Step function :

A step function is a mathematical function that takes on a finite number of constant values over a finite range of the input variable. It is also known as a staircase function or a piecewise constant function.

Advantages:

Easy to understand: Step functions are simple and easy to understand because they only change values at specific points. They are commonly used in modeling situations where the output is either on or off, such as in binary classification problems.

Efficient computation: Because step functions are piecewise constant, they are computationally efficient and can be evaluated quickly. This makes them useful in applications where speed is important, such as signal processing.

Disadvantages:

Limited applicability: Step functions are limited in their applicability because they only work well for problems that have discontinuous outputs. They are not suitable for modeling problems where the output changes continuously.

Non-differentiable: Step functions are non-differentiable at the points where the function changes value. This can make them difficult to use in optimization problems that require the gradient of the function. In such cases, approximations or alternative functions may be used instead.

Sigmoid function :

The sigmoid function is a mathematical function that maps any input value to a value between 0 and 1. It is commonly used in machine learning and neural networks as an activation function to introduce non-linearity into the output of a neuron.

Advantages :

Non-linearity: The sigmoid function introduces non-linearity into the output of a neuron, which allows the neural network to model complex relationships between inputs and outputs.

Smoothness: The sigmoid function is a smooth and continuous function, which makes it easier to compute gradients during the training process of a neural network.

Disadvantages :

Vanishing gradients: The gradient of the sigmoid function becomes very small for very large or very small input values, which can lead to vanishing gradients during backpropagation and slow down the training of deep neural networks.

Outputs are not zero-centered: The output of the sigmoid function is not centered around zero, which can make it difficult for other layers in a neural network to learn meaningful representations.

Tanh function :

The TANh (hyperbolic tangent) function is a mathematical function commonly used in neural networks as an activation function. It maps input values from negative infinity to positive infinity to a range between -1 and 1.

Advantages :

It is a widely used activation function in neural networks because it is continuously differentiable and provides strong non-linear properties.
The TANh function is zero-centered, which means that the mean of its output values is close to zero, which can improve the performance of neural networks.

Disadvantages :

The TANh function is prone to the vanishing gradient problem, which can occur when the function's derivative becomes too small, causing the gradient to vanish, and slowing down or halting the learning process.
The TANh function can also suffer from the saturation problem, which occurs when the function's output becomes saturated at either +1 or -1, making it difficult for the neural network to learn further

Relu :
The Rectified Linear Unit (ReLU) function is a commonly used activation function in neural networks. Its equation is:

Avantages:

It is computationally efficient: the ReLU function is a simple function to compute, and it does not require any expensive operations like exponential calculations.

It helps to prevent the vanishing gradient problem: the ReLU function does not saturate for positive inputs, which means that it does not cause the gradient to become too small, helping to prevent the vanishing gradient problem.

However, the ReLU function also has two main disadvantages:

It is not differentiable at x = 0: the ReLU function is not differentiable at x = 0, which can cause problems when using certain optimization algorithms.

It suffers from the "dying ReLU" problem: if the input to a ReLU neuron is negative, the output of the neuron will be 0, which can cause the neuron to "die" and stop contributing to the network's output.

Elu:
The Exponential Linear Unit (ELU) function is another activation function that is similar to the ReLU function. Its equation is:

Advantages :
It is smooth and differentiable everywhere: the ELU function is a smooth function that is differentiable everywhere, which makes it easier to use with certain optimization algorithms.

It helps to prevent the "dying ReLU" problem: the ELU function does not cause neurons to "die" like the ReLU function can, which can help improve the performance of neural networks.

Disadvantages:

It is computationally more expensive than the ReLU function: the ELU function requires the calculation of an exponential function, which is more computationally expensive than the simple max function used in the ReLU function.

It is not as widely used as the ReLU function: while the ELU function has some advantages over the ReLU function, it is not as widely used and its benefits may not always outweigh the added computational cost.

Selu :
The SELU (Scaled Exponential Linear Unit) function is a type of activation function used in neural networks. It is designed to address some of the shortcomings of other commonly used activation functions such as the ReLU function.

Advantages :

Self-normalizing property: The SELU function has a self-normalizing property, which means that the output of the function has a mean of 0 and a standard deviation of 1 for inputs with a mean of 0 and a standard deviation of 1. This property can help to prevent vanishing gradients and improve the performance of deep neural networks, especially those with many hidden layers.

Better performance: The SELU function has been shown to perform better than other commonly used activation functions, such as the ReLU and sigmoid functions, on a variety of tasks, including image classification and natural language processing. This is because the SELU function allows for faster and more efficient learning, which can lead to better accuracy and faster training times.

Disadvantages :

Limited use: The self-normalizing property of the SELU function only applies to inputs with a mean of 0 and a standard deviation of 1. This means that the function may not be as effective on data that does not meet this criteria, such as data with a skewed distribution.

Complexity: The SELU function is more complex than other commonly used activation functions, such as the ReLU and sigmoid functions. This can make it more difficult to implement and optimize in some cases, especially for large-scale deep neural networks. Additionally, the choice of the alpha and scale parameters can have a significant impact on the performance of the function, which can make it more difficult to tune.