Activation functions

1. **Step Function**:

Advantages:

- Simple and easy to understand.
- Useful in modeling real-world phenomena where the value of the function changes abruptly at specific points.
- Easy to compute and implement.

Disadvantages:

- Limited accuracy and applicability.
- Discontinuous nature may lead to convergence problems.

2. **Sigmoid Function**:

Advantages:

- Smooth and continuous.
- Outputs are bounded between 0 and 1, making it useful for binary classification problems.
- Easy to compute and implement.

Disadvantages:

- Vanishing gradient problem may occur.
- Outputs outside the range [-2, 2] are close to zero, leading to a saturation of the gradients and slower learning.

3. **TanH Function**:

Advantages:

- Smooth and continuous.
- Outputs are bounded between -1 and 1, making it useful for symmetric data.
- Easy to compute and implement.

Disadvantages:

- Vanishing gradient problem may occur.
- Similar saturation problem as the sigmoid function.

4. Rectified Linear Units (ReLU) Function: Advantages:

- Simple and computationally efficient.
- Fast convergence due to the non-saturating nature of the function.
- Effective in preventing vanishing gradients.

Disadvantages:

- Outputs are not bounded, leading to the exploding gradients problem.
- Outputs can be fragile to noise.

5. **Exponential Linear Units (ELUs) Function:**

Advantages:

- Smooth and continuous.
- Outputs are negative for negative inputs, making it more biologically plausible.
- Effective in preventing vanishing gradients.

Disadvantages:

- Outputs are not bounded, leading to the exploding gradients problem.
- More computationally expensive than ReLU.

6. **S-shaped Linear Unit (SLU) Function:**

Advantages:

- Smooth and continuous.
- Outputs are bounded between 0 and 1, making it useful for binary classification problems.
- Effective in preventing vanishing gradients.

Disadvantages:

- Not as widely used and tested as other activation functions.
- May require more computation and time compared to other functions.

The ELU function is similar to the ReLU function for **x > 0**, but it has a smooth curve for **x < 0**, which allows it to produce negative output values. This can be useful in certain types of neural networks, such as those used for image processing or natural language processing. The SELU activation function is similar to the ELU function, but with a normalization step to ensure that the mean and variance of the output are approximately constant during training. One advantage of the SELU function over the ELU function is that it has been shown to produce better results on certain types of neural networks, especially those with many layers. This is due to the normalization property, which helps prevent the vanishing/exploding gradient problem that can occur during training. the SELU function is similar to the ELU function for **x > 0**, but it has a steeper curve for **x < 0**, which allows it to produce negative output values more easily. Additionally, the function has a normalization property, which helps prevent the vanishing/exploding gradient problem during training. This can be especially useful for deep neural networks with many layers.