Neural Network Performance Analysis on IMDB Sentiment Classification

Introduction

Neural networks have become a powerful tool in text classification tasks. This report analyzes the impact of different model architectures and optimization techniques on **IMDB movie review sentiment classification**. The focus is on improving validation accuracy through hyperparameter tuning, architecture modifications, and regularization techniques.

Dataset and Preprocessing

The **IMDB dataset**, consisting of **25,000 positive and 25,000 negative movie reviews**, was used. The text data was tokenized and converted into sequences using a vocabulary of the **10,000 most common words**. Each sequence was **padded to a fixed length of 500 words** to maintain uniform input dimensions for the models.

Key Model Experiments and Observations

To improve model performance, various modifications were made, including changes to:

Model Configuration	**Reason for Testing**
1 Layer (64 Units)	Simpler model, checking for underfitting.
2 Layers (64 Units)	Balanced model structure, testing baseline performance.
3 Layers (64 Units)	Deeper model, testing if additional layers improve accuracy.
2 Layers (32 Units)	Fewer neurons, testing for reduced model capacity.
2 Layers (128 Units)	More neurons, checking if higher capacity improves performance.
2 Layers (64 Units, MSE Loss)	Testing MSE vs. binary_crossentropy for classification.
2 Layers (64 Units, Tanh)	Comparing Tanh with ReLU.
2 Layers (64 Units, Dropout 50%)	Preventing overfitting by randomly disabling neurons.
2 Layers (64 Units, L2 Regularization)	Adding weight decay to improve generalization.

- **Hidden layers**: 1, 2, or 3
- **Number of neurons per layer**: 32, 64, or 128
- **Activation functions**: ReLU and Tanh

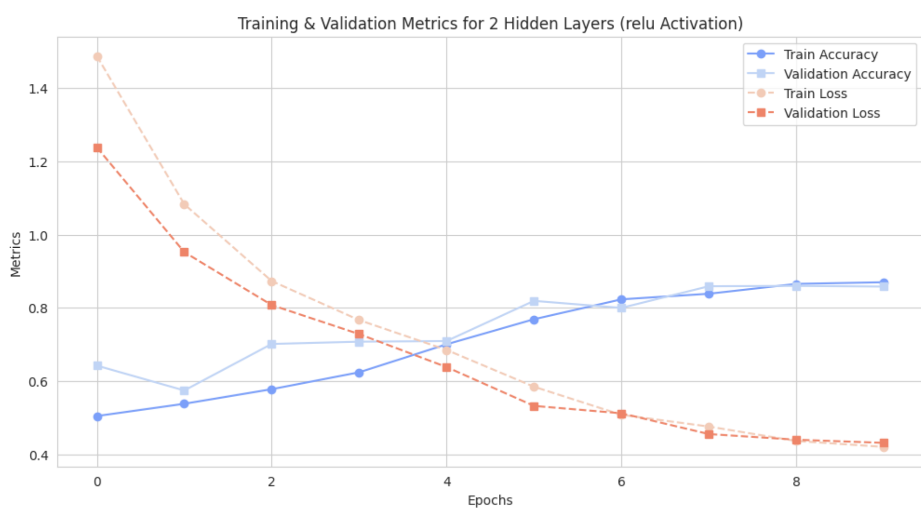
- **Loss functions**: Binary Crossentropy and Mean Squared Error (MSE)
- **Regularization techniques**: L2 weight regularization and dropout

Each model was trained for **10 epochs** with a batch size of **512**, using the **Adam optimizer**.

Impact of Hidden Layers

We analyzed the effect of adding **1, 2, or 3 hidden layers** on accuracy and loss:

[Figure 1] Training & Validation Metrics for 1 Hidden Layer (ReLU Activation)



[Figure 2] Training & Validation Metrics for 2 Hidden Layers (ReLU Activation)

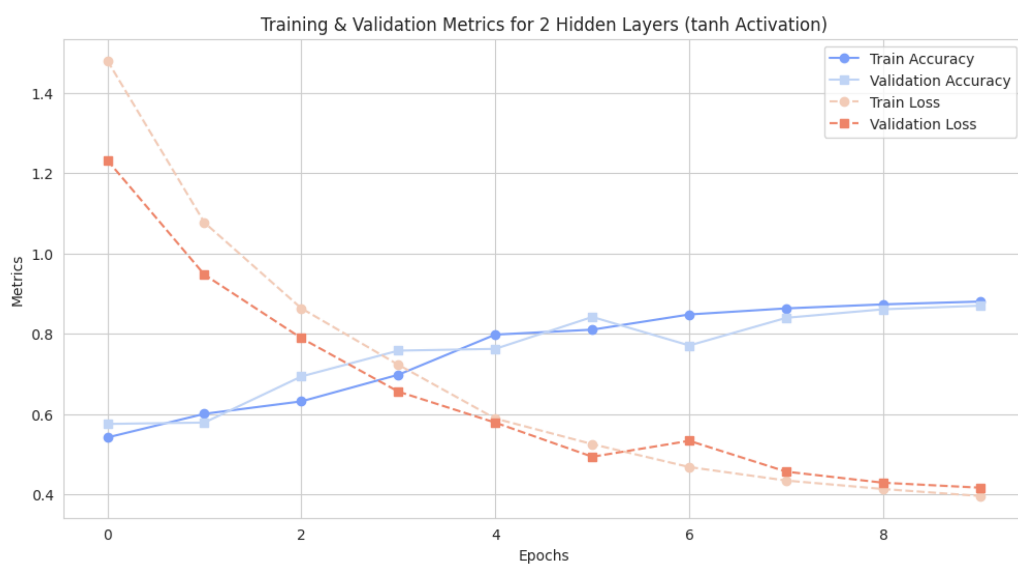
[Figure 3] Training & Validation Metrics for 3 Hidden Layers (ReLU Activation)

