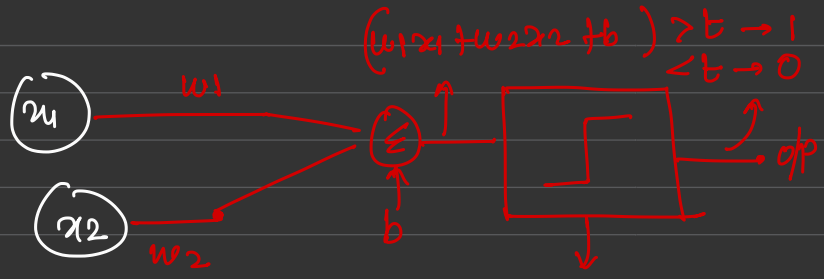
Activation Functions

XOR gate

A	B	o/p
0	0	0
1	0	1
0	1	1
1	1	0

Perceptron couldn't train the logic on the XOR



Activation function ⟶ thresholding function

Cannot capture non-linear patterns

Activation Functions:

1. Sigmoid
2. Relu (Rectified Linear Unit)
3. Leaky Relu
4. tanh (hyperbolic tangent)
5. Softmax
6. Parameteric Relu

Sigmoid activation: Binary classification Problem

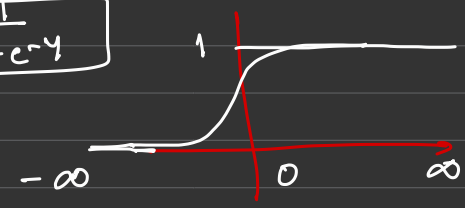
Squishing function

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

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

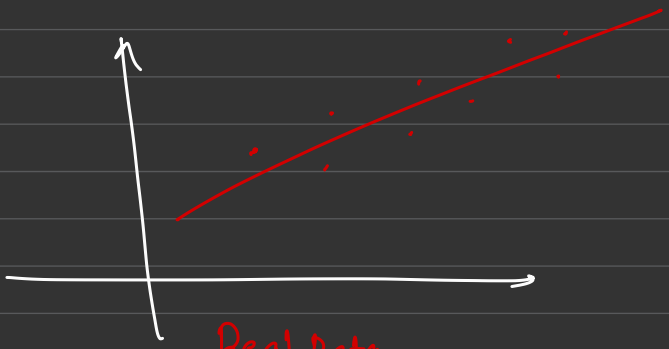
$$\sigma(y) = \frac{1}{1+e^{-y}}$$

0 - 1

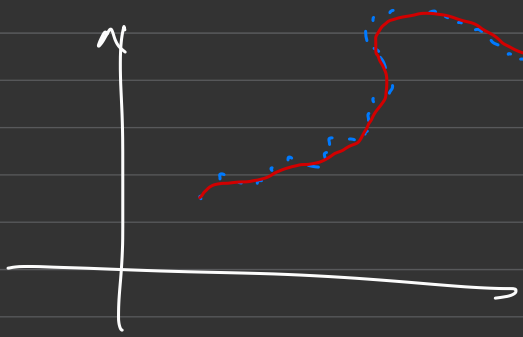


$-\infty \leq x \leq \infty$

$0 < \sigma(x) < 1$



Real Data



derivative of the sigmoid

$w_{new} = w_{old} - \alpha \, dJ/dw$

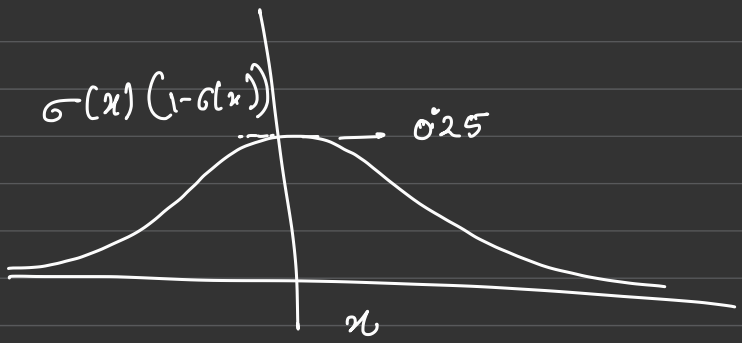
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

