Summary Report

Hyewon Lee

Overview

This report evaluates multiple neural network configurations for the IMDB movie review sentiment classification task. The objective was to identify a model that achieves strong predictive performance while maintaining stable generalization to unseen data.

Final Selected Model

- Hidden layers: 2
- Units: 64
- Activation: ReLU (a nonlinear transformation that improves learning capacity)
- Loss: Binary Crossentropy (designed for two-class classification problems)
- Dropout: 0.5 (a regularization technique that reduces overfitting by randomly disabling neurons during training)
- Epochs: 6
- Validation Accuracy: 0.8927
- Test Accuracy: 0.8810

Modeling Approach

Multiple model configurations were systematically evaluated by varying:

- Network depth (number of hidden layers)
- Number of units per layer
- Activation functions
- Loss functions
- Regularization strength (dropout)
- Training duration

Model selection was based strictly on validation accuracy to ensure fair evaluation and prevent test-set bias.

Key Findings

Across all experiments, models using **ReLU activation and binary cross-entropy loss** consistently outperformed alternatives such as tanh activation and mean squared error loss. The best performing configuration consisted of **two hidden layers** with **64 units per layer**, a **dropout rate of 0.5**, and **six training epochs**, achieving a validation accuracy of 0.8927 and

a test accuracy of 0.8810, demonstrating strong predictive capability on unseen data.

Increasing model complexity beyond this level did not produce meaningful performance improvements. Instead, moderate depth combined with appropriate regularization yielded the most stable and generalizable results. Moreover, validation performance stabilized early while training accuracy continued to improve, indicating that additional training provided limited benefit and increased the risk of overfitting.

For the IMDB sentiment classification task, a moderately sized neural network with regularization is recommended. This configuration balances predictive accuracy with generalization performance, avoiding unnecessary model complexity while maintaining strong classification capability.

Table 1. Top 10 Hyperparameter Configurations (Ranked by Validation Accuracy)

	Rank	Hidden Layers	Units	Activation	Loss Function	Dropout	L2	Epochs	Validation Acc	Test Acc
0	1	2	64	relu	binary_crossentropy	0.5	0	6	0.8927	0.8810
1	2	3	32	relu	binary_crossentropy	0.0	0.0001	5	0.8919	0.8822
2	3	2	32	relu	binary_crossentropy	0.5	0.0001	7	0.8912	0.8852
3	4	1	64	relu	mse	0.5	0	7	0.8911	0.8826
4	5	2	32	relu	mse	0.5	0	7	0.8909	0.8843
5	6	1	32	relu	binary_crossentropy	0.5	0.0001	8	0.8906	0.8843
6	7	2	64	relu	mse	0.5	0	6	0.8905	0.8849
7	8	1	64	relu	mse	0.2	0.0001	6	0.8904	0.8820
8	9	1	64	tanh	binary_crossentropy	0.5	0.0001	6	0.8904	0.8819
9	10	2	32	relu	mse	0.2	0	6	0.8903	0.8828

Table 2. Average Performance by Number of Hidden Layers

	n_hidden_layers	avg_val_acc	avg_test_acc	n_runs
0	1	0.887139	0.879638	72
1	2	0.886231	0.878788	72
2	3	0.885389	0.878861	72

Table 3. Average Performance by Loss Function

	loss	avg_val_acc	avg_test_acc	n_runs
0	binary_crossentropy	0.886470	0.878940	108
1	mse	0.886035	0.879251	108

Table 4. Average Performance by Activation Function

	activation	avg_val_acc	avg_test_acc	n_runs
0	relu	0.887013	0.879910	108
1	tanh	0.885493	0.878281	108

Table 5. Average Performance by Dropout Rate

	dropout_rate	avg_val_acc	avg_test_acc	n_runs
2	0.5	0.887585	0.880826	72
1	0.2	0.885921	0.878379	72
0	0.0	0.885253	0.878081	72