← Go Back to Deep Learning

Course Content

Activation Functions

Activation functions are introduced to learn complex patterns in the data. The activation function decides what is to be fired to the next neuron. It takes input from previous layers and converts it to some form of input for the next layers.

The most important feature of activation functions is to introduce non-linearity into a neural network because if there is no non-linearity it's just a linear algorithm or we can say the model just tries to fit a straight line but there are more complex patterns in data that cannot be identified by a linear algorithm.

We can use different types of activation functions as given below:

The Sigmoid Function: It is one of the most widely used non-linear activation functions. Sigmoid transforms the values into a range between 0 and 1. It can be interpreted as the probability of a particular class. The mathematical expression for sigmoid: $f(z) = \frac{1}{(1+e^{-z})}$

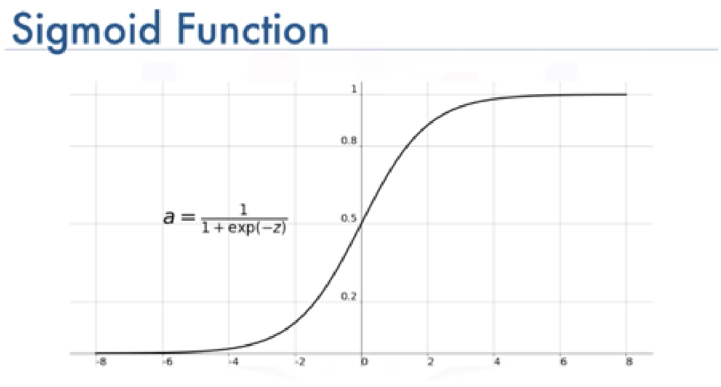

Image Source

The Tanh Function is very similar to the sigmoid function. The only difference is that it is symmetric around the origin. The range of values, in this case, is from -1 to 1. Thus the inputs to the next layers will not always be of the same sign. The tanh function is defined as:

$\tanh(z) = 2sigmoid(2z) - 1$ or $\tanh(z) = \frac{2}{(1+e^{-2z})} - 1$

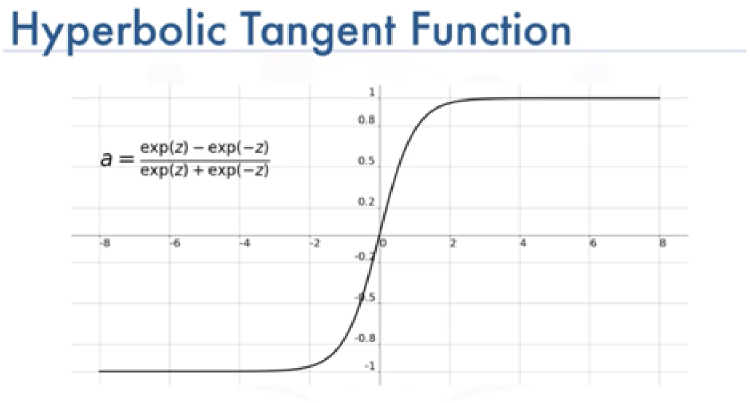

Image Source

The ReLU Function is another non-linear activation function that has gained popularity in deep learning. ReLU stands for Rectified Linear Unit. The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time. This means that the neurons will only be activated if the output of the linear transformation is greater than 0. The plot below will help you understand ReLU better: $f(z) = max(0,\ z)$

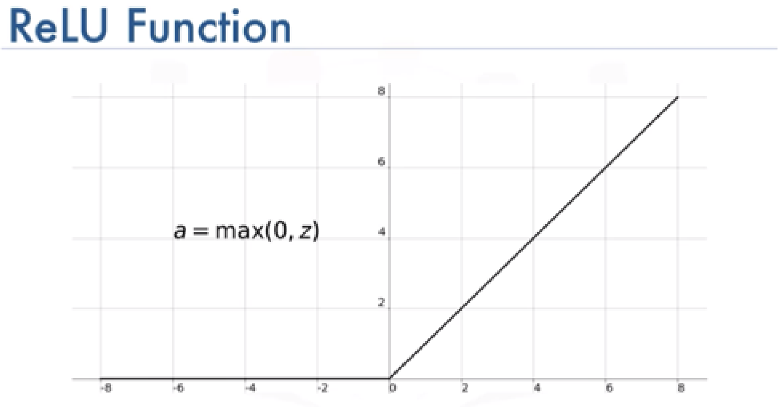

Image Source

The Softmax Function

The Softmax function returns the probability of each class. Here's the equation for the Softmax activation function:

$$softmax\,(z_i) = \frac{exp\,(z_i)}{\sum_j\ exp\,(z_j)}$$

Here, the Z represents the values from the neurons of the output layer. The exponential acts as the non-linear function. Later these values are divided by the sum of exponential values in order to normalize and convert them into probabilities.

❮ Previous

Next ❯

Help