1. Activation Function

The purpose of the activation function is to introduce non-linearity into the output of a neuron. When we use activation function to distinguished the output result which are not to be linearly separable are called Non Linearity. We update the weights of neurons on the basis of output loss or error rate, activation functions helps us to update that weights and biases. That whole process is called back-propagation.

Some common types of Activation function

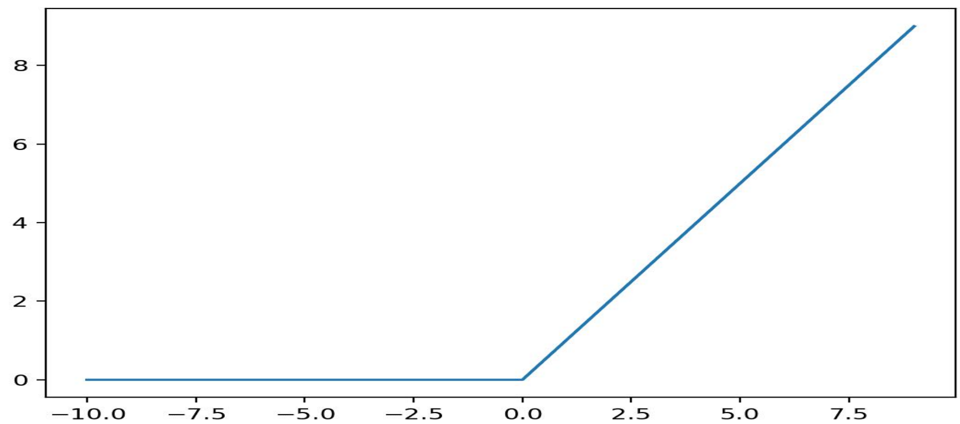
1. Rectified Linear Activation (ReLU)
2. Logistic (Sigmoid)
3. Hyperbolic Tangent (Tanh)

1. Rectified linear activation function (ReLU):

The ReLU function calculation formula is followed:

Max(0.0,x)

Where x is the input. It means it takes only select only positive number. If x is negative then it takes 0.0 otherwise x is returned.



2. Sigmoid

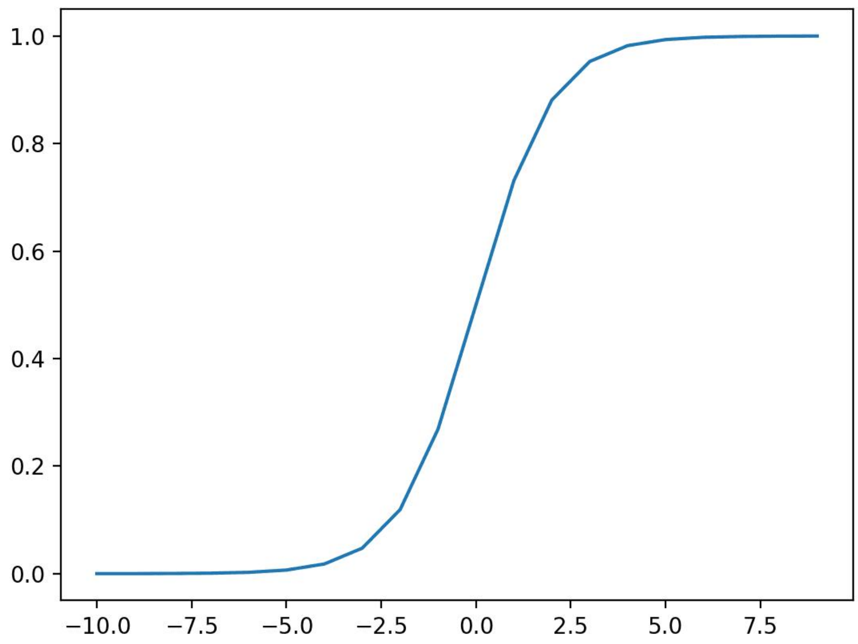
We can calculate sigmoid activation function by following formula:

1.0 / (1.0 + e^-x)

Input: Any Real Number

Range: (0,1)

The range is between 0 to 1 , you can put any larger number that means it goes closer to 1. For smaller number it goes near to 0. But it remain between 0 and 1.



