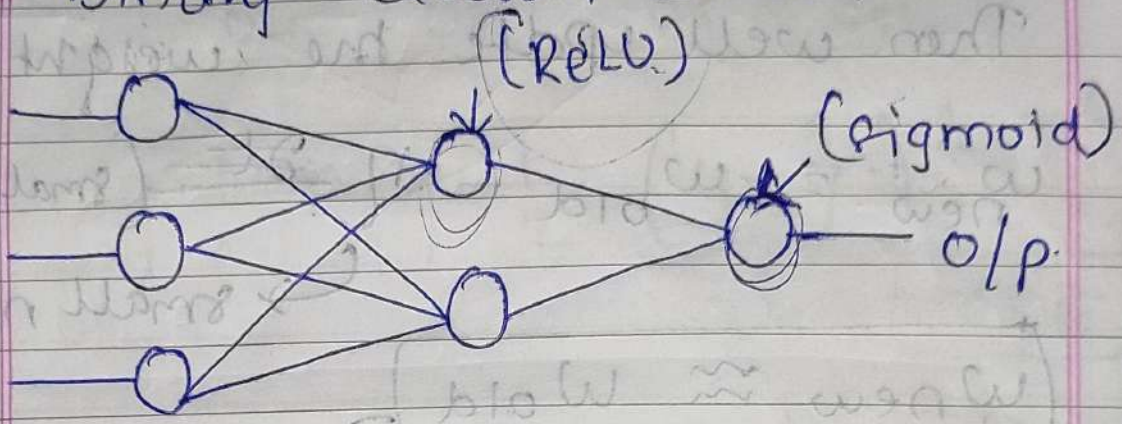
→ Use Another Activation function

1. Sigmoid
2. Tanh.
3. Relu
4. Leaky Relu
5. Re Relu

} Activation functions

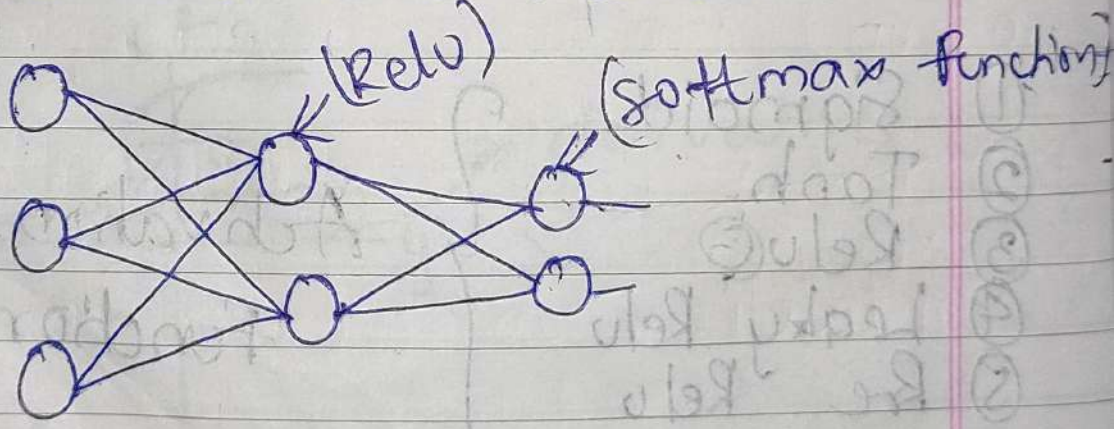
Technique which Activation function we should use.

* Binary classification.



* Always Remember whenever you are doing a binary classification problem your output should have the sigmoid Activation function.

* Multiclass classification.

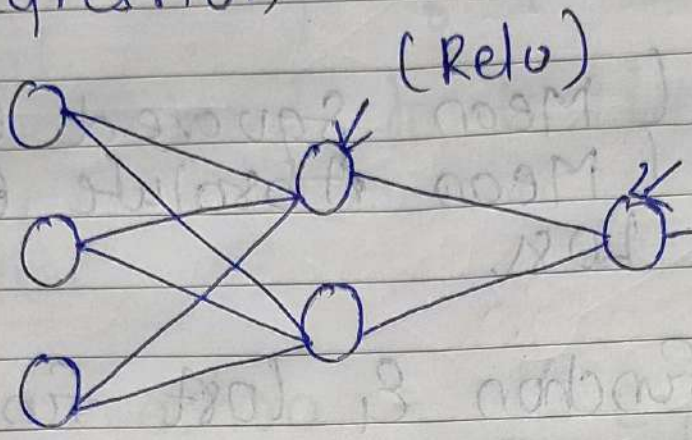


Loss function

↓

Linear Activation function

* Regression



* Loss function.

Deep Learning ANN.

↓ Regression

↓ classification.

Exp	Deg	Sal	...	Play	Study	P/f
10	phd	...	...	10	2	P
—	—	—		2	10	F
—	—	—		—	—	P

Regression Problem

classification.

Regression.

1. MSE (Mean Squared Error)
2. MAE (Mean Absolute Error).
3. Hubus Loss.

Loss function & lost function.

① MSE

Loss function $= \frac{1}{2}(y - \hat{y})^2$.

Lost function $= \frac{1}{2}\sum_{i=1}^{n}(y - \hat{y})^2$

→ Also called quadratic Equation.

Advantages of quadratic Equation

① Differentiable.

② It has only 1 local or Global minima.

③ It converges faster.

* Disadvantage of quadratic Equation.
Not
① Robust to Outliers

* MAE

Loss function $= \frac{1}{2}(y-\hat{y})^2$

Lost function $= \frac{1}{2}\sum_{i=1}^{n}(y-\hat{y})$

* Advantage.

① Robust to Outlier

* Hubes loss.

hyperparameter

Hubes loss is a combination of MSE and MAE

Loss fun $= \frac{1}{2}(y-\hat{y})^2$ if $|y-\hat{y}| \leq \delta$

Use when outliers are not present.

$\delta |y-\hat{y}| - \frac{1}{2}\delta^2$, otherwise

↑

use when outliers are present.

Classification.

Cross Entropy

↓

Binary Classification → Binary Cross Entropy → Binary Classification

Categorical Cross Entropy → Multi class classification

① Binary Cross Entropy.

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

↓↑

Logistic Regression

$$\text{Loss} = \begin{cases} -\log(1-\hat{y}) & \text{if } y = 0 \\ -\log(\hat{y}) & \text{if } y = 1 \end{cases}$$

① Categorical Cross Entropy.

f_1	f_2	f_3	O/P.
2	3	4.	Good
5	6	7	Bad.
8	9	10	Neutral.

Convert to One Hot Encoding.

↓

Good	Bad.	Neutral.
1	0	0
0	1	0
0	0	1

$$L(x_i, y_i) = -\sum_{j=1}^{c} y_{ij} * \ln(\hat{y}_{ij})$$

i = rows.

j = colums.

$y_i = [y_{i1}, y_{i2}, y_{i3} \ldots\ldots y_{ic}]$

$y_{ij} = \begin{cases} 1 \text{ if the element is in class} \\ 0 \text{ otherwise} \end{cases}$,

Conclusions.

Relu, Softmax → Multi class. → Categorical Cross Entrophy.

Relu, Sigmoid → Binary. → Binary Cross Entrophy.

Linear Regression

Relu, Linear Activation

↓

Loss function = MSE, MAE, Hubel