Zummit Infolabs

Activation Functions in Neural Network

By Shivam Bhardwaj   April 16, 2021   No Comments

Deep neural networks have been successfully used in diverse emerging domains to solve real world complex problems with more deep learning(DL) architectures, being developed to date. To achieve these state-of-the-art performances, the DL architectures use activation functions (AFs), to perform diverse computations between the hidden layers and the output layers of any given DL architecture.

So, many times we are in confusion about which activation function is to be used. I had tried my best to clear all the doubts in this concise discussion.

Artificial Neural Networks & Activation Functions

The typical artificial neural networks (ANN) are biologically inspired computer programmes, designed by the inspiration of the workings of the human brain. These ANNs are called networks because they are composed of different functions, which gathers knowledge by detecting the relationships and patterns in data using past experiences known as training examples in most literature.
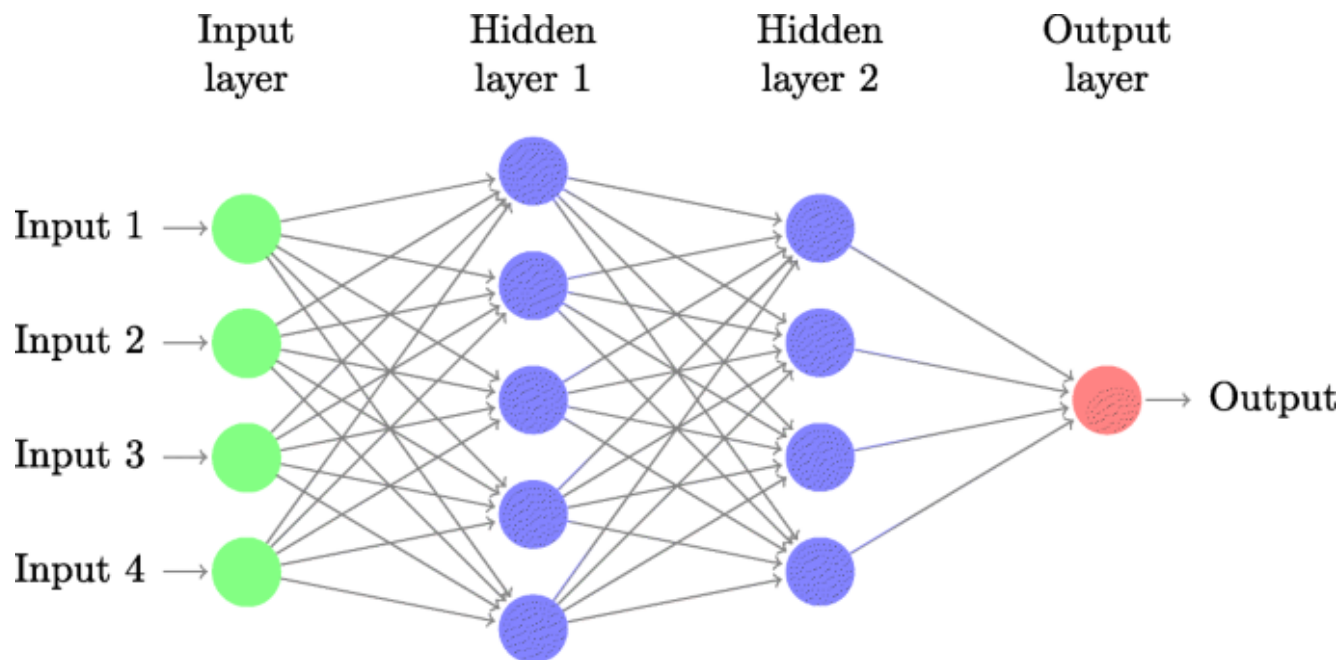
Basically, input features along with weights and biases are used to compute linear function first. Then, this linear function is used by activation function as input and calculated activations are fed as input to the next layer.

1.Rectified Linear Unit(ReLU)

The ReLU is a faster learning activation function, which has proved to be the most successful and widely used function. It offers the better performance and generalization in deep learning compared to the Sigmoid and tanh activation functions.

