(64061) ADVANCED MACHINE LEARNING

Assignment – 2 Neural Networks

Name: Chandra Lekha Suggala

Student ID: 811344864

E-mail ID: csuggala@kent.edu

Date: 03/04/2025

Assignment Objective:

The goal in this assignment is to tune a neural network model's performance for the IMDB sentiment classification task. This is done by testing different configurations via the adjustment of parameters such as the number of hidden layers, hidden units, activation functions, and loss functions. Regularization methods such as dropout are used to avoid overfitting.

Dataset: IMDB Sentiment Classification

IMDB dataset, which has movie reviews labeled as positive or negative, is used for this task. The dataset has 50,000 reviews, out of which 25,000 are training examples and 25,000 are test examples. The model's task is to classify the sentiment of the reviews.

1. Number of Hidden Layers Effectiveness:

Objective: To compare the performance of networks with 1, 2, 3,4 and 5 hidden layers.

Effect:

• The model converges rapidly with a single hidden layer, and the validation accuracy is near the training accuracy, indicating good generalization.

• Adding more hidden layers increased the complexity but resulted in only incremental (or even slight reduction) improvements in accuracy, perhaps due to overfitting.

2. Change in Number of Hidden Units:

Objective: To analyze the effect of changing the number of hidden units per layer (32, 64, 128, 264) on model performance

Effect:

• Accuracy increases steadily as the number of hidden units increases.

• The model with 256 hidden units had learning curves that were smoother, indicating stronger and more stable training compared to models with fewer units.

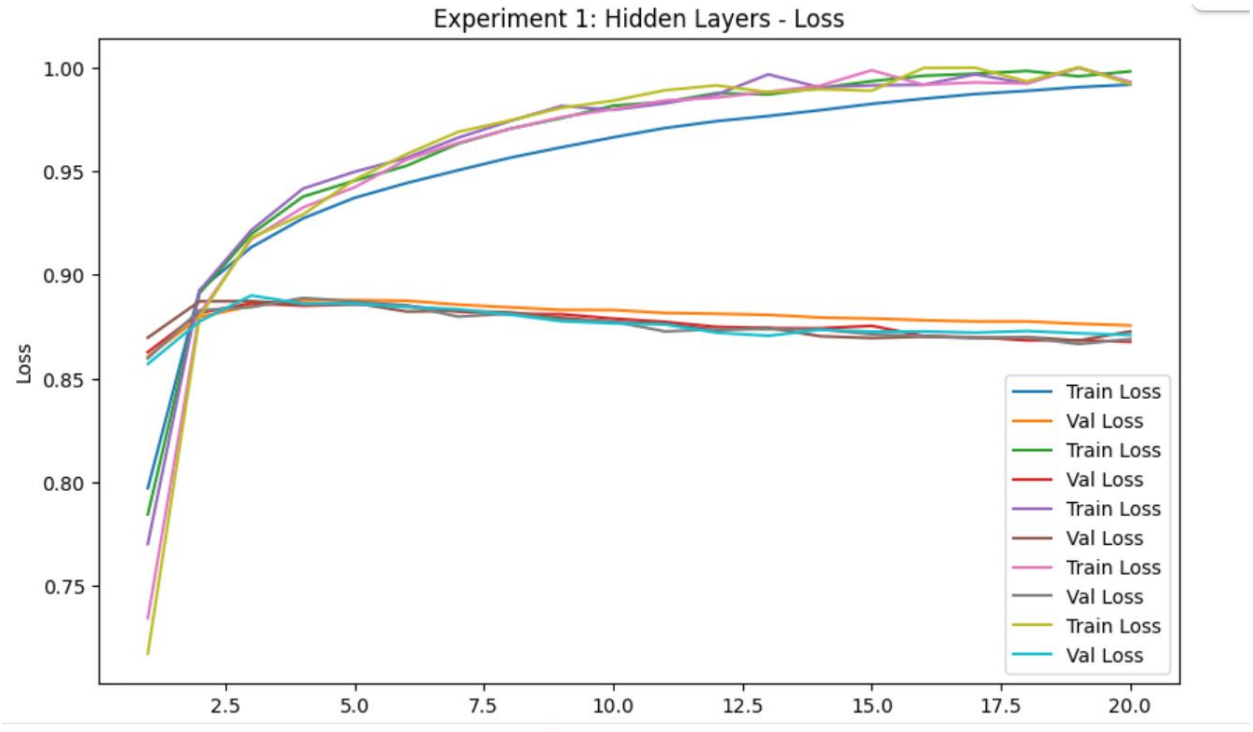
Experimental Setup:

