⇒ ACTIVATION FUNCTIONS IN NN :-

⇒ What are Activat^n Funct^n ?

→ In ANN, each neuron forms a weighted sum of its i/p & passes the resulting scalar value through a function referred as an activat^n funct (or) transfer funct^n.

→ If a neuron has n i/p's the o/p (or) activat^n of a neuron is

$$a = \underset{\substack{\downarrow \\ \text{activat}^n\ \text{funct}^n}}{g}(\omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3 + --- + \omega_n x_n).$$

→ It decides whether a neuron should be activated (or) not, and how strongly it should be activated.

⇒ Why do we need activat^n funct^n?

⇒ If we don't apply A.F, the network will become a linear regression model, because o/p of the NN without AF, is simply linear combinat^n of i/p's.

⇒ ∴ To capture the non-linear nature of the data, we use A.F.

⇒ Ideal Activatn Function :(characteristics):

1) Non - linear . → To capture non-linearity.

2) Differentiable → to apply Gradient descent. → (then to apply backpropogatn)

3) Computationally Inexpensive ⇒ calculatng derivative, should be simple & fast.

4) Zero - Centered : → the o/p from activatn functn should be normalized / zero-centered (mean = 0).

→ Ex : tanh

5) Non - saturating :- Saturatn occurs in A.F when the o/p of the functn approaches its max (or) min value. → (causes Vanishing gradient problem)

→ Like sigmoid squeeze o/p's to [0, 1] & tanh to [-1, 1]

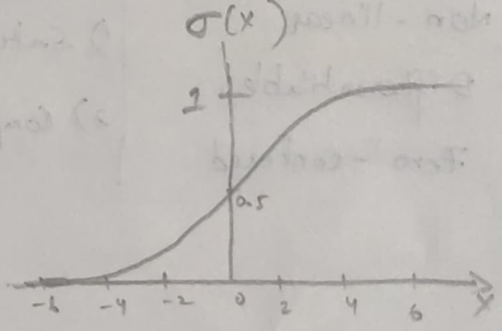
⇒ ReLU is a non-saturating, which operates in big range.

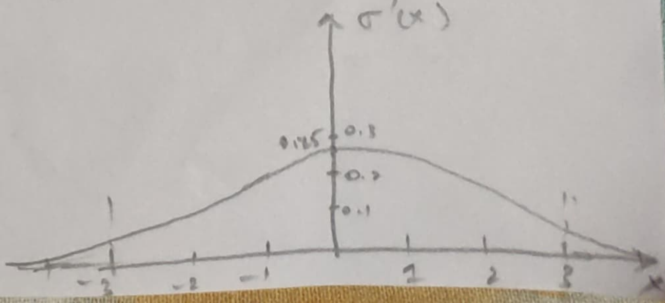
$f(x) = \max(0, x)$

⇒ SIGMOID ACTIVATION FUNCTN :

$$\boxed{\sigma(x) = f(x) = \frac{1}{1+e^{-x}}}$$

$\max(\sigma'(z)) = 0.25$





Advantages :-

1) $\sigma(x) \in [0, 1]$ → can be treated as probability

→ Can be used as "o/p layer" (for Binary classificat

2) Non - Linear

3) It is Differentiable → Grad. Descent can be applied

Disadvantages :-

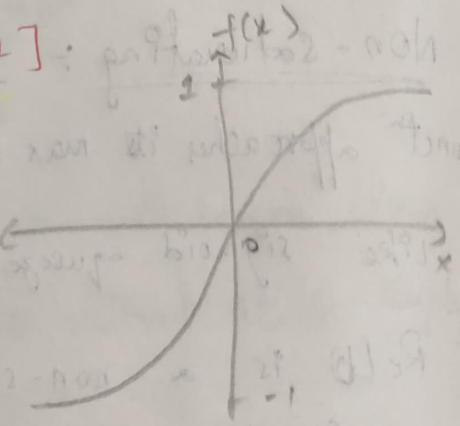
⇒ 1) Saturating functn → can cause vanishing gradient.

⇒ 2) Non-zero centered.

⇒ Tanh Activatn functn :- $\in [-1, 1]$

→ $$\boxed{f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}}$$

$$f'(x) = (1 - \tanh^2(x))$$



⇒ Advantages.

1) Non - linear

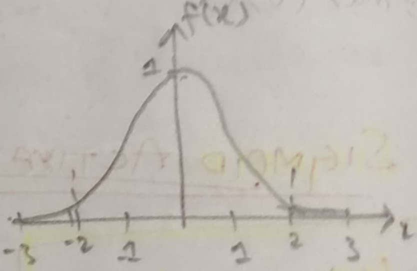
2) Differentiable

3) Zero - centered

DisAdv

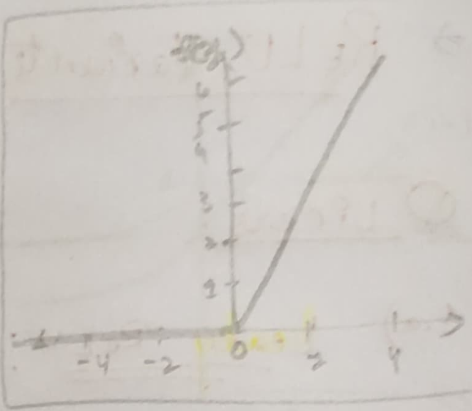
1) Saturatn functn

2) Computatnally expensive



⇒ RELU Activation functn :

$$\boxed{f(x) = \max(0, x)}$$



⇒ Adv.

1) Non - Linear

2) Non- saturated in +ve region.

3) Computationally inexpensive

4) Convergence → faster

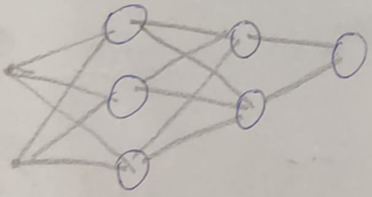
DisAdv.

1) Not completely differentiable (Not differentiable at '0')

2) Non -zero centered.

3) Dying ReLU problem.

⇒ DyING RELU Problem :

⇒ 

⇒ For some neurons, the value of o/p becames ~~for~~ for '0' for any given i/p, (dead neuron)

⇒ These dead neuron, is forever dead.

⇒ If this happens to more than 50% neurons, the learning / capturing data will be less than 50%.

⇒

⇒ Solutions :

↳① set low learning rate

↳② bias → +ve value → 0.01

↳③ Don't use ReLU → instead use it's variants.