Along with overall speed of computation enhanced, ReLU provides faster computation since it does not compute exponentials and divisions.

f(x)=x if x≥0 and 0 if x<0
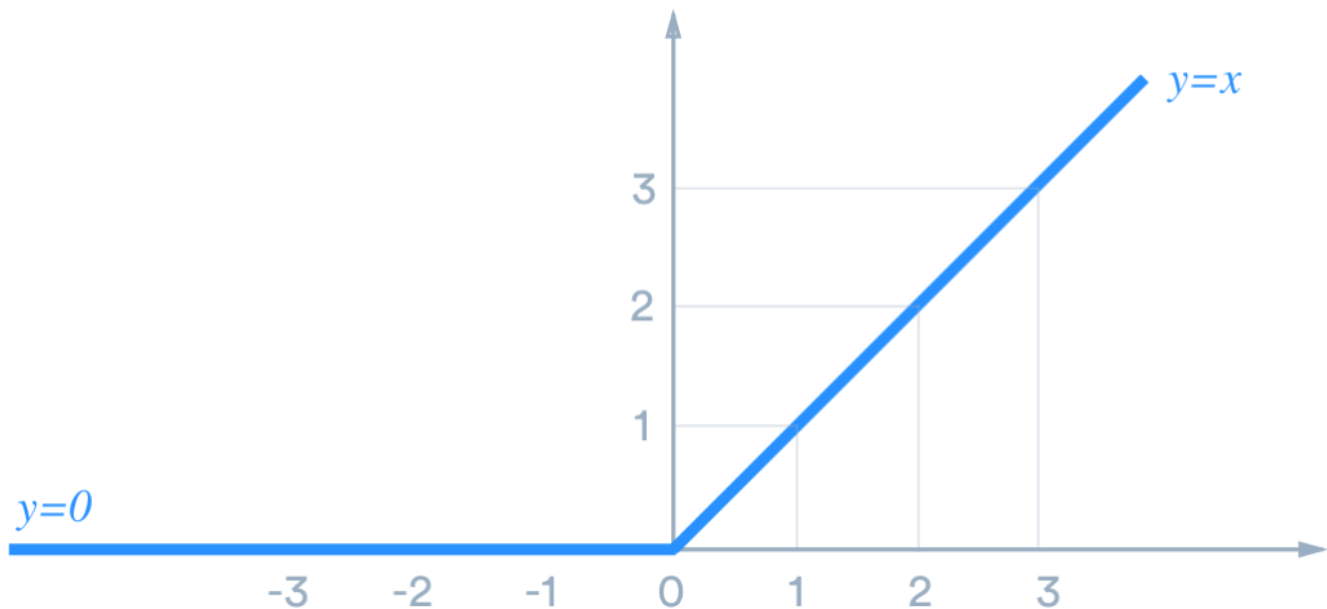
2.Sigmoid Function

The Sigmoid is a non-linear AF used mostly in feedforward neural networks. The sigmoid function appears in the output layers of the DL architectures, and they are used for predicting probability based output and has been applied successfully in binary classification problems.

It's major drawbacks are sharp damp gradients during backpropagation, gradient saturation, slow convergence and non-zero centred output thereby causing the gradient updates to propagate in different directions.

