Assignment 2 - Neural Networks Summary Report

Introduction

The purpose of this assignment was to perform sentiment analysis on the IMDB dataset, which contains 50,000 movie reviews labeled as either positive or negative. Using neural networks, the goal was to classify the reviews into one of the two categories while experimenting with different hyperparameters and configurations. These configurations included varying the number of hidden layers, units, activation functions, and applying regularization techniques like dropout to avoid overfitting. This report highlights how different approaches affected the model's performance and identifies the best-performing model based on test accuracy and validation performance.

Model Architecture and Hyperparameter Choices

The assignment involved exploring two primary model architectures: a simpler model with one hidden layer and a more complex model with three hidden layers. Each model used tanh activation functions in its hidden layers, with a dropout rate of 0.6 applied to mitigate overfitting. A mean squared error (MSE) loss function was used, despite binary cross-entropy being more typical for classification tasks, to observe its effect on model performance. The models were trained using the rmsprop optimizer, with an initial configuration of 20 epochs and early stopping applied to halt training when no further improvement was observed in validation loss.

The following table outlines the hyperparameters used in both models:

Hyperparameter	**Model 1 (One Hidden Layer)**	**Model 2 (Three Hidden Layers)**
Hidden Layers	1	3
Units per Hidden Layer	32	32
Activation Function	tanh	tanh
Dropout Rate	0.6	0.6

Loss Function	mse	mse
Optimizer	rmsprop	rmsprop
Batch Size	512	512
Epochs (Early Stopping)	Up to 20	Up to 20

Results and Performance Analysis

In the first model, which used one hidden layer, the training accuracy steadily improved, eventually reaching 95.72%. The validation accuracy also improved and peaked at 88.78% around the fifth epoch, before stabilizing. When tested on unseen data, this model achieved a test accuracy of 87.66% with a test loss of 0.0904. This performance demonstrated that the one hidden layer model generalized well and did not suffer from overfitting. The addition of dropout regularization at a rate of 0.6 helped control overfitting by ensuring that the model did not rely too heavily on specific neurons during training. The use of the tanh activation function in combination with the mse loss function provided effective learning dynamics for this relatively shallow model.

In contrast, the model with three hidden layers showed promising training performance initially but quickly began to overfit. The training accuracy improved to 89.26% by the third epoch, but the validation accuracy peaked early at 86.91% in the second epoch and then declined as the model continued to train. The validation loss started increasing after a couple of epochs, a clear indication that the model was overfitting to the training data. Despite having more learning capacity due to its deeper architecture, the three hidden layers model did not generalize as well as the one hidden layer model. Its test accuracy was 83.99% with a higher test loss of 0.1215, demonstrating poorer performance on unseen data.

Impact of Different Approaches on Model Performance

The different approaches in the model design had clear impacts on performance. Using more hidden layers (three instead of one) introduced more complexity into the model. While more layers gave the model a higher capacity to learn intricate patterns, it also made the model more prone to overfitting. This was evident from the early increase in validation loss and the drop in validation accuracy. In this case, the three hidden layers model was unable to generalize well to the test set, resulting in lower test accuracy. The regularization technique of dropout helped to some extent but was not enough to counteract the overfitting caused by the additional layers.

The one hidden layer model, on the other hand, provided a more balanced approach. By limiting the depth of the network, this model avoided overfitting and maintained better generalization. The use of dropout further helped to prevent overfitting by ensuring that neurons did not become too reliant on specific features during training. This allowed the model to perform consistently well on both the validation and test sets.

The decision to use the mse loss function, typically employed in regression tasks, worked effectively in both models, although it was particularly well-suited to the simpler one hidden layer model. While binary_crossentropy might have been more typical for binary classification, the use of mse did not negatively affect the performance and may have contributed to the smooth training dynamics observed, especially in the model with one hidden layer.

