Activation Function

- Activation function decides, whether a neuron should be activated or not by calculating weighted sum and further adding bias with it. The purpose of the activation function is to introduce non-linearity into the output of a neuron.

Explanation :-

We know, neural network has neurons that work in correspondence of weight, bias and their respective activation function. In a neural network, we would update the weights and biases of the neurons on the basis of the error at the output. This process is known as back-propagation. Activation functions make the back-propagation possible since the gradients are supplied along with the error to update the weights and biases.

**Why do we need Non-linear activation functions :-**
A neural network without an activation function is essentially just a linear regression model. The activation function does the non-linear transformation to the input making it capable to learn and perform more complex tasks.

There are several type of Activation But here i dicuss some of them:

```
1. Sigmoid (Binary Classification)
2. Tanh
3. Relu
4. Leaky Relu
5. Linear
6. Softmax (Use Multiclass Classification)
```
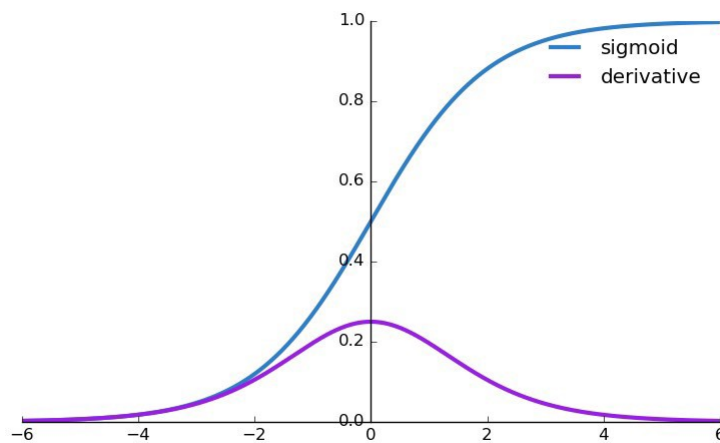
1. Sigmoid

- The Sigmoid Function curve looks like a S-shape.

$$\phi(z) = \frac{1}{1 + e^{-z}}$$
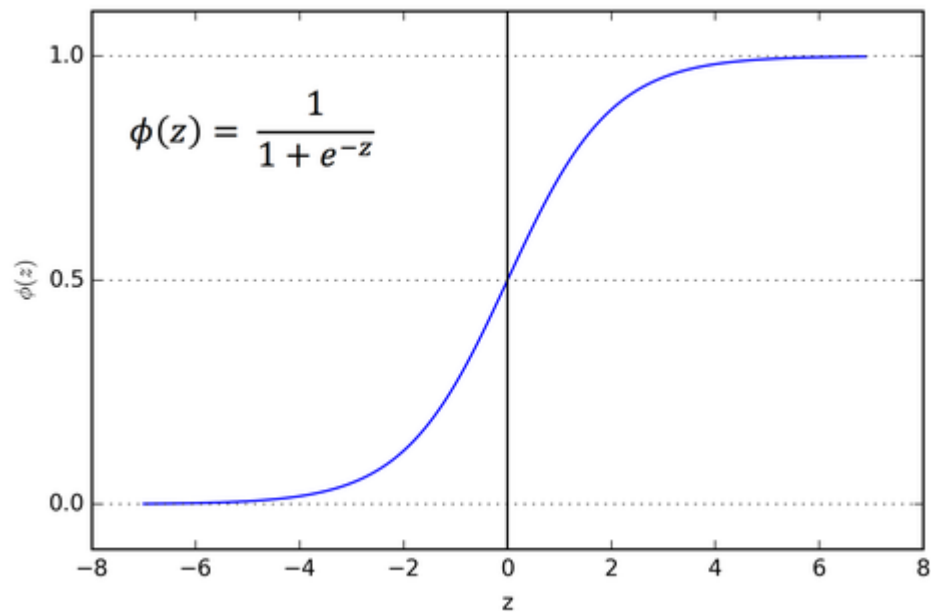
- The main reason why we use sigmoid function is because it exists between (0 to 1). Therefore, it is especially used for models where we have to predict the probability as an output.Since probability of anything exists only between the range of 0 and 1, sigmoid is the right choice.

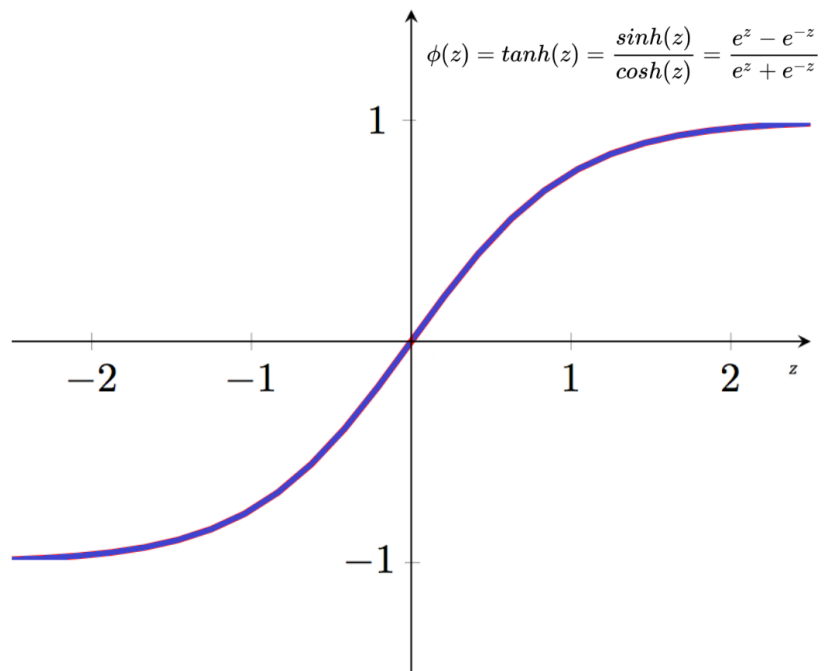**Derivative Of Sigmoid Function**

- Sigmoid Function Derivative range from 0 to 0.25

Uses : Usually used in output layer of a binary classification, where result is either 0 or 1, as value for sigmoid function lies between 0 and 1 only so, result can be predicted easily to be 1 if value is greater than 0.5 and 0 otherwise.

