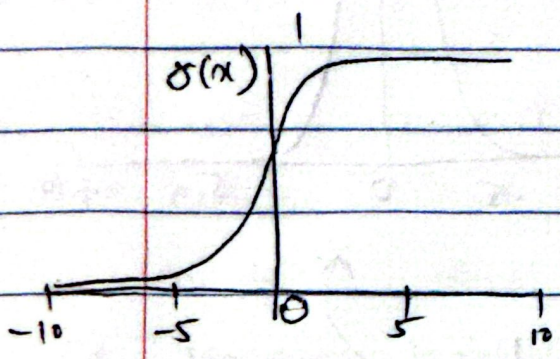
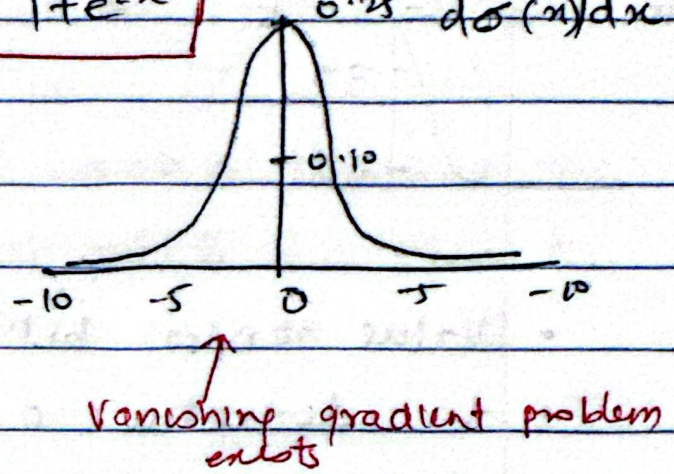
ACTIVATION FUNCTIONS

① Sigmoid

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





① When the input is slightly away from coordinate origin, the gradient of funct^n becomes very small almost zero in back propagat^n.

→ Adv

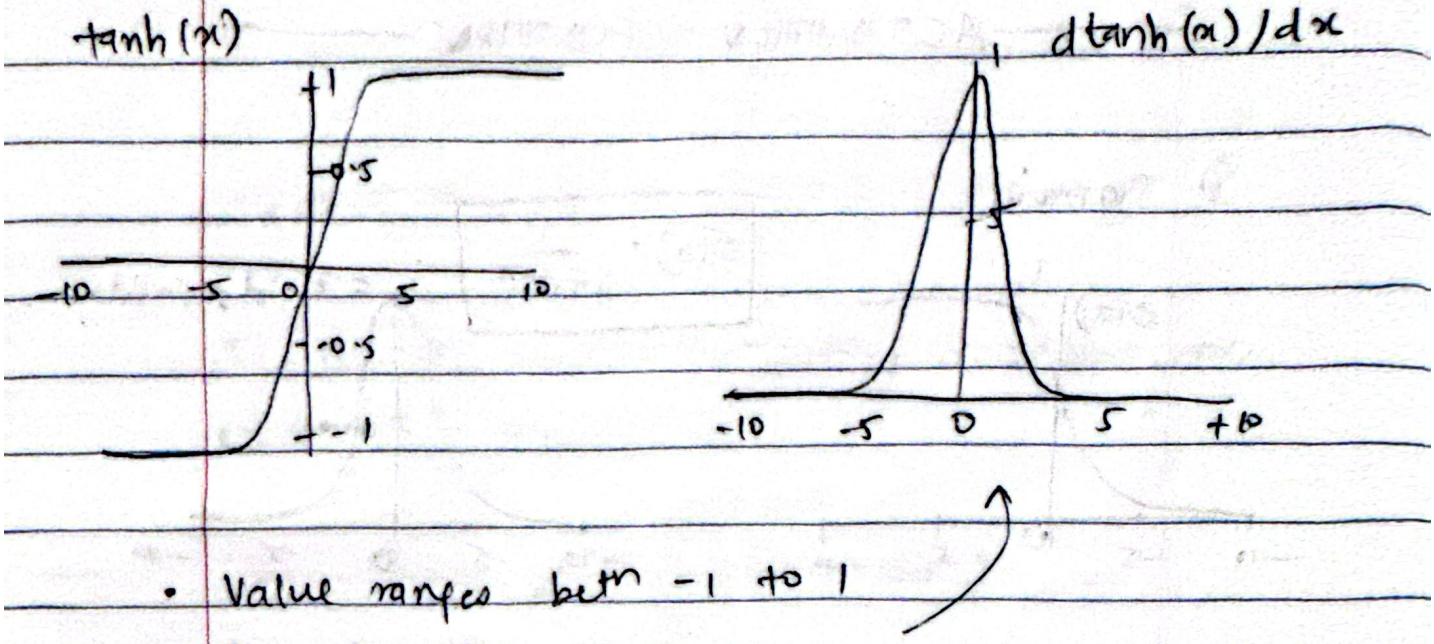
- Smooth gradient
- O/P values betn 0 & 1
- Clear predictions ie very close to 1 or 0

→ Disadv

- Prone to gradient vanishing
- Funct^n o/p is not zero centered
- Power operations are relatively time consuming

② tanh Function (Hyperbolic tangent function)

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



- Value ranges betn -1 to 1
- In derivative 0 to 1

The output interval of tanh is 1) and whole function is 0-centric, which is better than sigmoid.

