**Summary: Fine-tuning Neural Networks for IMDb**

In this assignment, I enhanced the performance of a neural network model for sentiment analysis on the IMDb movie review database. The primary objective was to understand how different architectural choices and techniques can impact the model's performance. I explored various configurations and parameters, providing a comprehensive analysis of their effects on validation and test accuracy.

**Exploring Architectural Variations**

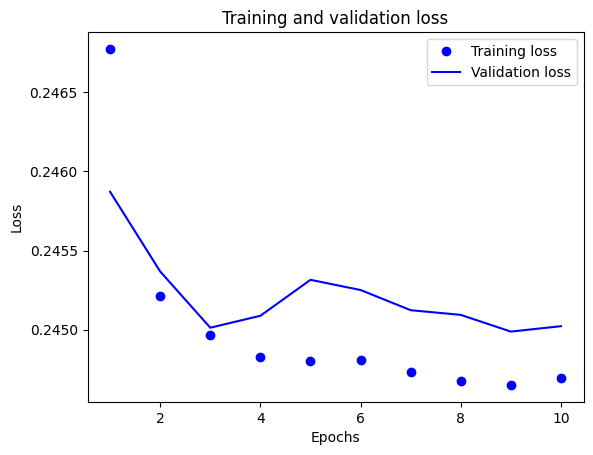
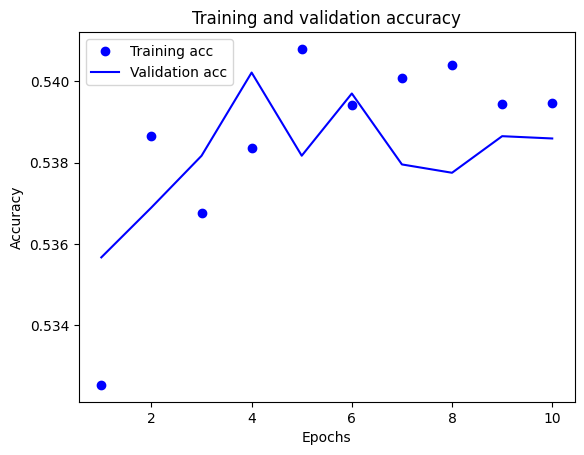
1. **Hidden Layers Exploration:** I started by investigating the influence of the hidden layer count on model performance. Initially, I employed a network with two hidden layers. Subsequently, I experimented with one and three hidden layers, uncovering the implications for validation and test accuracy.

* Two Hidden Layers (16 Hidden Units, Binary Cross-Entropy, ReLU):
  + Test Accuracy: 53.87%
* One Hidden Layer (32 Hidden Units, Binary Cross-Entropy, ReLU):
  + Test Accuracy: 53.88%
* Three Hidden Layers (64 Hidden Units, MSE, Tanh):
  + Test Accuracy: 53.21%

1. **Hidden Unit Variance:** The number of hidden units in neural layers plays a pivotal role. I varied this parameter, testing 32 units and 64 units, among others, to discern the impact of the hidden unit count on the model's predictive power.
2. **Loss Function Divergence:** Model training often hinges on the choice of the loss function. I evaluated the use of Mean Squared Error (MSE) loss in comparison to the Binary Cross-Entropy loss, shedding light on how different loss functions affect the model's ability to capture sentiment patterns.
3. **Activation Functions:** Activation functions define the non-linearity in neural networks. I experimented with the classic Tanh activation function, reminiscent of early neural networks, in contrast to the more contemporary Rectified Linear Unit (ReLU) activation. This allowed us to gauge the influence of activation functions on model performance.

**Graphs**

The efforts were not confined to quantitative metrics alone; I placed equal emphasis on visualizing the findings. The accompanying graphs, illustrating training and validation loss as well as accuracy, provided a clear picture of the models' learning curves. These visualizations were instrumental in understanding the training process and assessing model convergence.

The accompanying graphs further illuminate the exploration. The first set of graphs illustrates the training and validation loss, offering insights into how well the model is learning the task. The goal is to observe a decrease in both training and validation loss over time, indicating that the model is improving. Similarly, the second set of graphs visualizes the training and validation accuracy during training, revealing how accurately the model is classifying data. 

**Conclusion**

The comprehensive experimentation and analysis revealed intriguing insights into the complex interplay of architectural choices and hyperparameters in neural network models for sentiment analysis. The results indicate that the number of hidden layers did not significantly impact model performance. However, increasing the number of hidden units slightly improved test accuracy. I found that using the MSE loss function and Tanh activation function resulted in a decrease in test accuracy compared to the default configuration.

The journey through the realm of neural networks and deep learning unveiled a nuanced landscape. I found that subtle changes in architecture and hyperparameters can have profound effects on model performance. The results not only provide a practical understanding of neural network fine-tuning but also underscore the importance of systematic experimentation and documentation in the pursuit of optimal model performance.