ID: 20-43011-1

Name: Sowrov Ahmed

Section: A

Report on Activation Functions

Activation functions are an essential component in neural networks. They introduce non-linearity into the system, which enables the neural network to model complex relationships between inputs and outputs. The activation function determines the output of a neuron based on the weighted sum of inputs. In this report, I will discuss six commonly used activation functions: Step, Sigmoid, Tanh, Relu, Elu, and Selu.

1. Step Function:

Introduction: The step function is one of the simplest activation functions used in neural networks. It is a binary function that returns a 1 if the input is greater than or equal to zero, and 0 otherwise.

Formula: f(x) = 1 if x >= 0; 0 otherwise

Advantages:

- It is computationally efficient and easy to implement.
- It can be useful in some binary classification problems.

Disadvantages:

- The function is not continuous, which can create problems during gradient descent optimization.
- It can cause the model to get stuck in local minima during training.

2. Sigmoid Function:

Introduction: The sigmoid function is a smooth and bounded activation function that maps any input to a value between 0 and 1. It is commonly used in neural networks for binary classification problems.

Formula: f(x) = 1 / (1 + e^(-x))

Advantages:

- It produces a smooth output, which makes it easier to compute gradients during backpropagation.
- It is useful in binary classification problems where the output should be a probability value.

Disadvantages:

- It suffers from the vanishing gradient problem, which can slow down or even halt the learning process during training.
- It is not suitable for multi-class classification problems.

3. **Tanh Function:**

Introduction: The hyperbolic tangent (tanh) function is similar to the sigmoid function, but it maps any input to a value between -1 and 1. It is commonly used in neural networks for classification problems.

Formula: f(x) = (e^(x) - e^(-x)) / (e^(x) + e^(-x))

Advantages:

- It produces a smooth output, which makes it easier to compute gradients during backpropagation.
- It is useful in classification problems where the output should be a value between -1 and 1.

Disadvantages:

- It also suffers from the vanishing gradient problem, which can slow down or even halt the learning process during training.
- It is not suitable for multi-class classification problems.

4. **ReLU Function:**

Introduction: The rectified linear unit (ReLU) is a simple activation function that returns the input if it is positive, and 0 otherwise. It is one of the most popular activation functions in deep learning due to its simplicity and effectiveness.

Formula: f(x) = max(0, x)

Advantages:

- It is computationally efficient and easy to implement.
- It does not suffer from the vanishing gradient problem, which makes it suitable for deep neural networks.
- It has been shown to perform well in a variety of neural network architectures.

Disadvantages:

- It can cause a problem known as "dying ReLU" during training, where some neurons become inactive and stop learning.

5. **ELU Function:**

Introduction: The exponential linear unit (ELU) is similar to ReLU, but it has a non-zero output for negative input values. It is designed to improve the learning speed and stability of deep neural networks.

Formula: f(x) = x if x >= 0; alpha * (e^(x) - 1) if x < 0

Advantages:

- It can speed up the learning process in deep neural networks.
- It has been shown to outperform ReLU in some cases.
- It does not suffer from the dying ReLU problem.

Disadvantages:

- It is computationally more expensive than ReLU.

6. SELU Function:

Introduction: The scaled exponential linear unit (selu) is a self-normalizing variant of the elu function, which has been shown to improve the performance of deep neural networks.

Formula: f(x) = lambda * (e^x - 1), x <= 0

f(x) = lambda * x, x > 0

where lambda and alpha are constants that ensure the mean and variance of the output of each layer remain the same during training.

Advantages:

- The SELU function is self-normalizing, meaning that it can maintain a stable mean and variance of activations throughout the network.
- It does not suffer from the vanishing gradient problem.

Disadvantages:

- The SELU function can suffer from the exploding gradient problem.

Conclusion: Activation functions play a crucial role in the performance of neural networks. Different activation functions have different advantages and disadvantages, and their choice depends on the specific problem and the architecture of the network. In practice, relu and its variants are the most commonly used activation functions due to their simplicity and good performance on a wide range of problems. However, more advanced activation functions such as selu are also gaining popularity due to their ability to improve the performance of deep neural networks.