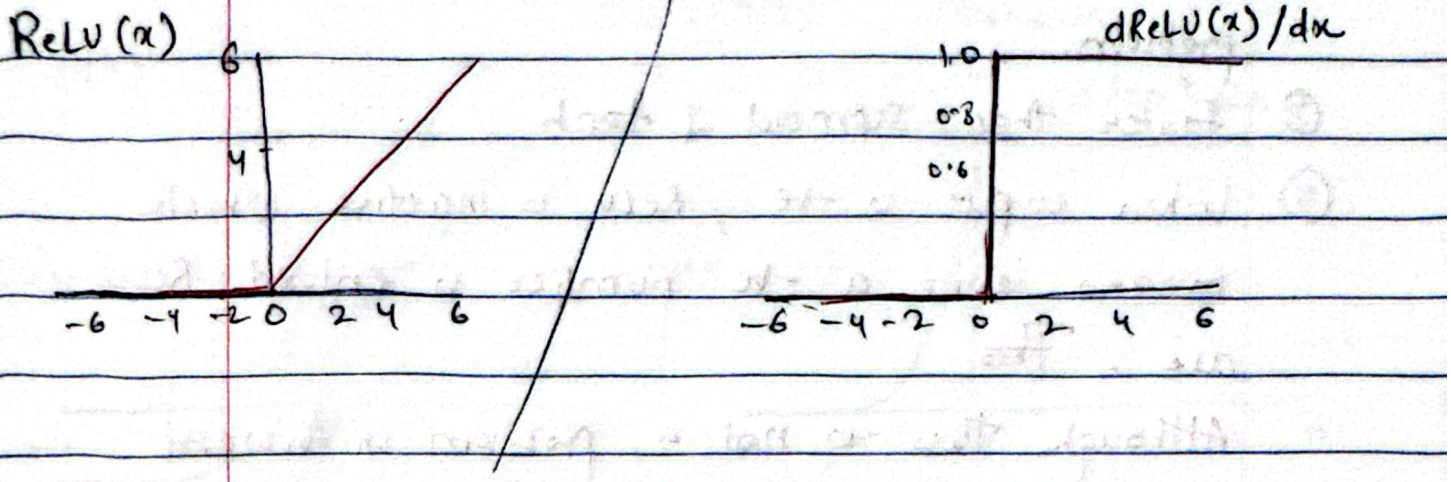
In binary classifi we used tanh for hidden layers & sigmoid for output layer

There will still be vanishing gradient problem

Solving vanishing gradient problem.

(3) Relu (Rectified Linear Unit)

$$ReLU = \max(0, x)$$ → quickest



- either it will be 0, or ~~whatever~~ wherever the x is 1 then you will get 1, 2 then 2
- derivative of Relu value is 0 or 1.

- Disadv

~~HL1~~ ~~ded~~ if the derivative is 0, ~~then~~ for eg. HL1 neuron then it will bcome dead during back propapatn

0

0

$\frac{\partial L}{\partial w_{old}} = 0$ then, $w_{new} \approx w_{old}$

- Other then this ReLU is exceptonal

~~Solu~~

- ReLU funct^n takes maximum value

① when input is +ve, there is no gradient saturat^n problem.

② faster then sigmoid & tanh

③ When input is -ve, ReLU is inactive which means once a -ve number is entered ReLU will die. ~~The~~

Although this is not a problem in forward propagat^n

In order to solve this dead Neuron problem we have ↓

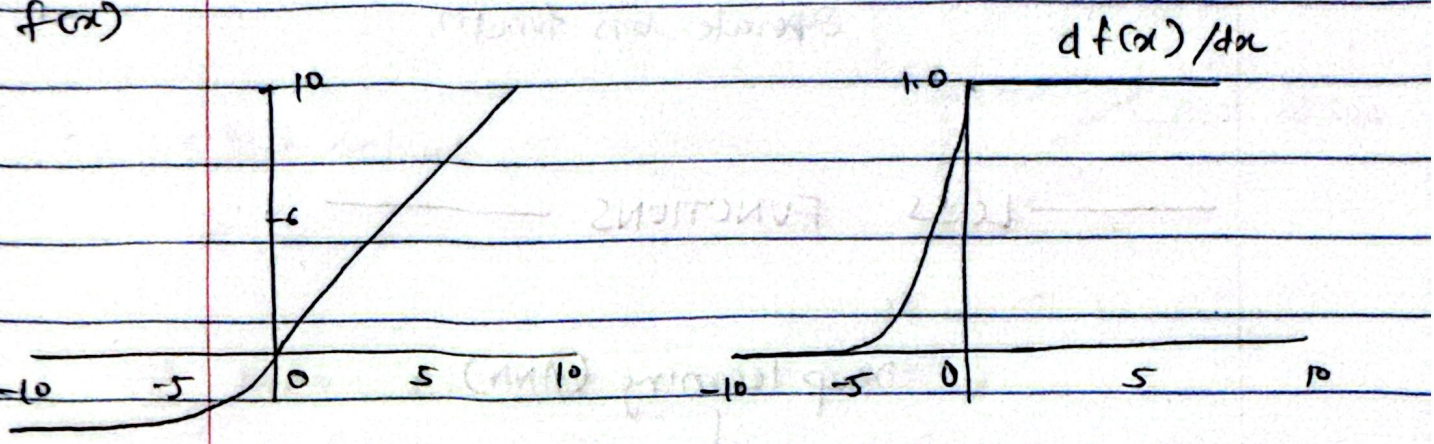
④ Leaky ReLU function

$$f(x) = \max(0.01x, x)$$

Solve dead Neuron problem.

⑤ ELU (Exponential Linear Units) functions

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha(e^x - 1), & \text{otherwise.} \end{cases}$$



- Computational expensive

⟶ Technique which Acti funct^n to use.

① Binary classification
Hidden layer → ReLU → Prelu, leaky
O/P layer → Sigmoid

② multiclass classificatn problems
O/P → Softmax Activat^n
Hidden → ReLU

③ Regression

Hidden → ReLU or variant of ReLU

o/p → Linear Activatn

↑

Seperate bcos functn