Activation Function

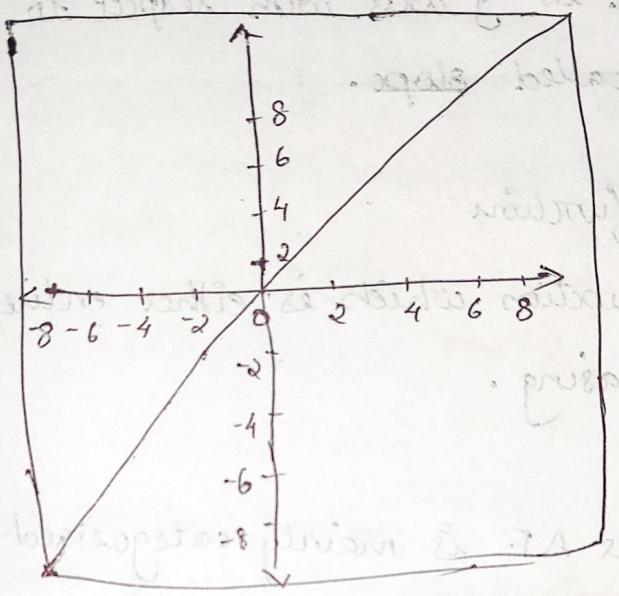
A function used to determine whether the output of a neuron either in an yes/no format. It is also called transfer function.

It decides whether a neuron should be activated or not by calculating weighted sum and further adding bias with it.

→ why we use activation function?

* It is used to determine the o/p of neural network like yes/no [1/0].
* It maps the resulting value in btw 0 to 1 or from -1 to 1 depending on the activation function.
* The purpose of activation function is to introduce non linearity into the o/p of a neuron.
* Neural n/w has neurons that work in correspondence of weight, bias and the respective activation fun^n.
* update the weights and bias of the neurons on the basis of error at the o/p. This process is known as back propogation.
* Activation function make the back propogation possible
* Activation fun^n can be basically divided into 2 types:
  1) Linear or Identity AF
  2) Non linear AF

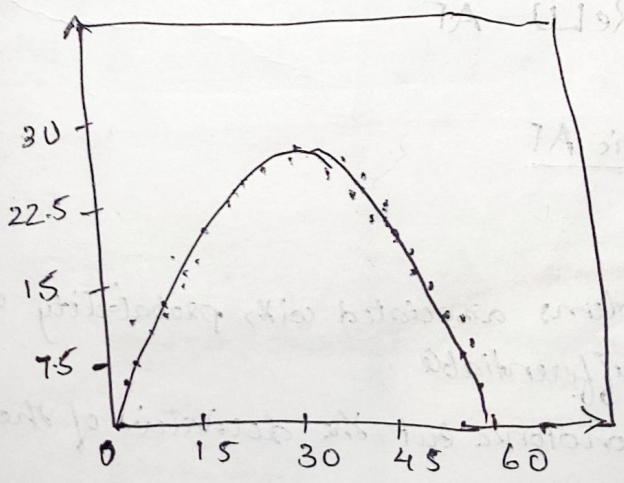
Linear AF : This fun^n is a line or linear. Therefore the o/p of the fun^n will not be confined (restricted) b/w any range $(-\infty, \infty)$.



Range = $-\infty$ to $+\infty$

$f(x) = x$

Non linear AF : These are the most used AF. It makes easy for the model to generalize or adapt with variety of data values and to differentiate b/w o/ps. It focuses on solving complex tasks.



Terminologies associated with non linear functions

1) Derivative or Differential

change in y axis with respect to x axis.
It is also called slope.

2) Monotonic function

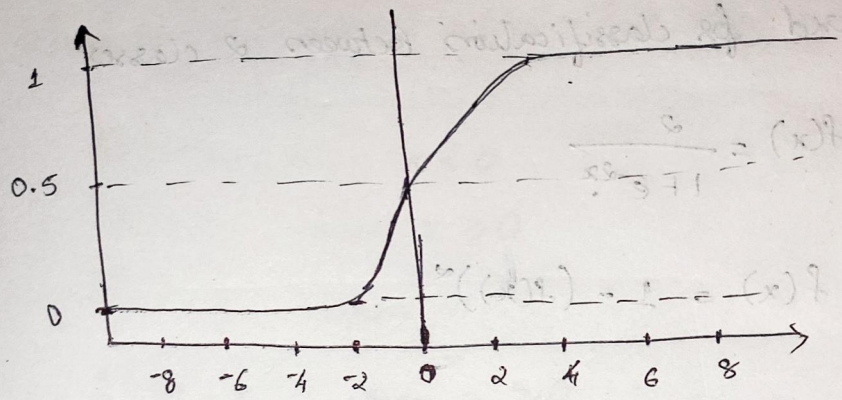
A function which is either entirely non increasing or non decreasing.

The non linear AF is mainly categorised on the basis of the range or curve.

