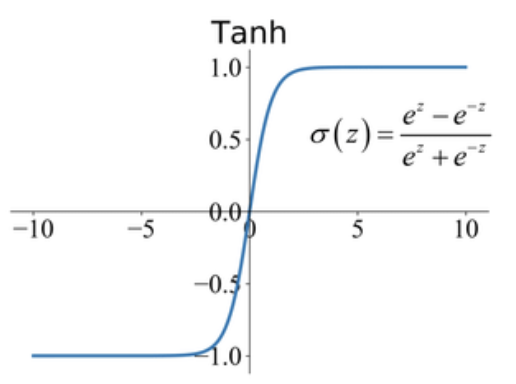
Tanh Activation

<https://paperswithcode.com/method/tanh-activation>

**Tanh Activation** is an activation function used for neural networks:

Historically, the tanh function became preferred over the [sigmoid function](https://paperswithcode.com/method/sigmoid-activation) as it gave better performance for multi-layer neural networks. But it did not solve the vanishing gradient problem that sigmoids suffered, which was tackled more effectively with the introduction of [ReLU](https://paperswithcode.com/method/relu) activations.



<https://www.aitude.com/comparison-of-sigmoid-tanh-and-relu-activation-functions/>

#### **Vanishing Gradient Problem**

Vanishing gradient problem mostly occurs during the backpropagation when the value of the weights are changed. To understand the problem we will increase the value of the input values in the activation function, At that time we will notice that the predicted output is available on the range of the selected activation function and maintain the threshold value.

For the sigmoid function, the range is from 0 to 1. We know that the maximum threshold value is 1 and the minimum value is 0. So when we increase the input values, the predicted output must lie near to the upper threshold value which is 1. So the predicted output must be less than or near to the 1.

We again increasing the input value and the output comes on the max threshold value and lies there. When the neuron outputs are very small for example ( −1<output<1), the patterns are created during the optimization will be smaller and smaller towards the upper layers. This causes them to make the learning process very slow, and make them converge to their optimum and this problem is known as the Vanishing Gradient Problem.

Y=Activationfunction(∑(weights∗input+bias))

In the nutshell, a neural network is a very dominant method and technology in machine learning which mimics how a brain perceives and operates.

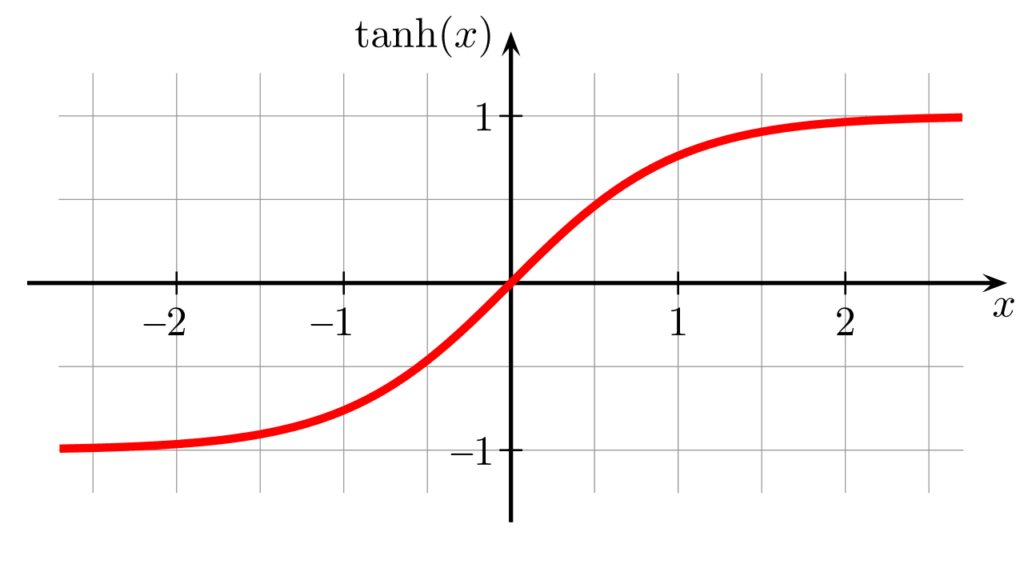
### **Hyperbolic Tangent Activation Function**

Tanh Activation function is superior then the Sigmoid Activation function because the range of this activation function is higher than the sigmoid activation function. This is the major difference between the Sigmoid and Tanh activation function. Rest functionality is the same as the sigmoid function like both can be used on the feed-forward network.

**Range** : -1 to 1

Equation can be created by:

*y*=*tanh*(*x*)



#### **Advantage of TanH Activation function**

Here negative values are also considered whereas in the sigmoid minimum range is 0 but in Tanh, the minimum range is -1. This is why the Tanh activation function is also known as the **zero centered activation function.**

#### **Disadvantage**

Also facing the same issue of Vanishing Gradient Problem like a sigmoid function.