Report on Activation Functions

NAME: MAHADI, MAHAMODUL HASAN
ID: 20-42768-1

***Abstract*—Activation functions play a crucial role in neural networks, serving as a non-linear transformation that introduces complex decision boundaries to allow the network to learn and make accurate predictions. The activation function takes the input signal of a neuron and applies a mathematical operation to it, determining whether or not the neuron should be activated and to what degree. There are several types of activation functions, including the widely-used sigmoid, ReLU, Step, ELU, TanH and SELU functions, each with its own advantages and disadvantages. The choice of activation function can have a significant impact on the performance of a neural network, and researchers continue to explore new activation functions that can improve accuracy, reduce training time, and address other challenges in deep learning.**

Keywords*— *activation functions, neural networks, non-linear transformation, sigmoid, ReLU, softmax.

I. INTRODUCTION

Activation functions are mathematical functions used in neural networks that introduce non-linearity in the output of a neuron. They are applied to the weighted sum of inputs to a neuron and determine the neuron's output, which is then passed on to other neurons or used as the final output of the network. Activation functions allow neural networks to learn complex patterns and relationships in the data, and they can be either continuous or discontinuous functions. Some popular activation functions include sigmoid, ReLU, and softmax functions, each with their own advantages and disadvantages depending on the specific application and type of neural network being used. The choice of activation function can have a significant impact on the performance and accuracy of a neural network..

II. ACTIVATION FUNCTIONS

A. *Step function*

The step function, also known as the Heaviside step function, is a simple activation function commonly used in neural networks. It is defined mathematically as follows:

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

The step function has a few advantages and disadvantages:

Advantages:

- The step function is simple to implement and can be computed quickly.
- It is a binary function, which makes it useful for binary classification problems.
- It is easy to interpret and understand, which makes it useful for educational purposes.

Disadvantages:

- The step function is not differentiable at x=0, which can cause problems when
- training neural networks using gradient descent algorithms.
- It is not suitable for regression problems or any task that requires the output to be a continuous value.
- The step function can suffer from the problem of vanishing gradients, where the gradients of the function become too small to be useful for training the network.

Thus, despite being a helpful activation function in some situations, 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 favoured.

B. *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 easy to work with and can be easily differentiated, which makes them useful in optimization algorithms like gradient descent.
- They are widely used in neural networks to map the output of a neuron to a probability distribution.
- Sigmoid functions are bounded, meaning that their outputs are always between 0 and 1, which can be useful in certain applications like probability calculations.

Disadvantages:

- Sigmoid functions are prone to saturation, which means that for large values of x, the output of the function becomes very close to 1, making it difficult for the function to learn further.
- The gradient of the sigmoid function becomes very small for large values of x, which can lead to slow convergence in optimization algorithms.
- Sigmoid functions are not symmetric around zero, which means that they can introduce bias into the output of a model.
-

C. *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:

- Like the sigmoid function, tanh is also a smooth function that can be easily differentiated, which makes it useful in training neural networks using backpropagation.
- It is a zero-centered function, which means that its outputs are centered around zero. This can help in preventing vanishing gradients during the training of deep neural networks.
- Tanh is bounded between -1 and 1, which makes it useful for normalization of data that has a wide range of values.

Disadvantages:

- Like the sigmoid function, tanh can suffer from the vanishing gradient problem, which can make it difficult to train deep neural networks.
- Tanh is not monotonic, which means that its derivative is not always positive or negative.
- This can make it more difficult to optimize using some optimization techniques.
- The output of tanh is not sparse, which means that it can be less efficient than other activation functions in terms of memory and computation requirements.
- Tanh function, which offers advantages over other functions like the sigmoid function, is a valuable activation function in neural networks overall. It does, however, have significant restrictions that must be taken into account while applying it in various situations.

D. *Relu function*

The ReLU (Rectified Linear Unit) function is a commonly used activation function in neural networks. It is defined as:

$$f(x) = \max(0, x)$$

Where x is the input to the function.

Advantages:

- The ReLU function is computationally efficient compared to other activation functions.
- It has been shown to work well in many types of neural networks, including deep neural networks.
- It is easy to implement and interpret.
- The ReLU function has a sparse output, which can help prevent overfitting.

Disadvantages:

- The ReLU function can suffer from the "dying ReLU" problem, where some
- neurons stop producing any output due to the input being negative. This can cause the network to become less expressive and impact its performance.
- The ReLU function is not symmetric, which can make it difficult to use in certain types of neural networks.
- The ReLU function is not differentiable at x=0, which can cause issues in some optimization algorithms that rely on gradient information.

In conclusion, the ReLU function is a well-liked activation function since it is computationally
effective and useful in many different kinds of neural networks. It does, however, have certain
drawbacks, such as the potential for the "dying ReLU" problem and difficulties with
optimization techniques.

E. *PReLU function*

PReLU (Parametric Rectified Linear Unit) is an activation function used in deep neural networks. It is an extension of the popular Rectified Linear Unit (ReLU) activation function, with the added parameter that allows the slope of the negative part of the function to be learned during training.

The PReLU activation function can be defined mathematically as follows:

$$f(x) = \max(0, x) + a * \min(0, x)$$

Where ‘a’ is a learnable parameter that determines the slope of the negative part of the function. If a=0, then PReLU reduces to the standard ReLU activation function.

Advantages of PReLU:

- Allows the negative slope of the function to be learned during training, which can improve the flexibility of the model.
- Can prevent the dying ReLU problem, where a large number of neurons can become inactive and stop learning during training.
- Has been shown to improve the performance of deep neural networks in image recognition and other applications.

Disadvantages of PReLU:

- Can lead to overfitting if the parameter a is not regularized properly.
- •Can increase the complexity of the model and the computational cost of training.

F. *EReLU function*

The Erelu function is a non-linear activation function used in deep learning. It is defined as:

$$\text{Erelu}(x) = \max(0, x) + \min(0, \alpha*(\exp(x)-1))$$

Where α is a hyperparameter that controls the curvature of the function.

Advantages:

- Non-linear: The Erelu function is non-linear, which means it can model complex relationships between inputs and outputs.
- Sparsity: The Erelu function can produce sparsity in the output, which can reduce overfitting and improve generalization.
- Smoothness: The Erelu function is smooth and differentiable everywhere, which allows for efficient gradient-based optimization.

Disadvantages:

- Complexity: The Erelu function is more complex than some other activation functions, which can increase computational cost and training time.

- Sensitivity to hyperparameters: The Erelu function's performance can be sensitive to the choice of hyperparameters, such as the value of α.
- Vanishing gradient: While the Erelu function can help prevent the vanishing gradient problem, it can still occur in certain cases, particularly with deep networks.

In deep learning, the Erelu function can be a beneficial activation function overall, especially for reducing overfitting and enhancing generalization. But, it has advantages and disadvantages like every activation function, therefore it is best to choose one based on the
demands of the current task