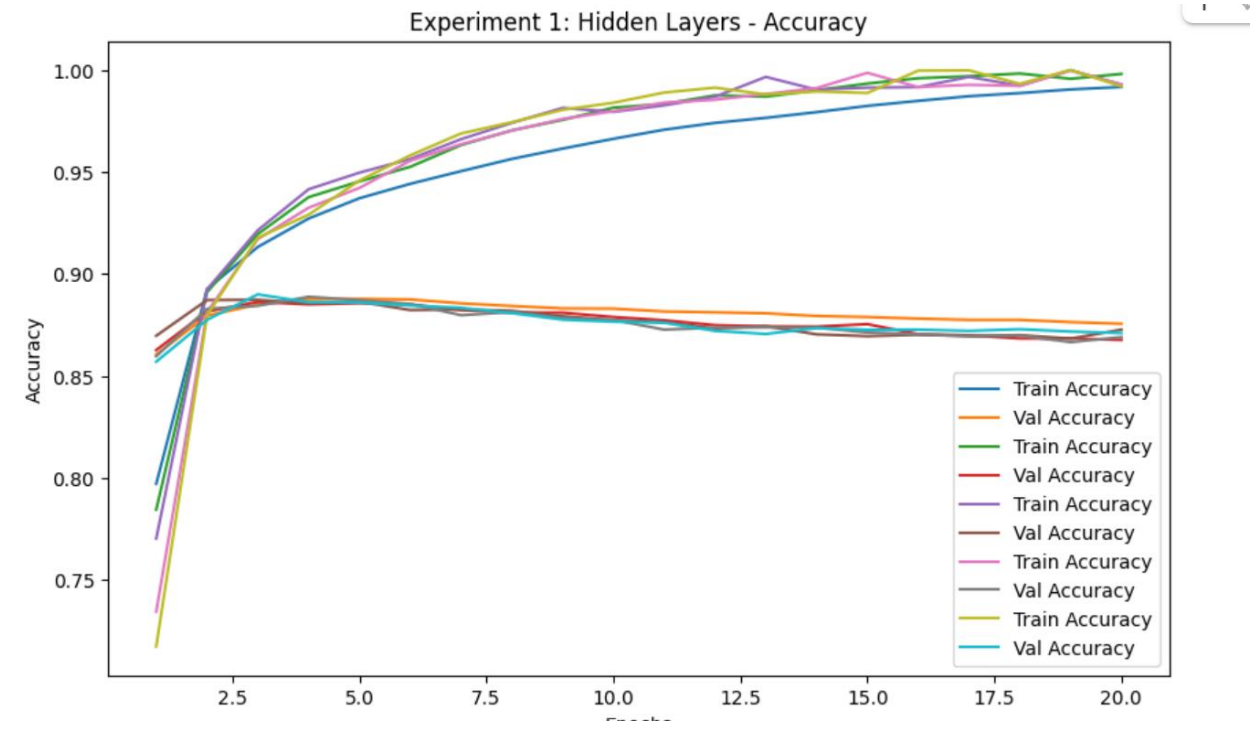
The neural network model is a basic feed-forward network with the following characteristics:

• Input layer: A 10,000-dimensional vector representing the binary multi-hot encoding of the top 10,000 most frequent words in the reviews.