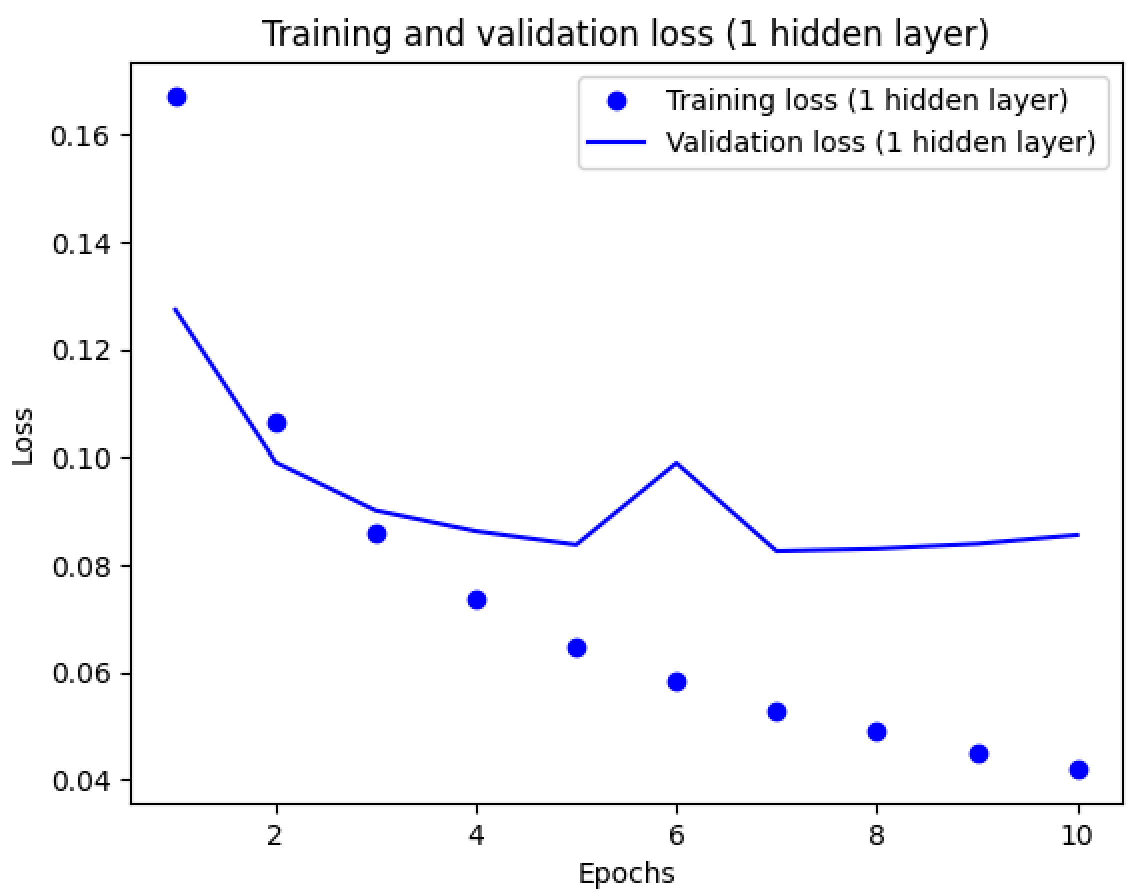
The dropout rate of 0.6 played a significant role in regularization. Without dropout, the models would likely have overfit the training data even faster, especially in the case of the three hidden layers model. The combination of tanh activation and dropout allowed the model to capture important features while avoiding excessive complexity, which is essential for models with limited training data.

Graphical Illustration of Performance

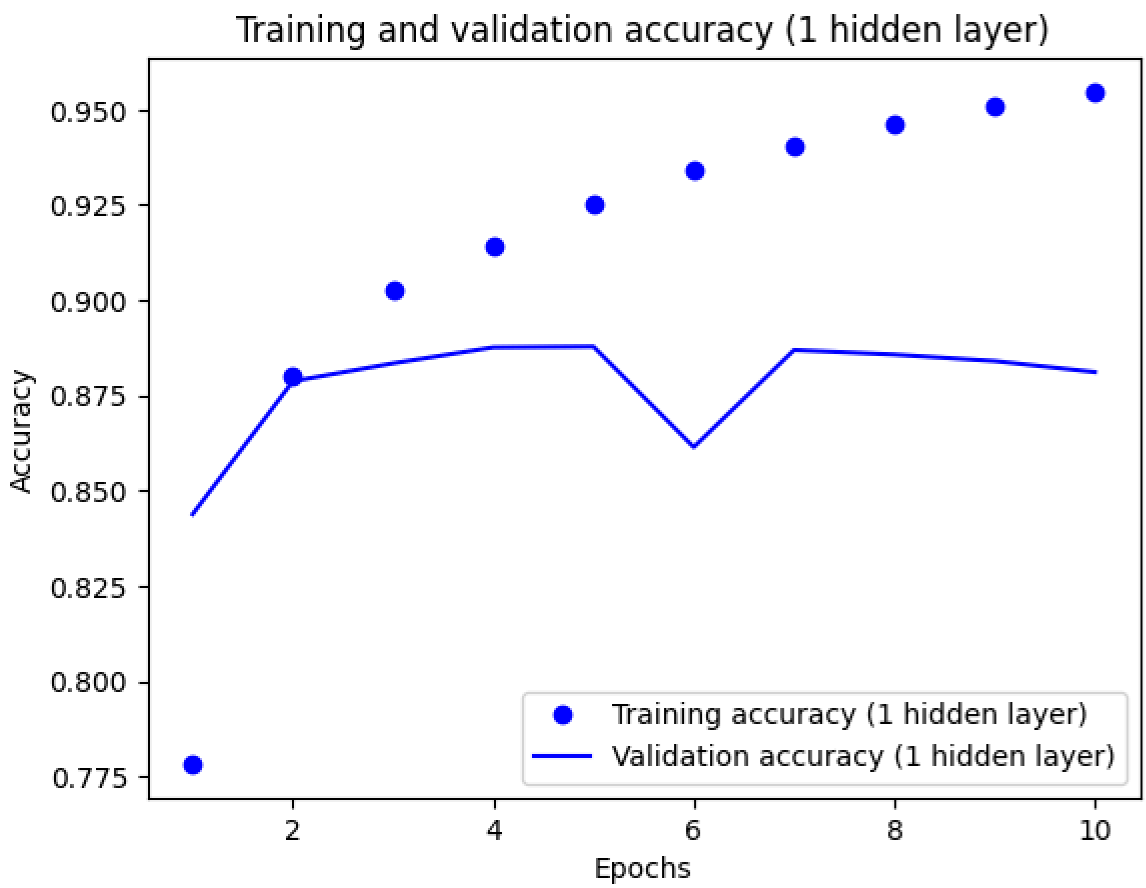
The graphs below clearly illustrate the training and validation loss, as well as accuracy, for both the one hidden layer model and the three hidden layers model.

