https://www.youtube.com/watch?v=gmjzbpSVY1A

Part 5 :

Y = sum(weight * input) + bias

the value of Y can be anything ranging from +infinity and -infinity. (linearity) The neuron really doesn't know the bounds of value. We decided to add activation functions for this purpose. Activation functions are decision making units of neural networks. They calculates net output of a neural node. Every layer has activation function.

Activation functions basically divided into 2 types :

- Linear Activation Function
- Non-linear Activation Function

Linear Activation Function :

f(x) = x
Range (-inf to +inf)

Step function is one of the most common activation function in neural networks. The function produces binary output. Step function can implemented primitive neural networks and it can separate linearly. Example : AND gate, OR gate. A = 1 if y> threshold, 0 otherwise.

Non-linear Activation Function :

It makes it easy for the model to generalize or adapt with variety of data and to differentiate between output.

The main terminologies needed to understand for nonlinear functions are:

Derivative or Differential: *Change in y-axis with respect to change in x-axis. It is also known as slope.*

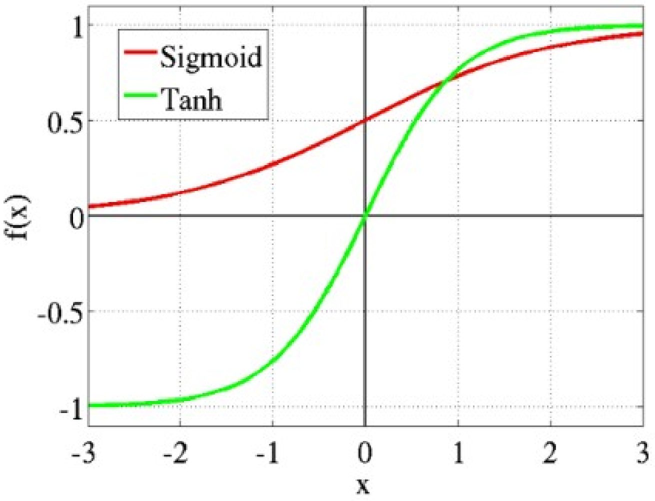
Monotonic function: *A function which is either entirely non-increasing or non-decreasing.*

Sigmoid or Logistic Activation Function :

The reason why we use sigmoid function is because it exist between 0 to 1. We are generally using this when we need to find probability as output. We can find the slope of the sigmoid curve at any points.

Tanh or hyperbolic tangent Activation Function

Tanh is also like logistic sigmoid but better. Range (-1 to 1), tanh is mainly used classification between two classes.



ReLU (Rectified Linear Unit) Activation Function

The Rectified Linear Unit Activation Function (ReLU) is the most used activation function in the world. Range (0 to infinity). The function and its derivative **both are monotonic**. But the issue is that all the negative values become zero immediately which decreases the ability of the model to fit or train from the data properly. So basically it doesn't map the negative values appropriately.

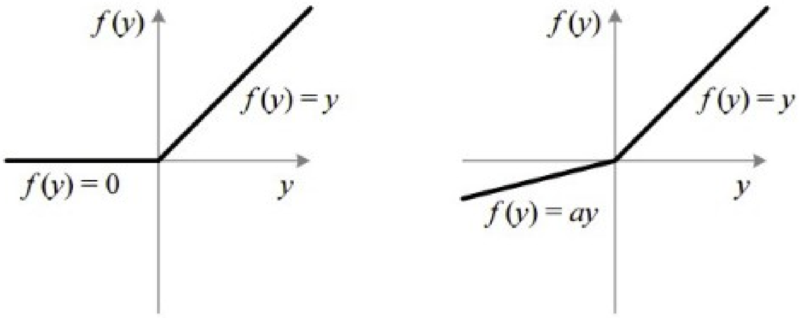


Fig : ReLU v/s Leaky ReLU