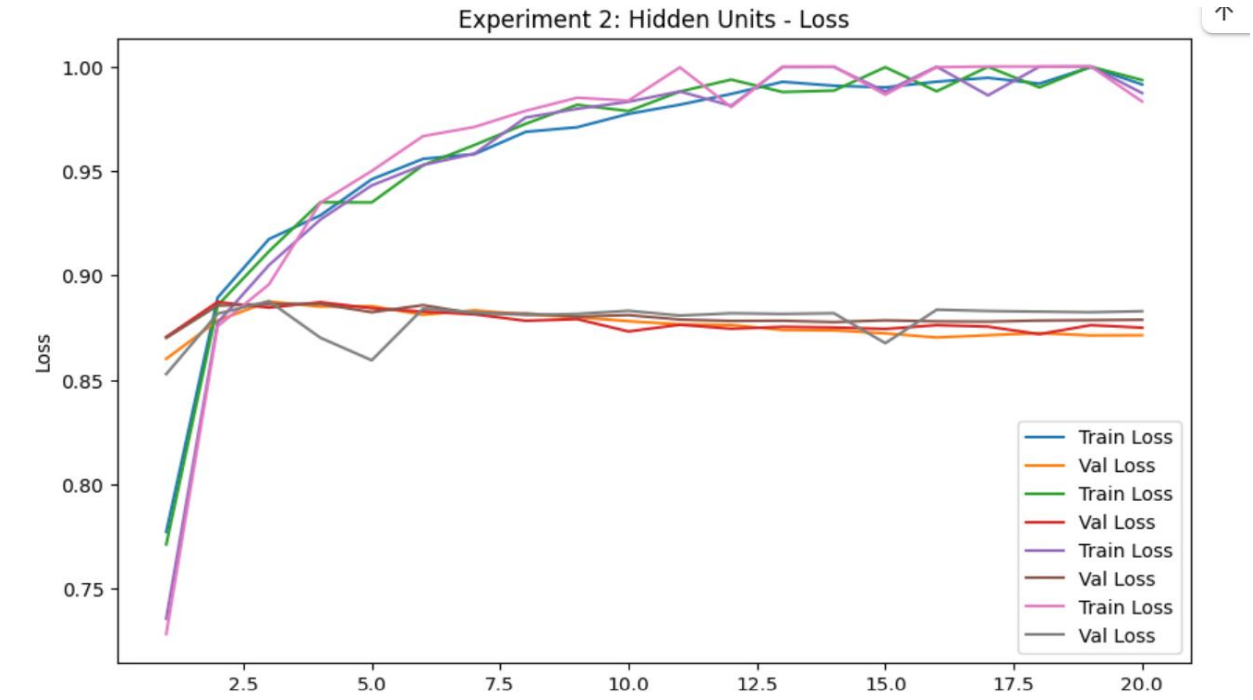
• Hidden layers: Experiment-dependent number of units and layers.

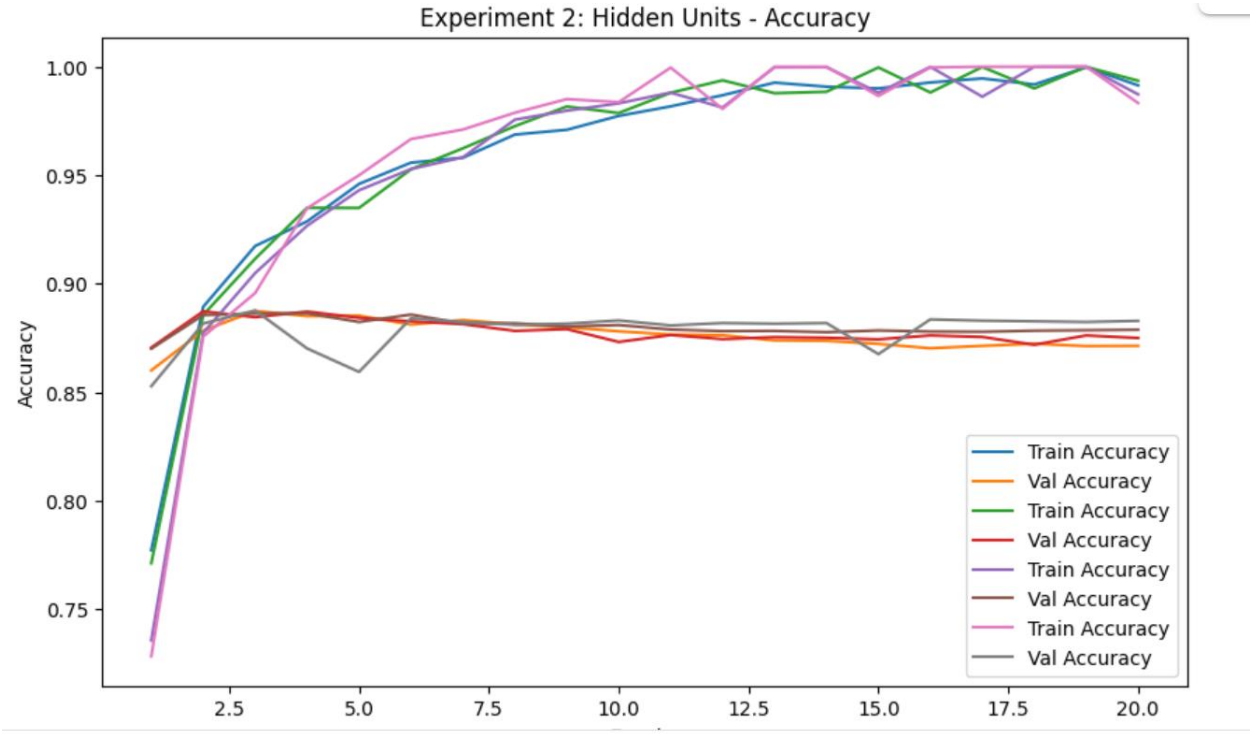
• Output layer: A single unit with sigmoid activation for negative or positive review classification.



