Hyperparameter Tuning Summary Report for IMDB Analysis

1. Introduction

A neural network model obtains results through its hyperparameter optimization process for sentiment analysis of IMDB data. Optimizing network architecture together with activation functions and loss functions and regularization methods served as the objective to find the perfect balance between accuracy and generalization performance.

2. Model Configurations and Performance

We tested multiple combination of architectural designs and hyperparameter settings to study their effects on modeling performance results. The results appear in the following tabular overview:

Model Type	Hidden Layers	Units per Layer	Activation	Loss Function	Regularization	Accuracy (%)	Loss
Baseline	2	16	ReLU	Binary Crossentropy	None	69	0.50
One Hidden Layer	1	16	ReLU	Binary Crossentropy	None	69	0.50
Three Hidden Layers	3	16	ReLU	Binary Crossentropy	None	70	0.49
32 Units per Layer	2	32	ReLU	Binary Crossentropy	None	71	0.49
64 Units per Layer	2	64	ReLU	Binary Crossentropy	None	69	0.49
128 Units per Layer	2	128	ReLU	Binary Crossentropy	None	70	0.50
MSE Loss Function	2	16	ReLU	Mean Squared Error	None	68	0.52
Tanh Activation	2	16	Tanh	Binary Crossentropy	None	70	0.50
Adam Optimizer	2	16	ReLU	Binary Crossentropy	None	72	0.48
L2 Regularization	2	16	ReLU	Binary Crossentropy	L2(0.001)	75	0.47

Dropout (50%)	2	16	ReLU	Binary Crossentropy	Dropout(0.5)	74	0.48
Hypertuned Model	3	32, 32, 16	ReLU	MSE	L2(0.0001) + Dropout(0.5)	75	0.46

- Varying **Hidden Layers:** Using one or three hidden layers showed that increasing model complexity didn't enhance performance. The model overfitted, with training accuracy rising while validation accuracy declined, indicating it was too complex for the task.
- Adjusting **Hidden Units:** Experimenting with different unit sizes (32, 64, etc.) revealed that adding more units led to overfitting. While training loss decreased, validation accuracy peaked early and then dropped, suggesting the model memorized data instead of generalizing.
- Using **MSE Loss Function:** Replacing binary_crossentropy with MSE initially showed stable loss trends, but validation accuracy declined after four epochs, highlighting issues with generalization.
- Tanh **vs. ReLU Activation:** The model with Tanh activation peaked early in validation accuracy before declining, whereas ReLU resulted in larger fluctuations and higher validation loss, making it potentially less suitable for this task.
- Applying **Regularization & Dropout:** Regularization and dropout improved generalization. Dropout, in particular, enhanced accuracy while maintaining stable validation performance, effectively reducing overfitting.

3. Key Observations and Findings

- Increasing the number of hidden layers slightly improved accuracy but also increased complexity.
- Increasing the number of units per layer from 16 to 32 improved accuracy (from 69% to 71%), but further increases (64 or 128 units) did not yield better results.
- Using the Mean Squared Error (MSE) loss function instead of Binary Crossentropy resulted in lower accuracy (68%).
- Switching from ReLU to Tanh activation function did not significantly impact performance.
- The Adam optimizer provided slightly better performance (72%) compared to RMSprop.
- Regularization techniques such as L2 and Dropout significantly improved generalization, achieving up to 75% accuracy.
- The hypertuned model, which combined L2 regularization, dropout, and a carefully chosen architecture, achieved the best performance with 75% accuracy and a loss of 0.46.

4. Conclusion

The most effective model implemented L2 regularization through dropout while selecting the right number of hidden units and layers. The particular configuration produced maximum accuracy levels together with optimized prevention of overfitting. Further research should test diverse learning rates together with batch normalization methods and alternative network architectures including CNNs and transformers to achieve maximum improvements.

5. Recommendations

- The implementation of L2 regularization with dropout helps to improve generalization capabilities.
- The performance improved minimally when network complexity surpassed a specific level through the addition of extra hidden layers and units.

- The preferred loss function for classification becomes binary cross-entropy loss rather than MSE for maximum accuracy.
- Additional optimization research and learning rate schedule exploration will help boost system performance.