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).

$$\phi(z) = tanh(z) = \frac{sinh(z)}{cosh(z)} = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

- 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 activation that works almost always better than sigmoid function is Tanh function also knows as Tangent Hyperbolic function. It's actually mathematically shifted version of the sigmoid function. Both are similar and can be derived from each other.
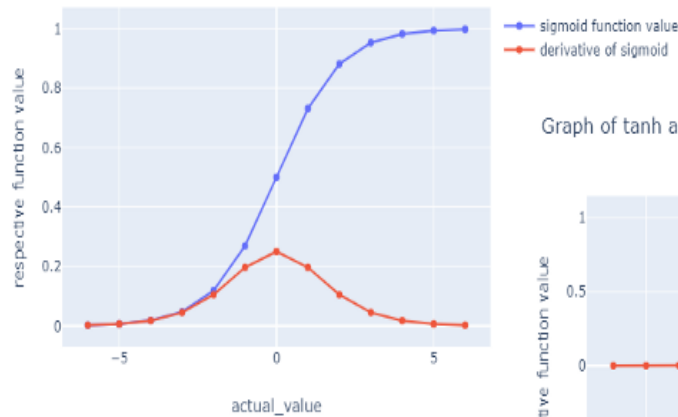
Derivative Of Tanh Function:-
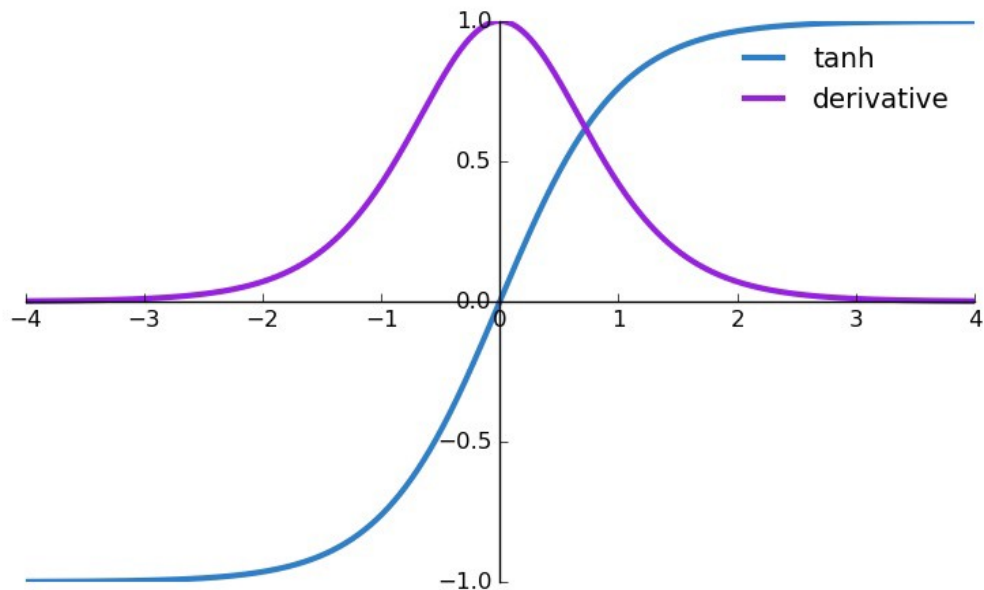
Sigmoid Function Derivative range from 0 to 1

Uses :- Usually used in hidden layers of a neural network as it's values lies between -1 to 1 hence the mean for the hidden layer comes out be 0 or very close to it, hence helps in centering the data by bringing mean close to 0. This makes learning for the next layer much easier.

Graph of Sigmoid and it's derivative

Graph of tanh and it's derivative
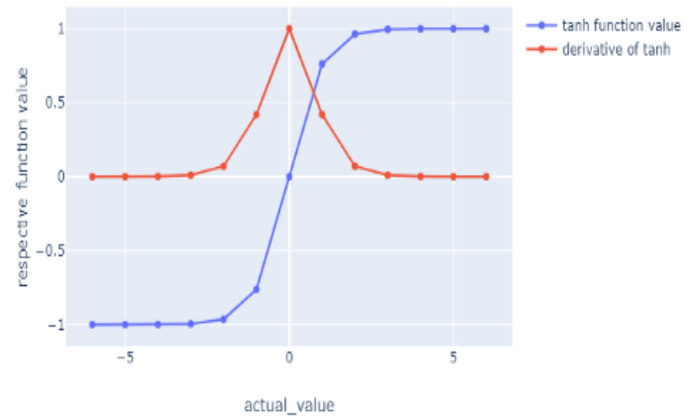
Point :-

1. tanh and logistic sigmoid are the most popular activation functions in 90's but because of their Vanishing gradient problem and sometimes Exploding gradient problem (because of weights), they aren't mostly used now.
2. These days Relu activation function is widely used. Even though, it sometimes gets into vanishing gradient problem, variants of Relu help solving such cases.
3. tanh is preferred to sigmoid for faster convergence BUT again, this might change based on data. Data will also play an important role in deciding which activation function is best to choose.