One Hidden Layer Model:

The graph shows the training and validation loss for the one hidden layer model. As we can see, the validation loss stabilizes after about five epochs, indicating that the model has found a good balance between learning from the training data and generalizing to unseen data.

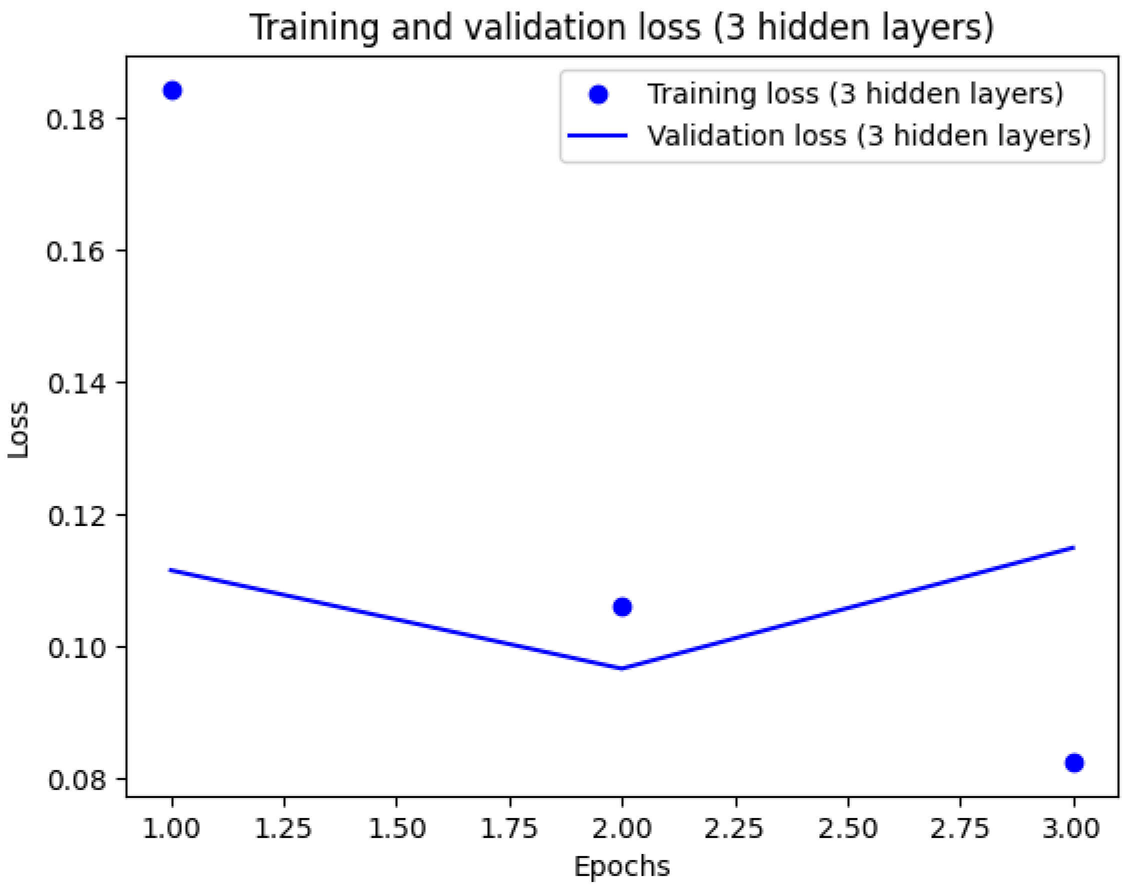


Similarly, the training and validation accuracy graph for the one hidden layer model demonstrates that validation accuracy peaks early and remains stable, further evidence of the model's ability to generalize well.