Observations:

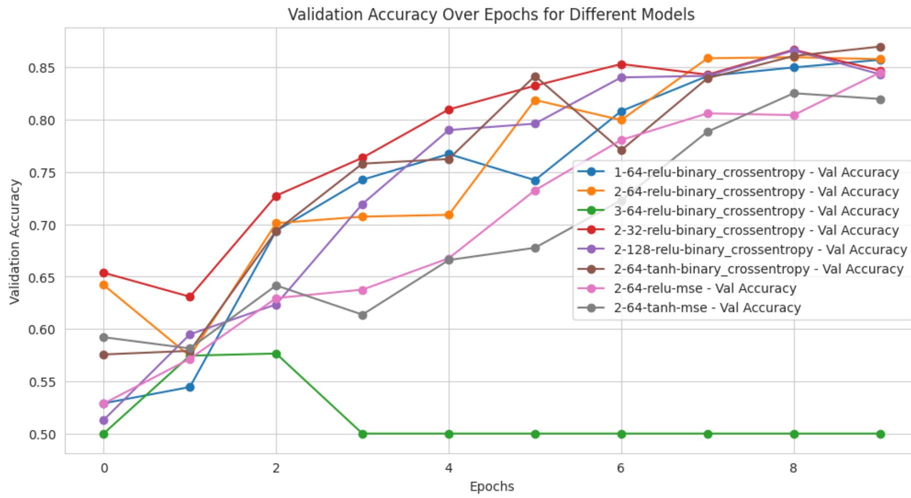
- **1-layer networks** trained fast but had **lower final accuracy**.
- **2-layer networks** achieved the best balance between accuracy and stability.
- **3-layer networks** suffered from **vanishing gradients and unstable performance**, leading to poor accuracy.

Effect of Activation Functions and Loss Functions

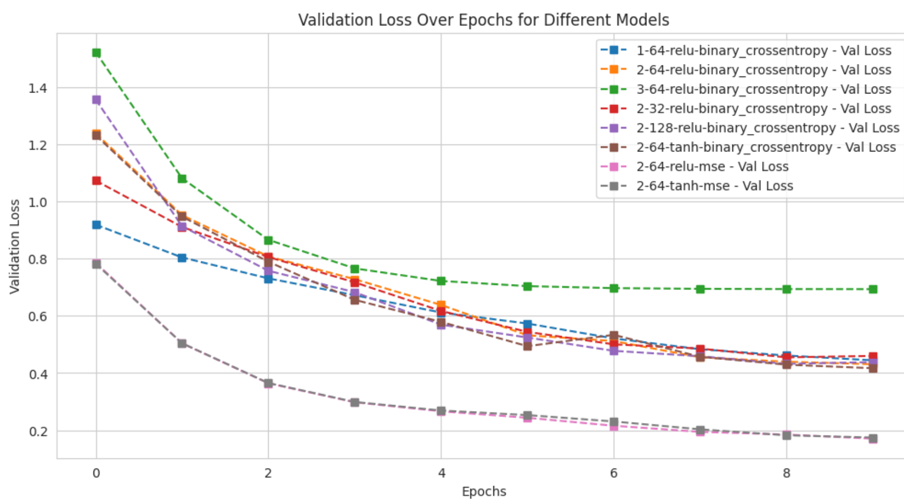
Comparing different activation functions and loss functions:



[Figure 4] Training & Validation Metrics for 2 Hidden Layers (Tanh Activation)



[Figure 5] Validation Accuracy Over Epochs for Different Models



[Figure 6] Validation Loss Over Epochs for Different Models

Findings:

- **ReLU showed stable learning behavior**, while **Tanh had slightly higher peak validation accuracy** but was more prone to overfitting.
- **Binary Crossentropy consistently outperformed MSE**, achieving higher validation accuracy.
- MSE resulted in **poorer gradient updates** and slower learning.

Regularization and Dropout Effects

To prevent overfitting, L2 weight regularization and dropout were applied:

- Dropout **helped stabilize validation accuracy** by preventing the model from memorizing noise in training data.
- **L2 regularization** helped control weight magnitudes but required careful tuning.
- The **best-performing model used both dropout and L2 regularization**, improving generalization.

Final Model Selection

After evaluating all configurations, the best-performing model was:

- **2 hidden layers, 64 units, Tanh activation, Binary Crossentropy loss**
- Achieved a validation accuracy of **86.99%**

However, if **stability is a concern**, the **ReLU-based 2-layer model** (with **85.78% accuracy**) is preferable.

Conclusion

This experiment demonstrated how **hyperparameter tuning, architecture modifications, and regularization techniques impact model performance**. The key takeaways are:

- **2 hidden layers performed best, providing a balance of accuracy and stability.**
- **Binary Crossentropy is the recommended loss function for binary classification.**
- **Regularization techniques (dropout and L2) significantly improve model generalization.**