Activation Functions Assignment

1. Explain the role of activation functions in neural networks. Compare and contrast linear and nonlinear activation functions. Why are nonlinear activation functions preferred in hidden layers?
Activation functions introduce non-linearity into neural networks, enabling them to learn complex patterns. Linear functions cannot capture these patterns, while nonlinear functions (e.g., ReLU, Sigmoid) allow for better learning of data relationships. Nonlinear functions are preferred in hidden layers because they enhance the network's capacity to model complex data.

2. Describe the Sigmoid activation function. What are its characteristics, and in what type of layers is it commonly used? Explain the Rectified Linear Unit (ReLU) activation function. Discuss its advantages and potential challenges. What is the purpose of the Tanh activation function? How does it differ from the Sigmoid activation function?
Sigmoid outputs values between 0 and 1, making it suitable for binary classification in output layers. ReLU is non-linear and outputs the input directly if positive, otherwise zero, making it computationally efficient but prone to "dead neurons." Tanh outputs values between -1 and 1, unlike Sigmoid, which outputs between 0 and 1, offering better gradients for optimization.

3. Discuss the significance of activation functions in the hidden layers of a neural network.
Activation functions in hidden layers enable the network to learn complex and non-linear relationships between inputs and outputs. They are crucial for the network's ability to approximate complicated functions.

4. Explain the choice of activation functions for different types of problems (e.g., classification, regression) in the output layer.
For classification tasks, softmax or sigmoid is used (softmax for multi-class, sigmoid for binary). For regression tasks, linear activation is often used in the output layer to produce continuous values.

5. Experiment with different activation functions (e.g., ReLU, Sigmoid, Tanh) in a simple neural network architecture. Compare their effects on convergence and performance.
ReLU typically leads to faster convergence and avoids the vanishing gradient problem, making it a good choice for deep networks. Sigmoid and Tanh may

suffer from slower convergence due to gradient vanishing, especially in deep networks. ReLU generally performs better in terms of training speed and efficiency.