

REPORT

Activation functions in Neural Networks:

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

Types of Activation Function:

- I. **Linear Function:** Linear function has the equation similar to as of a straight line $y = x$
No matter how many layers we have, if all are linear in nature, the final activation function of last layer is nothing but just a linear function of the input of first layer. Linear activation function is used at just one place – the output layer
- II. **Sigmoid Function:** It is a function which is plotted as ‘S’ shaped graph.
Equation is $A = 1/(1 + e^{-x})$
Non-linear. Notice that X values lies between -2 to 2, Y values are very steep. This means, small changes in x would also bring about large changes in the value of Y.
Usually it is 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.
- III. **Tanh Function :** The activation that works almost always better than sigmoid function is Tanh function also known as Tangent Hyperbolic function. It's actually mathematically shifted version of the sigmoid function. Both are similar and can be derived from each other. It's a non-linear function.

Usually it is used in hidden layers of a neural network as its values lie between -1 to 1 hence the mean for the hidden layer comes out to 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.

- IV. **RELU Function:** It stands for *Rectified linear unit*. It is the most widely used activation function. Chiefly implemented in *hidden layers* of Neural network.

Equation - $A(x) = \max(0, x)$. It gives an output x if x is positive and 0 otherwise. It is also a non-linear function

ReLU is less computationally expensive than tanh and sigmoid because it involves simpler mathematical operations. At a time only a few neurons are activated making the network sparse making it efficient and easy for computation.

- V. **Softmax Function :** The softmax function is also a type of sigmoid function but is handy when we are trying to handle multi- class classification problems.

Usually used when trying to handle multiple classes. the softmax function was commonly found in the output layer of image classification problems. The softmax function would squeeze the outputs for each class between 0 and 1 and would also divide by the sum of the outputs.