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

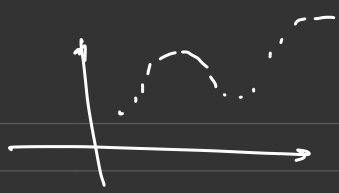
$$\sigma'(x) = (1+e^{-x})^{-2} \times (0 + e^{-x}) \times (-1)$$

$$\sigma'(x) = -(1+e^{-x})^{-2} \times e^{-x}$$

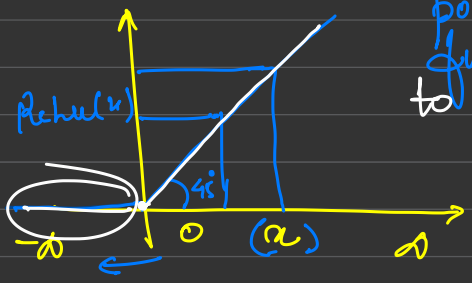
$$\sigma'(x) = \sigma(x) \times (1 - \sigma(x))$$



Relu → Rectified linear unit



$Relu(x) = max(x, 0)$ { $x \in (-\infty, \infty)$, $Relu \in (0, \infty)$



popularly used as an activation function in multi layer perceptrons to bring in non-linearity.

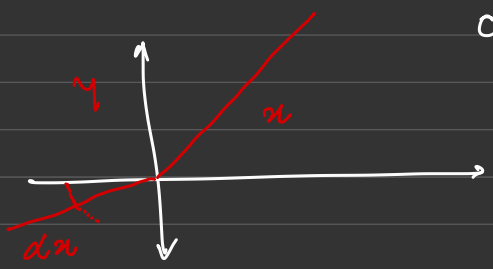
Parametric Relus:

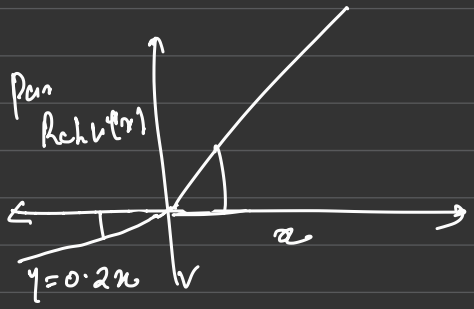
Relu $(x) \rightarrow \begin{cases} x & x \geq 0 \\ 0 & x = 0 \\ \alpha x & x < 0 \end{cases}$

$0 < \alpha < 1$

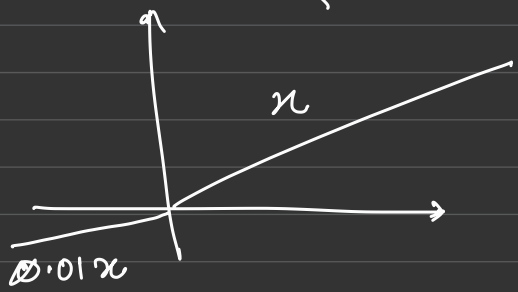
↓

learnt by the model.

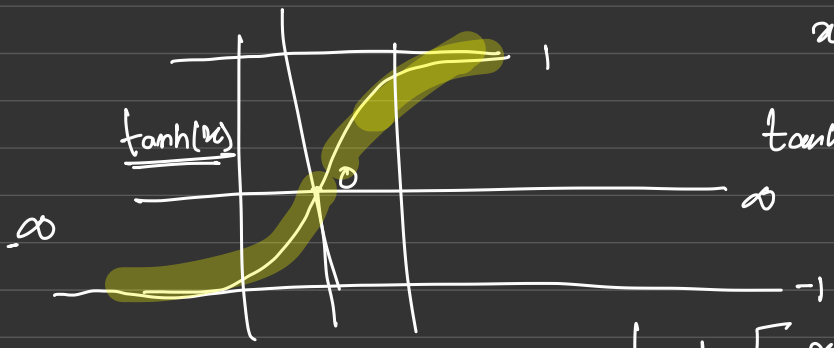




Leaky Relu(x) $\begin{cases} x & x > 0 \\ 0 & x = 0 \\ 0.01x & x < 0 \end{cases}$ → Leaky Relu



Hyperbolic Tangent (Tanh)