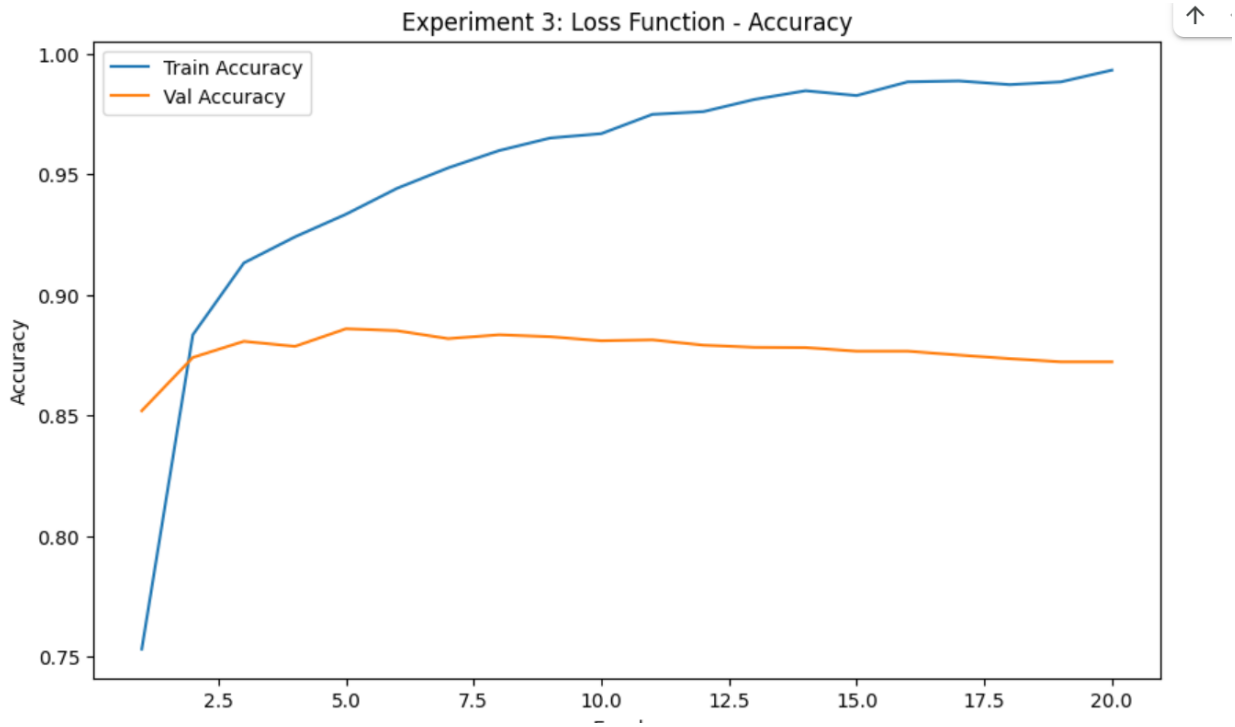


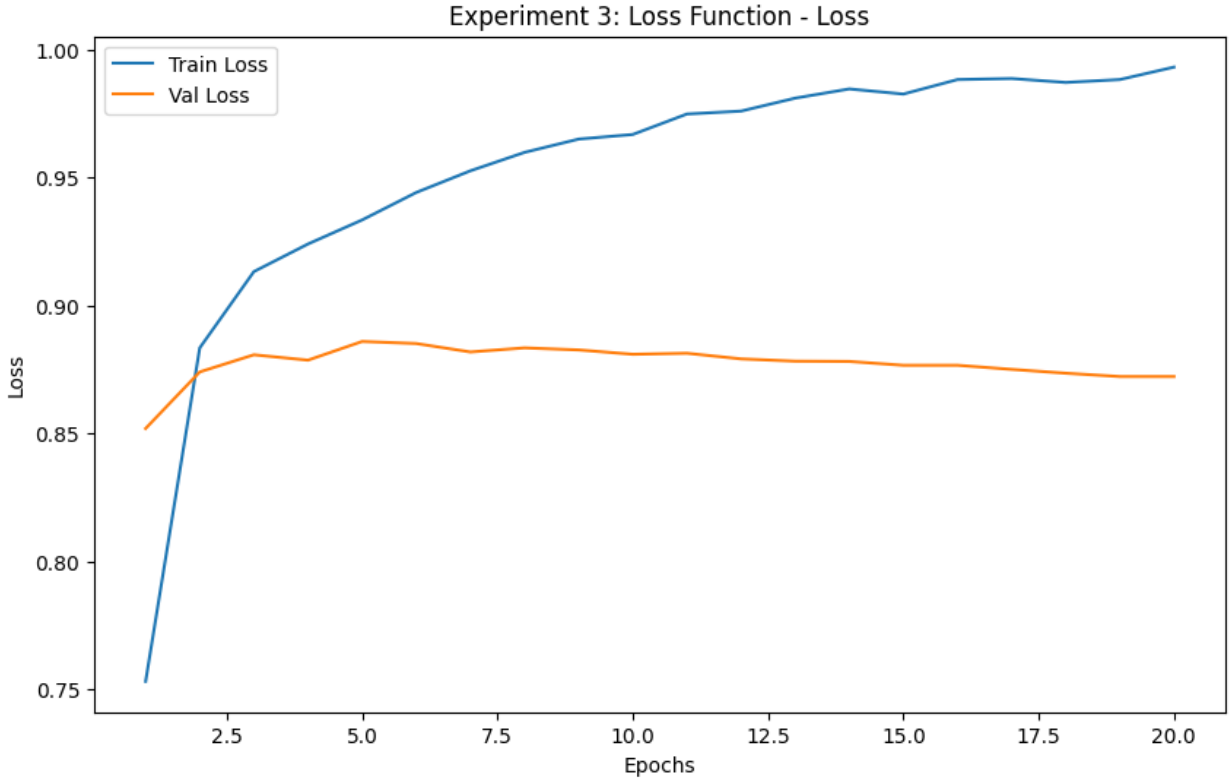
Analysis: The highest performance was achieved by the single hidden layer model, with the best validation accuracy (0.8755) and test accuracy (0.86572). More additional layers did not yield much better performance and, in a few cases, lowered performance slightly. This indicates that when using the IMDB dataset, an increase in network depth would lead to overfitting, and a simpler model with fewer layers gives better performance for this task