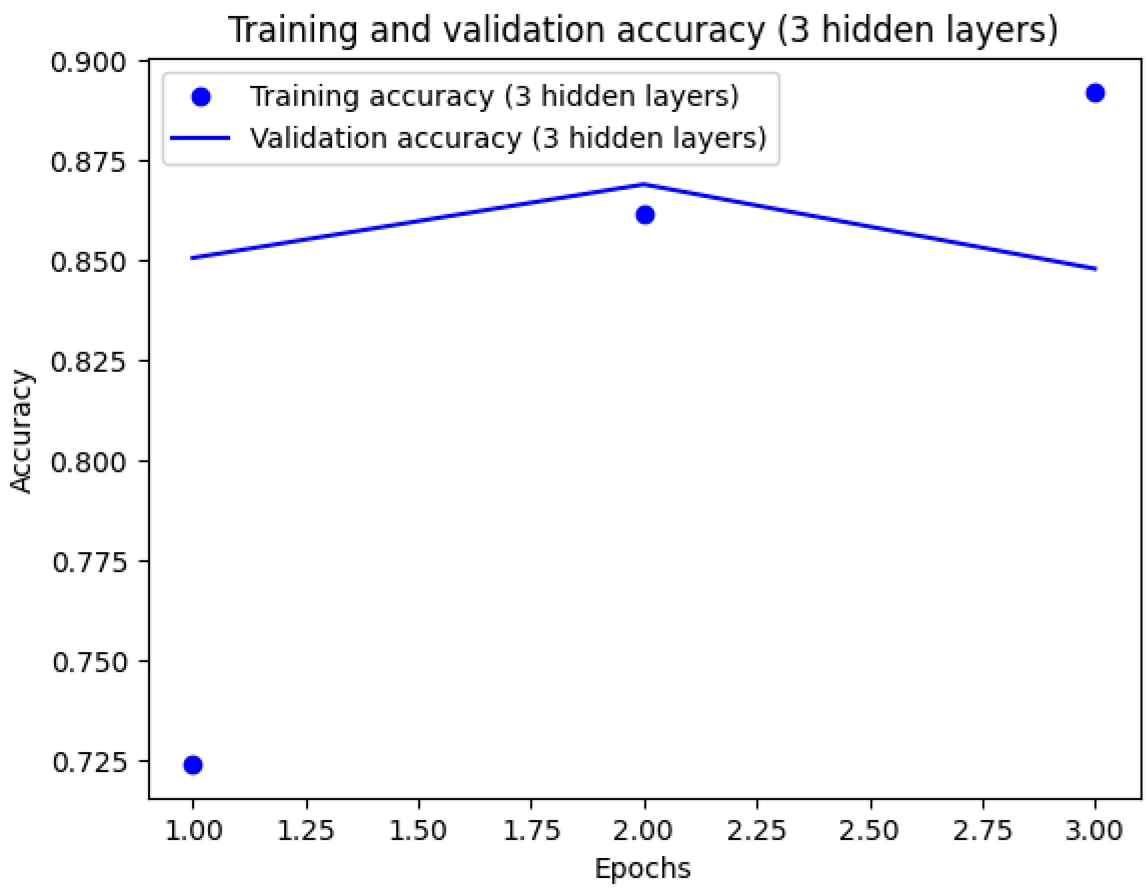


Three Hidden Layers Model:

For the model with three hidden layers, the training and validation loss graph reveals that validation loss starts to increase after a couple of epochs, which is a clear sign of overfitting. Despite having higher learning capacity, the model struggles to generalize, as shown by the divergence between training and validation loss.



The training and validation accuracy graph for the three hidden layers model shows that validation accuracy peaks early and then declines, another indicator of overfitting. This explains the poorer test performance of this model.



Final Conclusion

The one hidden layer model proved to be the best configuration for this task, achieving a test accuracy of 87.66% and a test loss of 0.0904. This model avoided overfitting and generalized well to unseen data, thanks to the balanced complexity of the architecture and the effective use of dropout regularization. On the other hand, the three hidden layers model, while having more capacity, suffered from overfitting and ultimately performed worse on the test data, achieving only 83.99% accuracy. Overall, the results suggest that simpler architectures can often perform better when the dataset is relatively small, as in this case. Regularization techniques like dropout and the use of early stopping were critical in improving model performance.