3. The hyperbolic tangent activation function (Tanh):

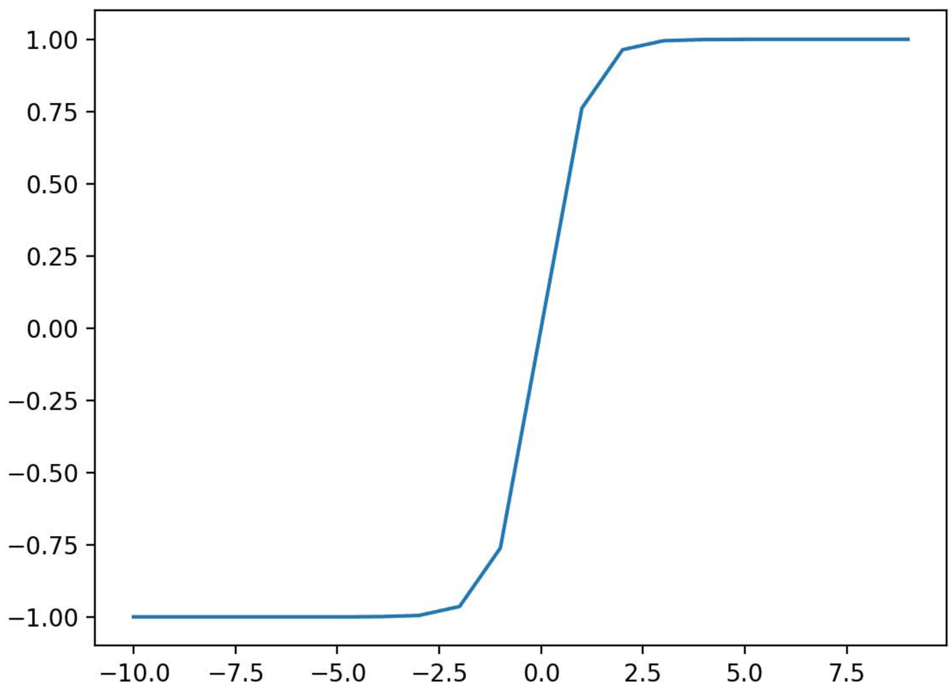
The formula for calculating Tanh is as following:

(e^x – e^-x) / (e^x + e^-x)

We can gave any real value as an Input.

The Output range is between -1 to 1.

This means the larger the input takes us closer to the +1 and the smallest the input takes us closer to -1.



Q: How to chose which activation function is good?

It depend on on architecture of Neural Network. It is cleared from

a) For Multilayer Perceptron (MLP):

ReLU activation function.

b) For Convolutional Neural Network (CNN):

ReLU activation function.

c) For Recurrent Neural Network (RNN):

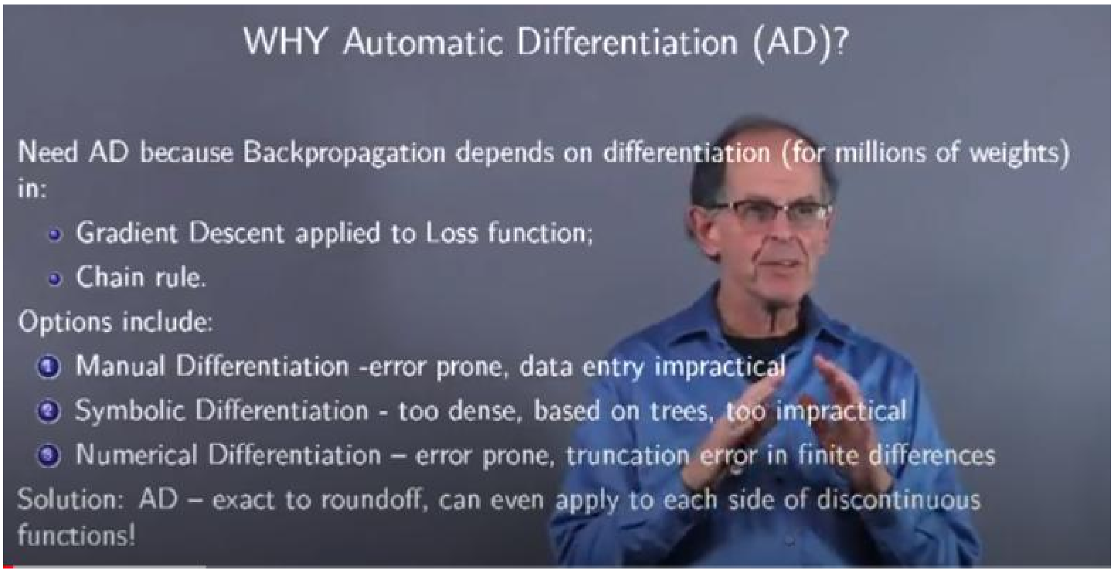
Tanh and/or Sigmoid activation function.

2. Dual numbers automatic differentiation

Automatic Differentiation (AD) is to calculate the derivative of outputs(loss) of a model with respect to its inputs via a series of methods. We need Automatic Differentation because back-propagation depends on differentiation (millions of weights) via chain rule.

A dual number is a multidimensional number. Where sensitivity of the function is propagated.

We used AD gives the exact round off solution by calculating derivatives without any numerical error.



3. Symbolic Differentiation

A symbolic differentiation means we have to find the derivative of given formula with respect to specified variable. Actually mean generating a new formula as output. Instead of adding values and proving the formula we have to use variable to produce a new formula this mean Symbolic Differentiation

4. Simple Derivatives

It's a continuous description of how function changes with a very small changes in other variables.

Basically, it's a instantaneous rate of change in any function with respect to variables is known as Derivatives

5. The Beta-Binomial model

Its a continuously updating the probability distribution of two parameters. It is the modeling the uncertainty about the probability of success of the given experiment.

If X is unknown, assume the value of X is between 0 and 1. Then the uniform distribution of all points means they are equally likely. So the probability of all data points are same.