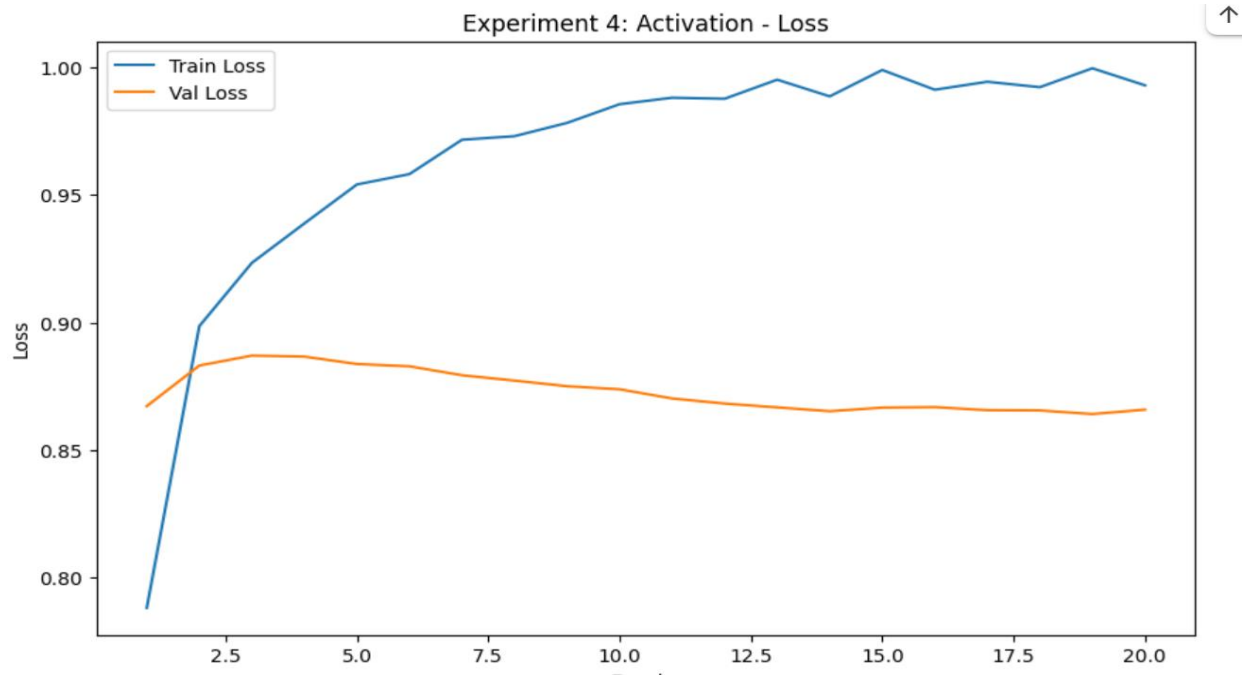


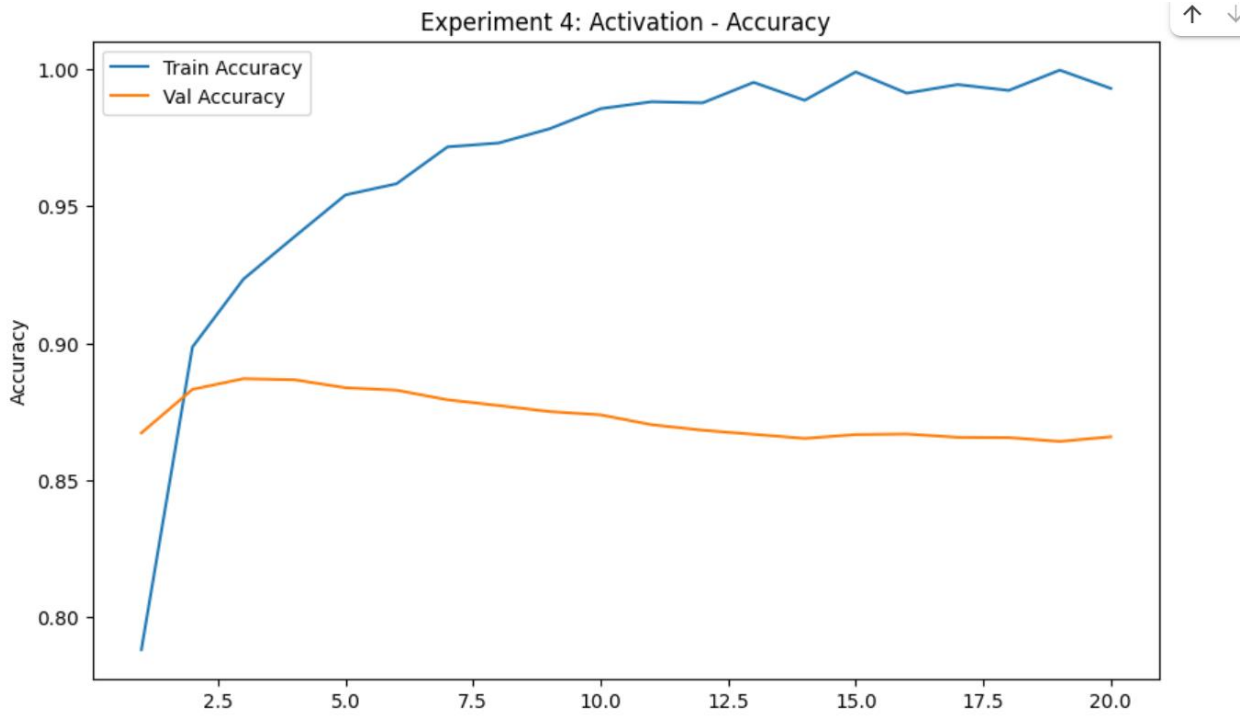
Analysis: Increasing hidden units improves validation and test accuracy, peaking at 256 units with **0.8798** validation and **0.87204** test accuracy. However, larger models risk overfitting, so avoiding unnecessary complexity is best.