f(x)=1/(1+e^(-x)
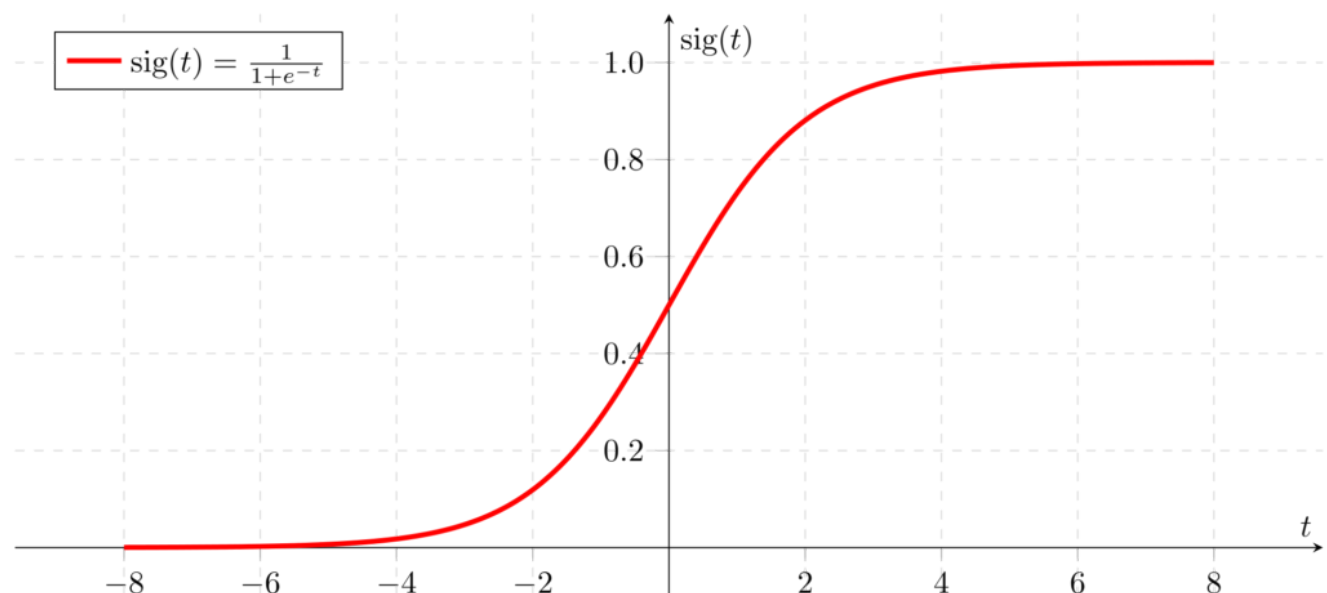
3.TanH Function

The tanh function became the preferred function compared to the sigmoid function in that it gives better training performance for multi-layer neural networks. However, the tanh function could not solve the vanishing gradient problem suffered by the sigmoid functions as well. The hyperbolic tangent function function, is a smoother zero-centred function whose range lies between -1 to 1.

