ASSIGNMENT 1 – ADVANCED MACHINE LEARNING

Author: Hemani Panchmatiya
Date: February 14, 2026

Executive Summary:

This report evaluates how different neural network architectures impact performance on the IMDB movie review sentiment classification task. The primary objective was to compare models using different activation functions, loss functions, network depths, and regularization techniques to determine which configuration achieves the best generalization and test accuracy.

Three neural network models were developed and tested. Results show that a simpler model with one hidden layer achieved the highest test accuracy (88.03%). Increasing model depth and complexity (additional layers and units) did not improve performance and, in some cases, led to overfitting. Although regularization techniques such as L2 regularization and dropout improved model stability, they did not significantly outperform the simpler architecture.

These findings highlight that increasing complexity does not necessarily lead to better performance, and well-designed simple models can often generalize more effectively.

Introduction:

This assignment explores how different neural network architectures affect performance on the IMDB binary classification task. The models were modified by changing the number of layers, units, activation functions, and loss functions. Regularization techniques were also applied to control overfitting.

Each model was evaluated using validation and test accuracy, and the results were compared to understand the impact of model complexity on generalization performance.

Methodology:

Three different neural network architectures were implemented and evaluated.

Model 1: One Hidden Layer (Baseline Model):

This model uses one hidden layer with ReLU activation and binary Crossentropy loss, which are well-suited for binary classification of positive and negative reviews. The training and validation plots show that the model learns quickly during the first few epochs. After some time, validation performance stops improving, indicating the potential start of overfitting. Early stopping was applied to halt training at the optimal point and improve generalization on new data.

Model 2: Three Hidden Layers (Increased Depth):

This model increased network depth (three hidden layers with 32 units each) and changed both the activation function (Tanh) and the loss function (Mean Squared Error) to evaluate whether additional complexity improves performance.

After running 20 epochs, training performance continued to improve while validation performance stopped improving after a few epochs. This indicates overfitting, where the model begins memorizing training data instead of learning general patterns. Additionally, Mean Squared Error is not ideal for binary classification problems, which negatively impacted performance. Therefore, increasing depth alone does not guarantee better accuracy and may increase the risk of overfitting.

Model 3: Two Hidden Layers with Regularization:

This model used two hidden layers with 64 units each. ReLU activation and binary Crossentropy loss were applied, which are appropriate for binary classification tasks. Although training accuracy improved, test accuracy did not improve significantly compared to the baseline model.

The training curves show that performance continued to increase while validation performance improved only slightly, indicating possible overfitting due to the increased number of parameters. To reduce overfitting, L2 regularization and dropout were implemented. L2 regularization helps control large weight values, while dropout randomly deactivates neurons during training, encouraging the model to learn more generalized patterns instead of memorizing training data.

These techniques helped stabilize validation performance and improve generalization. However, the model still did not outperform the simpler baseline architecture.

Test Result Accuracy:

Model	Architecture	Test Accuracy
Model 1	1 Hidden Layer (ReLU, BCE)	**88.03%**
Model 2	3 Hidden Layers (Tanh, MSE)	**84.74%**
Model 3	2 Hidden Layer (ReLU, BCE)	**87.74%**

Findings:

a) **Simpler Model Generalized Better:** The baseline model achieved the highest test accuracy and showed stable validation performance.
b) **Increasing Depth Did Not Improve Accuracy:** Additional layers and units increased training accuracy but did not improve test performance, indicating overfitting.
c) **Loss Function Matters:** Using Mean Squared Error instead of Binary Crossentropy reduced classification performance. Binary Crossentropy is more appropriate for binary classification tasks.
d) **Regularization Helps but Does Not Guarantee Improvement:** L2 regularization and dropout reduced overfitting but did not significantly outperform the simpler architecture.

Conclusion: The results demonstrate that the simple one-hidden layer model performed best overall. Increasing the number of layers and units did not significantly improve test accuracy. Using an inappropriate loss function (MSE) reduced performance. Larger models showed signs of overfitting, while regularization and dropout helped control complexity.

These findings confirm that simpler models can often generalize better than more complex architecture. Model selection should focus on balancing performance and complexity rather than assuming that deeper networks will always yield better results.