1) Sigmoid or logistic AF
2) Tanh or hyperbolic tangent AF
3) ReLU (Rectify Linear Unit) AF
4) Leaky ReLU AF

⇒ Sigmoid or Logistic AF

→ S shaped curve
→ It is used in problems associated with probability of an o/p
→ The function is differentiable.
→ The function is monotonic but the derivative of the function is not monotonic
→ Softmax activation function is used.



equation : $f(x) = \frac{1}{1+e^{-x}}$

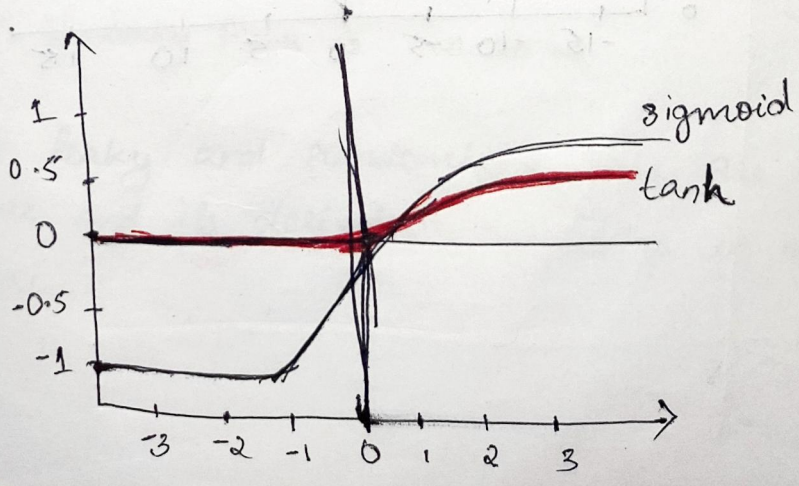
derivative : $f'(x) = f(x)(1-f(x))$

Softmax

→ sigmoid activation function
→ It is used for multiclass classification
→ range from 0 to 1.

⇒ Tanh

→ sigmoid activation function
→ It is ~~use~~ S shaped
→ differentiable
→ range from -1 to 1
→ function is monotonic while derivative is not monotonic.



Tan h is used for classification between 2 classes

$$f(x) = \frac{2}{1+e^{-2x}}$$

$$f(x) = 1 - (f(x))^2$$

Both Sigmoid and tan h is used in feed forward networks.

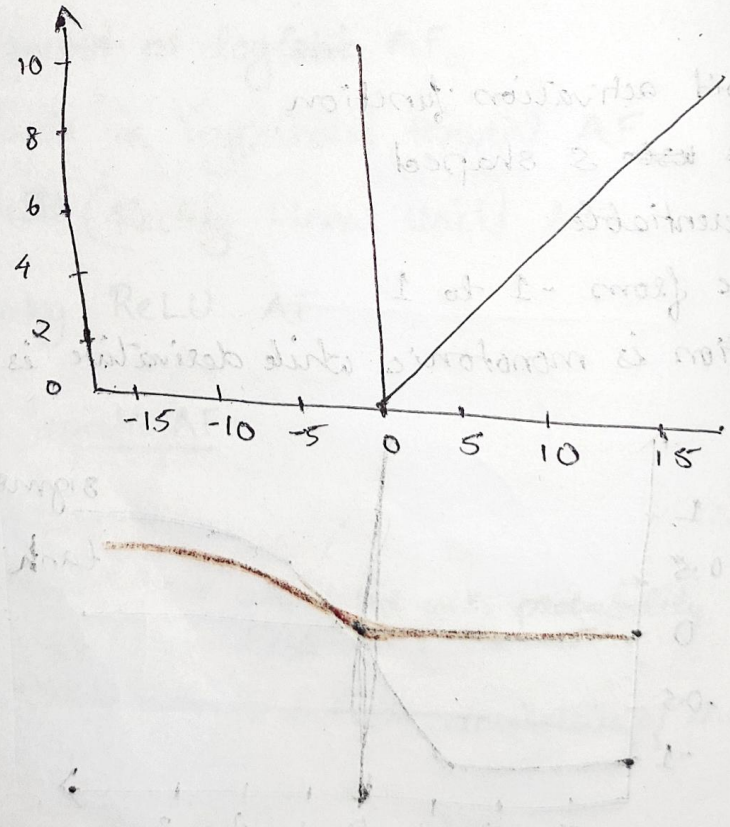
⇒ **ReLu**

→ Most commonly used

→ It is used in CNN (convolutional neural n/w)

→ Range from 0 to ∞

→ Function and its derivative, both are monotonic.



⇒ half rectified from bottom (-ve part not included)

$f(x) = 0$ when $x < 0$

$f(x) = x$ when $x \geq 0$

$f'(x) = 0$ when $x < 0$

$f'(x) = 1$, when $x \geq 0$

⇒ Leaky ReLu

→ Adding an extra parameter to ReLu

→ All the negative values become zero immediately in the graph which in turn affect the resulting graph but not mapping negative values appropriately.

⇒ To solve this problem leaky ReLu was used by adding an extra parameter to increase the range of the ReLu function.

→ Usually the value of the parameter a is 0.01.

→ If the value of 'a' is not 0.01, it is called randomised ReLu.

⇒ Range of leaky ReLu is $-\infty$ to ∞.

→ Both leaky and randomized ReLu are monotonic in nature and its derivative are also in monotonic in nature.

Equation :

Leaky ReLu

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

$$f'(x) = \begin{cases} \alpha & ; x < 0 \\ 1 & ; x \geq 0 \end{cases}$$