Gabriel Augusto Arsego

Assignment 1 – Neural Network Model Tuning Report

IMDB Sentiment Classification

This report evaluates how different neural network configurations affect sentiment classification performance on the IMDB movie review dataset. Several model variations were tested, including changes in the number of hidden layers, number of hidden units, activation functions, loss functions, and the use of dropout regularization.

The results indicate that simpler models performed as well as or slightly better than more complex ones. Increasing model depth or width did not significantly improve validation accuracy. Binary cross entropy was the most appropriate loss function for this binary classification problem. Overall, this summary shows that increasing complexity does not necessarily improve performance for this dataset.

Baseline Model

The baseline neural network consisted of:

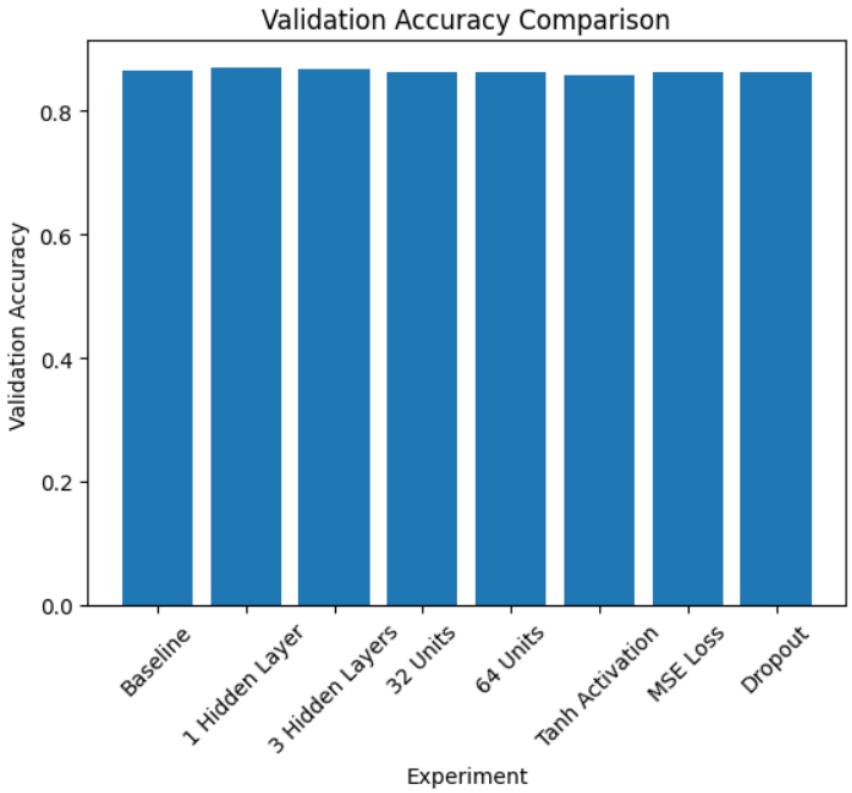
- Embedding layer
- Two hidden Dense layers (16 units each, ReLU activation)
- Sigmoid output layer
- Binary cross entropy loss
- Adam optimizer

These were the base results for validation and test accuracy:

- Validation Accuracy: **0.8660**
- Test Accuracy: **0.8632**

This model was modified 7 times to test other methods.

	Experiment	Validation Accuracy	Test Accuracy
0	Baseline	0.8660	0.86316
1	1 Hidden Layer	0.8698	0.86528
2	3 Hidden Layers	0.8678	0.86200
3	32 Units	0.8632	0.86260
4	64 Units	0.8636	0.86124
5	Tanh Activation	0.8572	0.85776
6	MSE Loss	0.8620	0.85676
7	Dropout	0.8626	0.85828



Analysis and Interpretation

Number of Hidden Layers

Reducing the model to a single hidden layer resulted in the highest validation accuracy (0.8698). Increasing depth to three hidden layers did not improve performance and slightly reduced validation accuracy. This suggests that increasing the complexity of the model was not necessary.

Number of Units

Increasing hidden units to 32 and 64 did not meaningfully improve validation accuracy. In fact, performance slightly decreased at higher unit counts. This indicates that the baseline model was already capable of capturing the underlying structure of the dataset.

Activation Function

Replacing ReLU with tanh reduced performance. The Tanh Activation had the lowest validation accuracy, which makes sense, considering it isn't a popular method anymore.

Loss Function

Using Mean Squared Error (MSE) instead of binary cross entropy resulted in slightly worse validation performance. Binary cross entropy is more appropriate for binary classification problems.

Dropout Regularization

Adding dropout did not significantly improve validation accuracy. This suggests that the baseline model was not overfitting to the point that it would need more regularization.

The most important findings were:

1. Simpler models performed best on this dataset.
2. Increasing depth and width did not improve performance.
3. Binary cross entropy remains the most appropriate loss function.
4. ReLU activation outperformed tanh.
5. Regularization did not significantly enhance generalization.

Conclusion and Recommendation

Based on the results, the best configuration for this dataset is a neural network with one hidden layer using ReLU activation and binary cross entropy loss. Higher complexity did not result in meaningful improvements and, in some cases, reduced performance.

For this classification task, a simpler model works best.