**Activation Functions in Neural Networks**

## https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

## **1. Sigmoid or Logistic Activation Function**

The Sigmoid Function curve looks like a S-shape.



The main reason why we use sigmoid function is because it exists between **(0 to 1).**Therefore, it is especially used for models where we have to **predict the probability** as an output.Since probability of anything exists only between the range of **0 and 1,** sigmoid is the right choice.

The function is **differentiable**.That means, we can find the slope of the sigmoid curve at any two points.

The function is **monotonic**but function’s derivative is not.

The logistic sigmoid function can cause a neural network to get stuck at the training time.

## **2. Tanh or hyperbolic tangent Activation Function**

tanh is also like logistic sigmoid but better. The range of the tanh function is from (-1 to 1). tanh is also sigmoidal (s - shaped).



The advantage is that the negative inputs will be mapped strongly negative and the zero inputs will be mapped near zero in the tanh graph.

The function is **differentiable**.

The function is **monotonic** while its **derivative is not monotonic**.

The tanh function is mainly used classification between two classes.

## **3. ReLU (Rectified Linear Unit) Activation Function**

The ReLU is the most used activation function in the world right now.Since, it is used in almost all the convolutional neural networks or deep learning.



As you can see, the ReLU is half rectified (from bottom). f(z) is zero when z is less than zero and f(z) is equal to z when z is above or equal to zero.

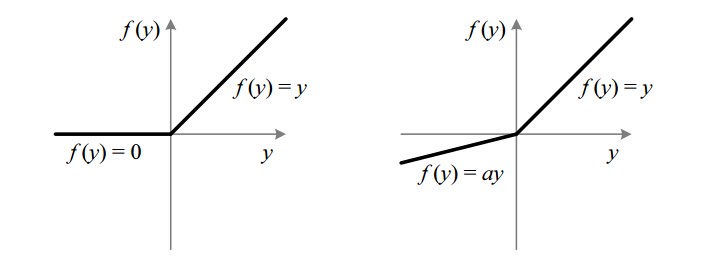
**Range:**[ 0 to infinity)

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. That means any negative input given to the ReLU activation function turns the value into zero immediately in the graph, which in turns affects the resulting graph by not mapping the negative values appropriately.

## **4. Leaky ReLU**

It is an attempt to solve the dying ReLU problem



The leak helps to increase the range of the ReLU function. Usually, the value of **a**is 0.01 or so.

When **a is not 0.01** then it is called **Randomized ReLU**.

Therefore the **range** of the Leaky ReLU is (-infinity to infinity).

Both Leaky and Randomized ReLU functions are monotonic in nature. Also, their derivatives also monotonic in nature.