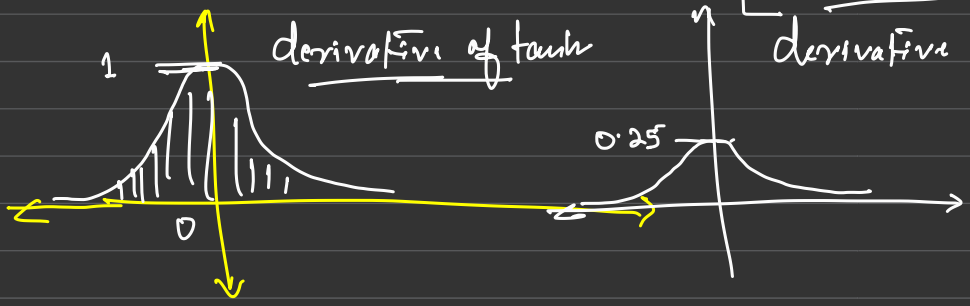


$x \in (-\infty, \infty)$

$\tanh(x) \in (-1, 1)$

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

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

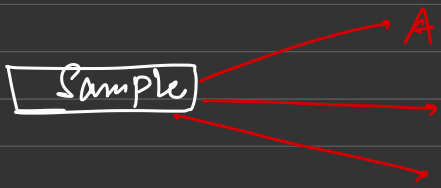


tanh is an activation function which is popularly used in the NN for creating non-linear features

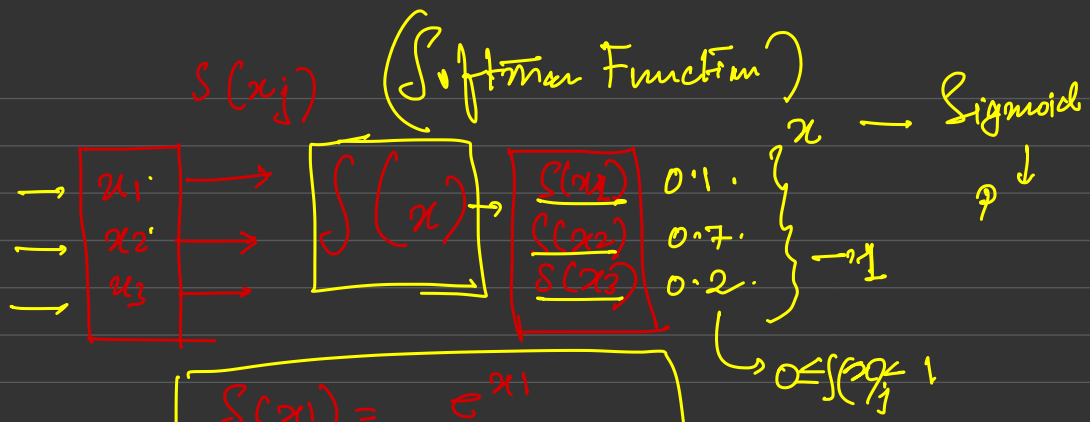
Softmax activation function

Special case of activation function

Multiclass classification problem



Softmax function
basically a more generalized form of sigmoid

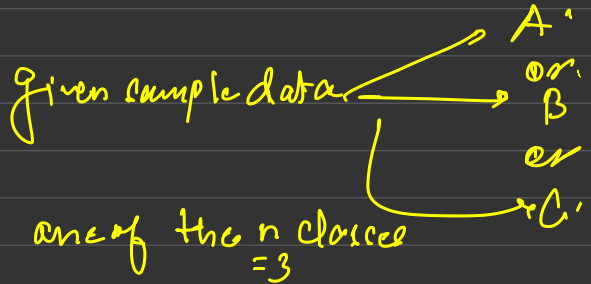


$$S(x_1) = \frac{e^{x_1}}{e^{x_1} + e^{x_2} + e^{x_3}}$$

$$S(x_2) = \frac{e^{x_2}}{e^{x_1} + e^{x_2} + e^{x_3}}$$

$$S(x_3) = \frac{e^{x_3}}{e^{x_1} + e^{x_2} + e^{x_3}}$$

$$S(x_j) = \frac{e^{x_j}}{\sum_{i=1}^{n} e^{x_i}}$$



Output in a multiclass classification problem