What are Activation Functions?

A **neural network** activation function is a function that introduces nonlinearity into the model. A neural network without an activation function is just a linear regression model. The activation function does the non-linear transformation to the input, making it capable of learning and performing more complex tasks.

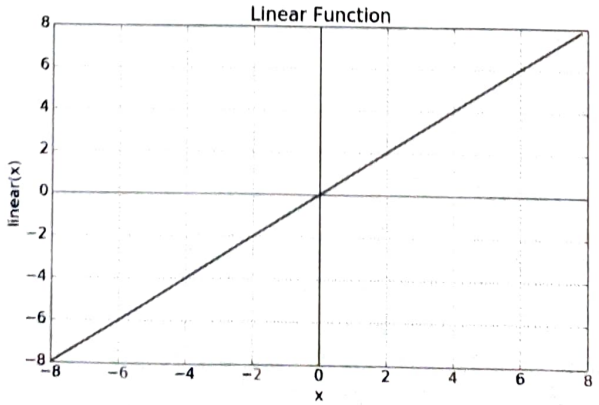
Why use activation functions?

1. Activation functions' main objective is to add non-linearities into the network so that it can model more intricate and varied interactions between inputs and outputs. In the absence of activation functions, the network would only be capable of performing linear transformations, which cannot adequately represent the complexity and nuances of real-world data. Since neural networks need to implement complex mapping functions, non-linear activation functions must be used to introduce the much-needed nonlinearity property that allows approximating any function.

2. Normalizing each neuron in the network's output is a key benefit of utilizing activation functions. Depending on the inputs it gets and the weights associated to those inputs, a neuron's output can range from extremely high to extremely low. Activation functions make ensuring that each neuron's output falls inside a defined range, which makes it simpler to optimise the network during training.

Types of Activation Functions:

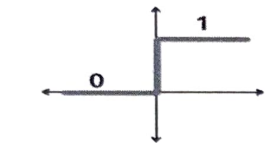
1. Linear Function/Identity Function



In the Linear process, activation is directly proportional to the input.

2. **Binary step function:**

This function is used in single layer networks to convert the net input to output. The output is binary i.e. 0 or 1. The t represents the threshold value.

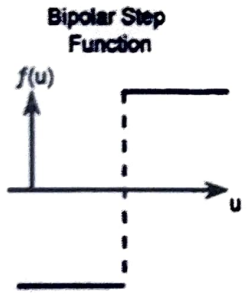


Binary Step Activation Function

$$f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

3. **Bipolar step function:**

In the Bipolar Step Function, if the value of Y is above a certain value known as the threshold, the output is +1and if it's less than the threshold then the output is -1. It has bipolar outputs (+1 to -1). It can be utilized in single-layer networks.

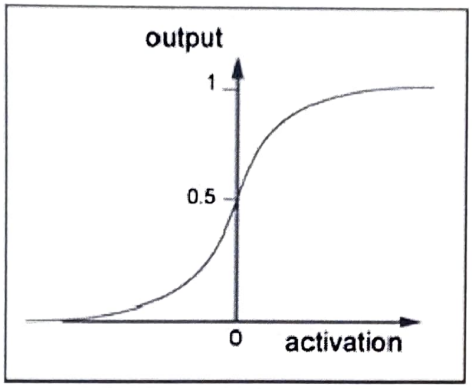


$$f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

4. **Sigmoidal Functions:**

It is also called S-shaped functions. Logistic and hyperbolic tangent functions are commonly used in sigmoid functions. There are two types of sigmoid functions.

- Binary Sigmoid function (or Logistic function)
  Binary Sigmoid Function or Sigmoid function is a logistic function where the output values are either binary or vary from 0 to 1. It is differentiable, non-linear, and produces non-binary activations.

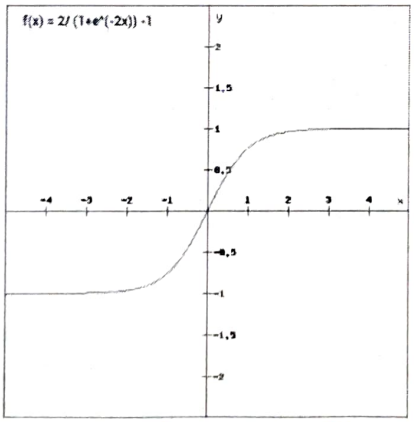


$$f(x) = \frac{1}{1+e^{-\lambda x}}$$

λ = steepness parameter

$$\therefore \text{Derivative: } f'(x) = f(x)[1-f(x)]$$

- Bipolar Sigmoid function (or Hyperbolic Tangent Function or Tanh):
  Hyperbolic Tangent Function or Tanh is a logistic function where the output value varies from -1 to 1. Also known as Bipolar Sigmoid Function. The output of Tanh centers around 0 and sigmoid's around 0.5. Tanh Convergence is usually faster if the average of each input variable over the training set is close to zero.

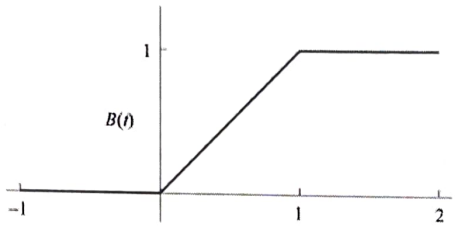


$$f(x) = \frac{1-e^{-\lambda x}}{1+e^{-\lambda x}}$$

$$\therefore \text{Derivative: } f'(x) = [1+f(x)][1-f(x)]$$

5. Ramp Function:

ReLu stands for the rectified linear unit (ReLU). It is the most used activation function in the world. It output 0 for negative values of x. This is also known as a ramp function. The name of the **ramp function** is derived from the appearance of its graph.



$$f(x) = \begin{cases} 1 & \text{if } x > 1 \\ 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \end{cases}$$

How to choose activation functions?

Consideration	Activation Function
Non-linearity	Sigmoid, Tanh, ReLU, Leaky ReLU, ELU, SELU
Derivability	Sigmoid, Tanh, ReLU, Leaky ReLU, ELU, SELU
Range of output values	Sigmoid, Softmax
Computational efficiency	ReLU, Leaky ReLU, ELU, SELU
Saturation	ReLU, Leaky ReLU, ELU, SELU