Report on Activation Functions for MNIST Dataset

Introduction

In this report, we evaluate the performance of three popular activation functions—Sigmoid, Tanh, and ReLU—within a simple neural network architecture for the task of handwritten digit classification using the MNIST dataset. The primary objective is to determine which activation function yields the highest accuracy and to analyse the implications of each function on the model's performance.

Results Summary

The following table summarizes the accuracy achieved by each activation function:

Activation Function	Accuracy
Sigmoid	0.9769
Tanh	0.9765
ReLU	0.9775

Observations

1. **ReLU Activation Function**:
    - **Accuracy**: 0.9775
    - **Performance**: The ReLU activation function achieved the highest accuracy among the three functions tested. This suggests that ReLU is effective in allowing the model to learn complex patterns in the data.
    - **Characteristics**: ReLU introduces non-linearity while maintaining computational efficiency. It avoids the vanishing gradient problem, which can hinder training in deeper networks. This characteristic likely contributed to its superior performance in this task.
2. **Sigmoid Activation Function**:
    - **Accuracy**: 0.9769
    - **Performance**: The Sigmoid function performed well, achieving an accuracy close to that of ReLU. However, it is slightly less effective than ReLU in this context.
    - **Characteristics**: Sigmoid outputs values between 0 and 1, which can be interpreted as probabilities. However, it suffers from the vanishing gradient problem, especially in deeper networks, which can slow down learning and affect performance.

3. **Tanh Activation Function**:
    - **Accuracy**: 0.9765
    - **Performance**: The Tanh function performed similarly to Sigmoid, with an accuracy slightly lower than both Sigmoid and ReLU.
    - **Characteristics**: Tanh outputs values between -1 and 1, providing a zero-cantered output which can help in certain scenarios. However, like Sigmoid, it is also susceptible to the vanishing gradient problem.

Recommendations

Based on the observed accuracy, **ReLU** is recommended as the best-performing activation function for the MNIST dataset. ReLU's efficiency in computation and its ability to mitigate vanishing gradient problems make it ideal for deep learning models, particularly in tasks like image classification, where it outperforms other functions in both speed and accuracy.

Justification:

- **Efficiency:** ReLU is computationally less expensive compared to Sigmoid and Tanh, as it involves simple thresholding at zero.
- **Gradient Flow:** ReLU helps to maintain gradient flow throughout the network, preventing the vanishing gradient problem common with Sigmoid and Tanh.
- **Sparse Activation:** ReLU results in sparse activation, meaning that only a few neurons are activated at a time, which can lead to more efficient learning and better generalization.

Conclusion

In conclusion, while all three activation functions demonstrated good performance on the MNIST dataset, the ReLU activation function emerged as the best performer. Its combination of high accuracy, computational efficiency, and ability to mitigate the vanishing gradient problem makes it the preferred choice for this image classification task. Future work could involve exploring variations of ReLU, such as Leaky ReLU or Parametric ReLU, to further enhance model performance.