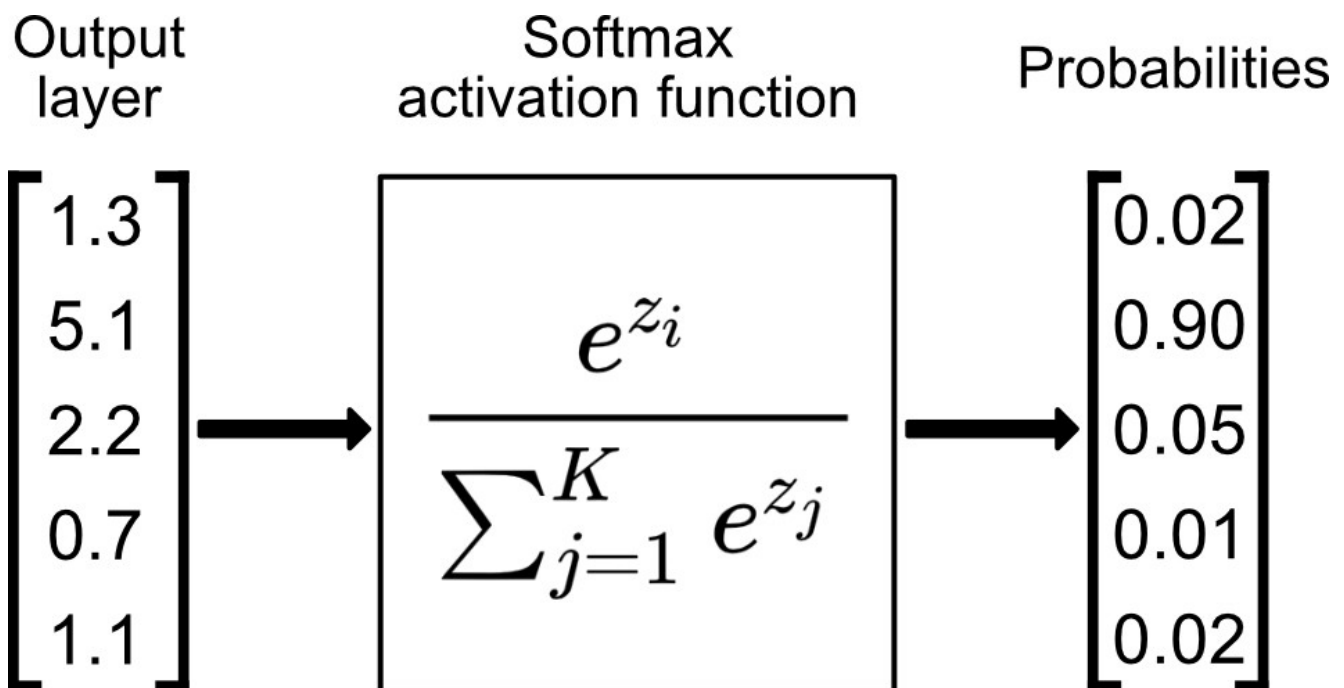$f(x)=(e^x-e^{(-x)})/(e^x+e^{(-x)})$
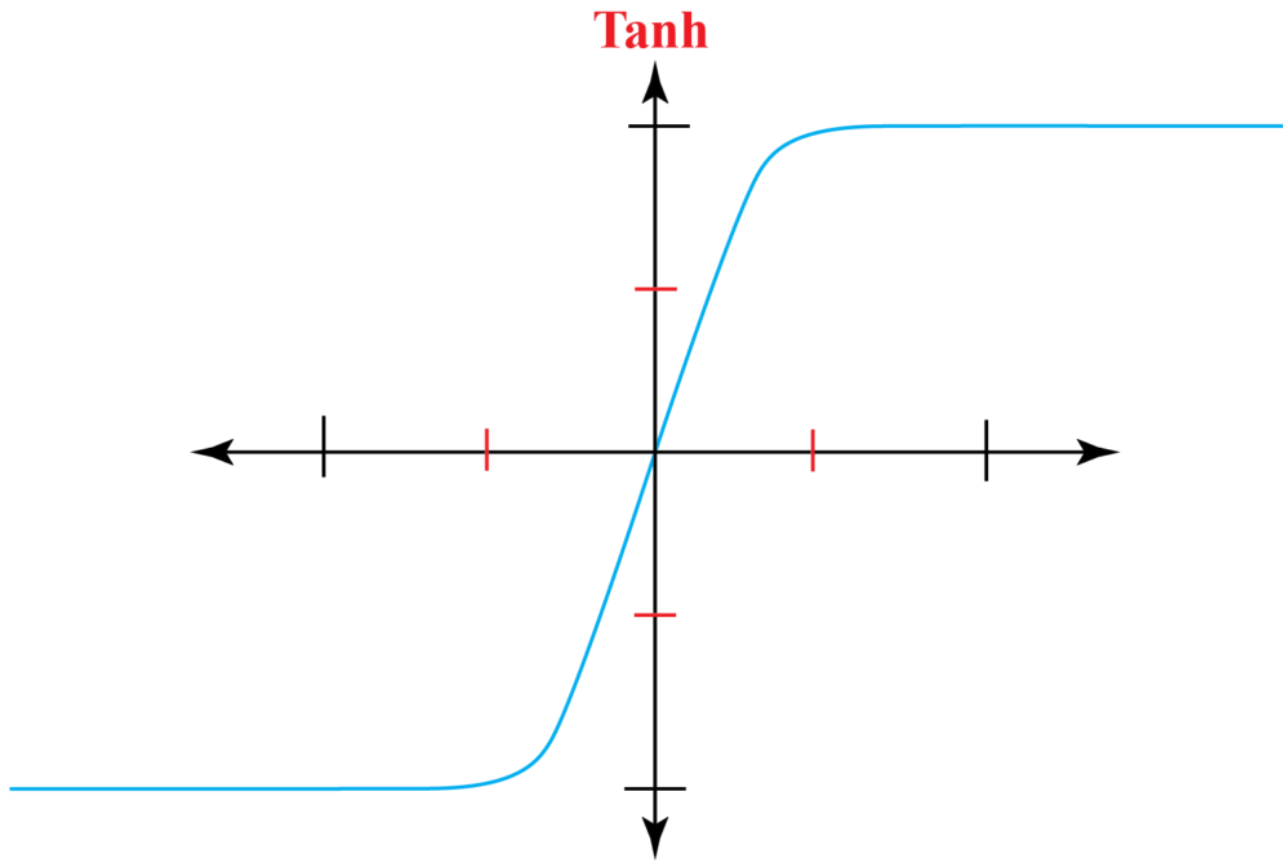
4.ELiSH