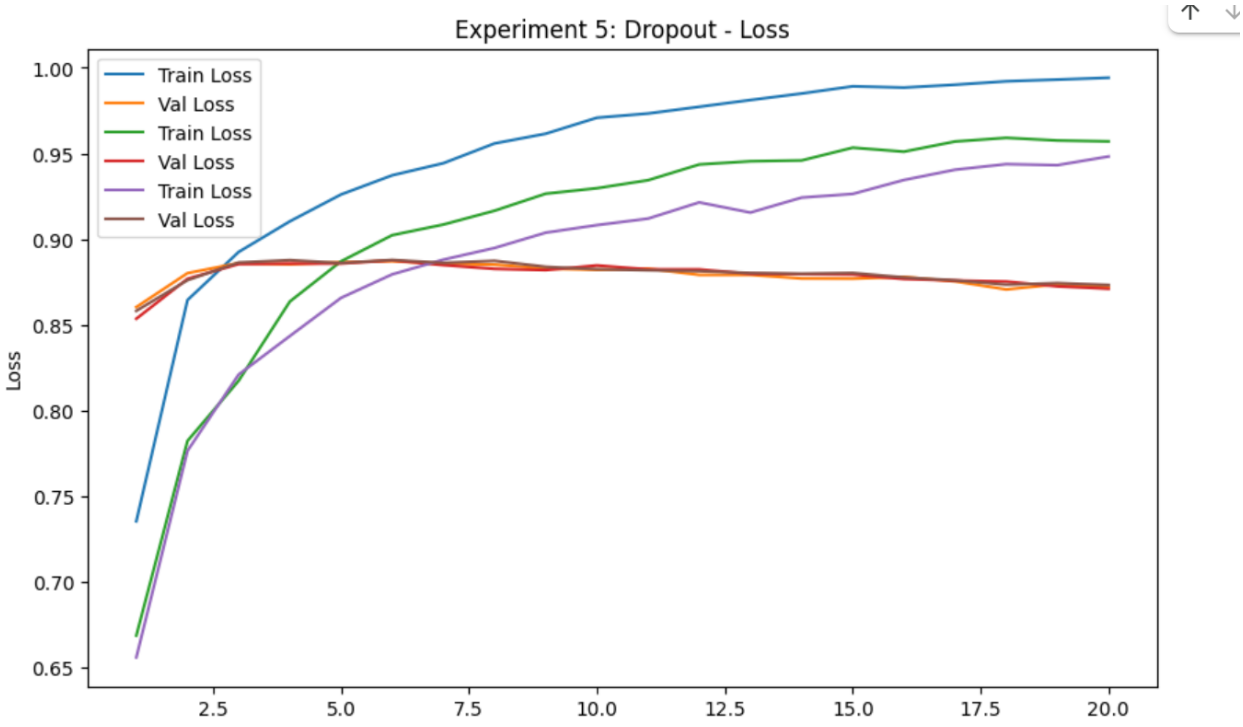


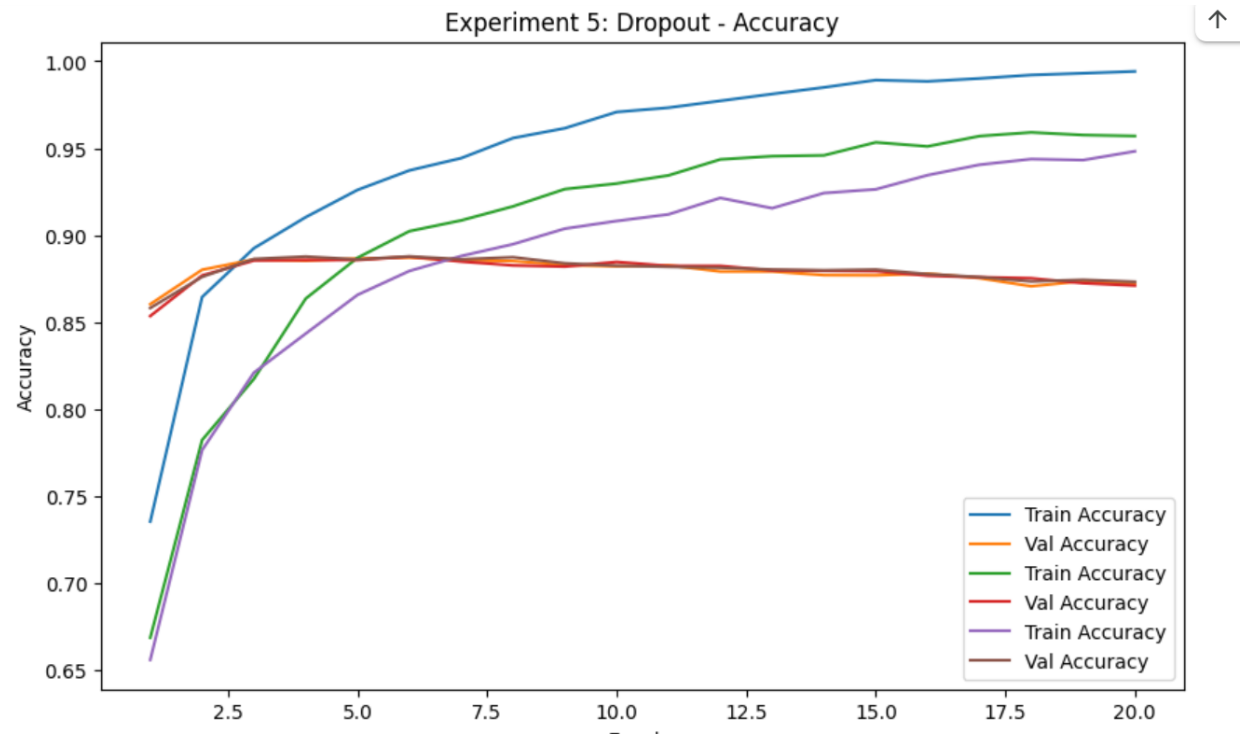
Analysis: Binary Crossentropy has greater validation and test accuracy compared to MSE. This is because Binary Crossentropy is designed for classification problems, and it attempts to minimize the difference between predicted probabilities and actual labels. MSE, which is more suited for regression problems, attacks the classification problem in a regression manner, and therefore, accuracy is lower.





Analysis: Increasing hidden units generally improves validation and test accuracy, with peak performance at 256 units (**0.8798** validation, **0.87204** test accuracy). However, larger models are more prone to overfitting, making it essential to avoid unnecessary complexity.





Analysis: A generally helps in reducing overfitting, rates that are too high may hinder learning; thus, 0.7 seems to be optimal in this case. Overall, all dropout rates show similar

performance, indicating the model's robustness to regularizationdropout rate of 0.7 achieves the highest validation accuracy, suggesting it provides a good balance between model complexity and generalization. While dropout

Optimal Model Selection

- Hidden Layers: 1 Layer
- Hidden Units: 256 Units
- Loss Function: Binary Crossentropy
- Activation Function: ReLU
- Regularization: Dropout Rate of 0.7

Conclusion

The neural network's best fit model is a single hidden layer with 256 units, best balancing complexity and performance, using the Binary Crossentropy loss function, which produced the highest validation and test accuracies, the ReLU activation function, which further improved the model's performance and contributed to an impressive validation accuracy, and a dropout rate of 0.7 for regularization, which effectively minimized overfitting while maintaining strong accuracy metrics. All things considered, this configuration produced a test loss of 0.4936 and a test accuracy of 0.8619, making it the most efficient model for the dataset.