

Name: Argho Proshad Singha

Id: 20-42906-1

Sec: B

## **Report on Activation Function**

In artificial neural networks and machine learning, an activation function is a mathematical function that is applied to a neuron's output to introduce nonlinearity. The activation function assists in determining whether or not a neuron should fire based on input from other neurons in the network. In this study, I'll discuss six distinct activation functions, their mathematical underpinnings, advantages, and disadvantages.

### **1.Step function:**

The step function is a simple activation function commonly used in neural networks. It is defined mathematically as follows:

$$H(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$$

The step function has a few advantages and disadvantages:

Advantages:

- The step function is rapid to compute and easy to implement.
- Because it is a binary function, it can be used to solve binary classification issues.
- It is useful for educational reasons because it is simple to interpret and comprehend.

Disadvantages:

- The step function can experience vanishing gradients, when the gradients of the function become too small to be effective for training the network.
- It is not ideal for regression problems or any task that requires the output to be a continuous number.

The step function is not frequently used in conventional neural networks due to its drawbacks. Due to their

smoother gradients and capacity to handle a larger range of issues, other activation functions, such as the sigmoid or ReLU functions, are often favored.

## **2. Sigmoid function:**

The sigmoid function is a mathematical function that maps any input to a value between 0 and 1. The most commonly used sigmoid function is the logistic function, which is defined as:

$$f(x) = 1 / (1 + e^{(-x)})$$

where x is the input to the function.

Advantages:

- Sigmoid functions are important in optimization methods like gradient descent because they are simple to use and can be discriminated easily.
- In neural networks, they are frequently employed to convert the output of a neuron into a probability distribution.

Disadvantages:

- Sigmoid functions are susceptible to saturation, which makes it challenging for the function to learn further when x is large and the result is close to 1.
- At large values of x, the gradient of the sigmoid function shrinks dramatically, which might cause optimization techniques to converge slowly.
- Because sigmoid functions are asymmetric near zero, they have the potential to influence a model's output.

## **3. Tanh function:**

A mathematical function called the hyperbolic tangent function (tanh) converts input values into output values between -1 and 1. The equation is as follows:

$$\tanh(x) = (e^x - e^{(-x)}) / (e^x + e^{(-x)})$$

where  $e$  is the mathematical constant approximately equal to 2.71828, and  $x$  is the input value.

Advantages:

- Tanh, like the sigmoid function, is a smooth function that is simple to discriminate, which makes it helpful for backpropagation neural network training.
- Its outputs are centered around zero because it is a zero-centered function. This can aid in avoiding vanishing gradients when deep neural networks are being trained.

Tanh can be used to normalize data with a broad range of values because it is bounded between -1 and 1.

Disadvantages:

- Tanh's derivative is not necessarily positive or negative since it is not monotone. This may make employing some optimization strategies more challenging.
- Because tanh's output is not sparse, its memory and compute needs may be higher than those of other activation functions.

Tanh function is an important activation function in neural networks overall because it has advantages over other functions, such as the sigmoid function. It does, however, have important limitations that must be considered while using it in various circumstances.

#### **4. Relu function:**

. It is defined as:

$f(x) = \max(0, x)$  where  $x$  is the

input to the function.

Advantages:

- It is simple to implement and interpret;
- It has been demonstrated to function effectively in a variety of neural networks, including deep neural networks.
- The sparse output of the ReLU function can aid in preventing overfitting.

Disadvantages:

- Because the ReLU function is not symmetric, using it in some kinds of neural networks can be challenging.
- Because the ReLU function is not differentiable at  $x=0$ , some optimization techniques that depend on gradient data may encounter difficulties.

ReLU is a popular activation function since it is computationally efficient and useful in a variety of neural network architectures. It has some limitations, though, including the potential for the "dying ReLU" problem and issues with optimization strategies.

## **5. ELU function:**

The ELU (Exponential Linear Unit) step function is an activation function used in artificial neural networks. It is a variant of the ReLU (Rectified Linear Unit) function and is defined as:

$$\text{ELU}(x) = \{ x \text{ if } x \geq 0, \alpha * (\exp(x) - 1) \text{ if } x < 0 \}$$

where  $\alpha$  is a hyperparameter that determines the negative saturation value.

Advantages:

- Smoother gradients: The ELU function has a continuous derivative, which helps avoid the vanishing gradient problem that can occur with other activation functions like the sigmoid function.
- Better performance: Empirical evidence suggests that ELU activation functions can lead to better performance in deep neural networks compared to other activation functions.
- Faster convergence: The ELU function has been shown to help networks converge faster than other activation functions.

Disadvantages:

- Computationally expensive: The ELU function requires computing the exponential function, which can be computationally expensive.
- Unstable for large negative inputs: The ELU function can be unstable for large negative inputs, which can lead to numerical instability in the network.
- Not as well-known: The ELU function is not as well-known as other activation functions like the sigmoid and ReLU functions, which can make it harder to find resources and support for implementing it in neural networks.

## **6. SELu function:**

It is a type of activation function used in neural networks. Unlike other activation functions, such as ReLU, SELu has a smooth, continuous curve.

The formula for the SELu activation function is:

$$f(x) = 1.0507 * (e^x - 1), x < 0$$
$$f(x) = x, x \geq 0$$

Advantages:

- The SELu activation function is a self-normalizing function, which means that regardless of the input distribution, each neuron's output has a zero mean and unit variance. The issue of vanishing/exploding gradients, which can occur in deep neural networks, may be lessened as a result of this.
- It has been demonstrated that SELu outperforms other activation functions, such as ReLU and tanh, on a range of deep learning tasks, including speech recognition and image classification.

Disadvantages:

- Because to the exponentiation involved, the SELu activation function is computationally more expensive than others, such as ReLU.
- 
- Not all forms of data will respond effectively to SELu. While it has been demonstrated to be effective for many deep learning tasks, it may not be the best option for all types of data. To find the optimum activation function for a given task, some testing may be required.