The Exponential linear Squashing activation function known as the ELiSH function is one of the most recent AF, proposed by Basirat and Roth, 2018. The ELiSH shares common properties with the Swish function. The ELiSH function is made up of the ELU and Sigmoid functions.
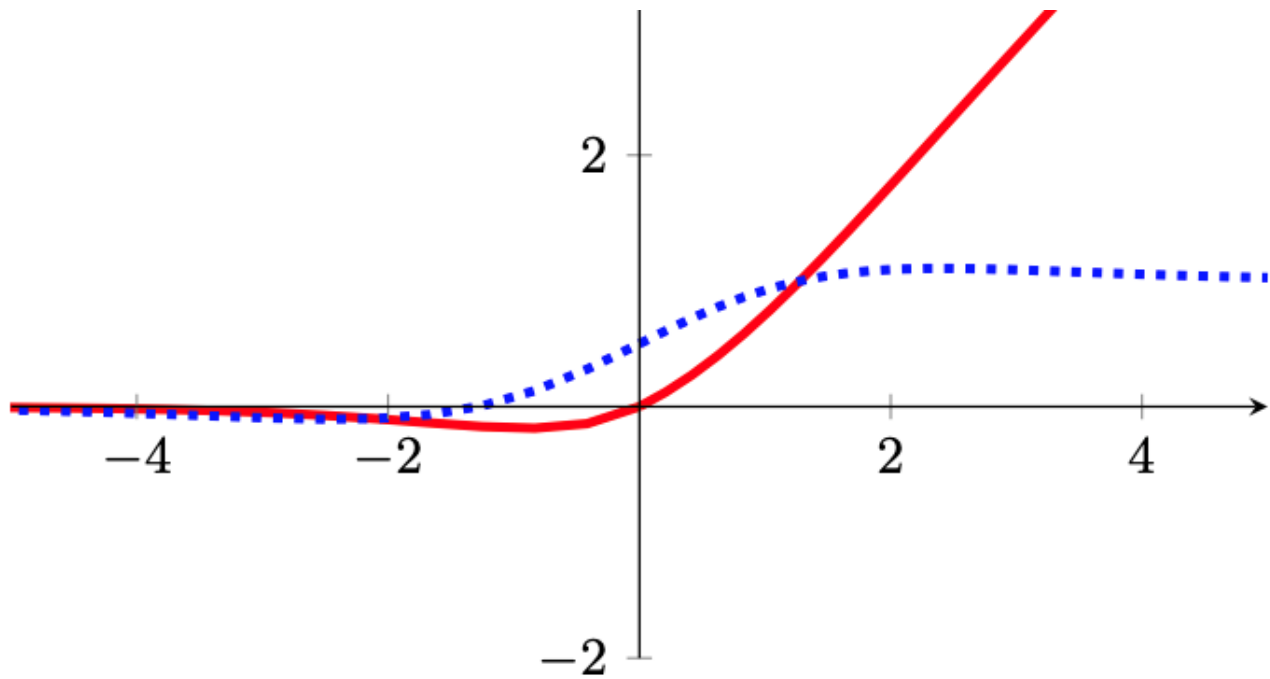
$$y(x) = \texttt{ELiSH}$$

5.Softmax Function

The Softmax function produces an output which is a range of values between 0 and 1, with the sum of the probabilities been equal to 1. The Softmax function is used in multi-class models where it returns probabilities of each class, with the target class having the highest probability.

$$\text{Output layer} \quad \begin{bmatrix} 1.3 \\ 5.1 \\ 2.2 \\ 0.7 \\ 1.1 \end{bmatrix} \rightarrow \boxed{\frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}} \rightarrow \begin{bmatrix} 0.02 \\ 0.90 \\ 0.05 \\ 0.01 \\ 0.02 \end{bmatrix} \quad \text{Probabilities}$$

These were the five important functions being used in the neural network architectures. You can also create your own neural network and can check its feasibility by training your dataset.

Thanks and Keep Learning!

Leave a Reply

Logged in as Shivam Bhardwaj. Log out?

Comment

Post Comment

Home

Products

Contact

© 2021 Proudly powered by WordPress