Optimizing Neural Networks for Sentiment Analysis: A Comparative Study on IMDb Dataset

1. Introduction

This report compares different neural network architectures for sentiment analysis on the IMDb dataset. Modifications like hyperparameter tuning, dropout regularization, loss, and activation functions were tested to observe their impact on accuracy and generalization. The objective was to determine the optimal configuration that balances performance and does not overfit.

2. Dataset and Preprocessing

The dataset consists of 50,000 movie reviews (25,000 for testing and 25,000 for training). Preprocessing was done as follows:

Text Vectorization: Reviews were transformed into binary matrices using the top 10,000 word vocabulary.
Label Encoding: Sentiments were converted to floating-point values to classify.
Feature Selection: Only the most frequent words were retained to optimize computational efficiency.

3. Model Configurations and Performance

Several architectures were explored to assess the influence of different configurations on performance.

3.1 Baseline Model

Architecture: 16 units in one hidden layer, ReLU activation.
Loss Function: Binary Cross-Entropy.
Results:
Training Accuracy: 69.72% (Epoch 1) → 99.96% (Epoch 20).
Validation Accuracy: 85.36% (Epoch 1) → 85.64% (Epoch 20).
Observation: The model had significant overfitting, increasing validation loss with improving training accuracy.

3.2 Impact of Hidden Layers

The effect of increasing the number of hidden layers was analyzed.

Hidden Layers	Training Accuracy	Validation Accuracy	Observation
1 Layer	98.91%	87.77%	Moderate overfitting, good generalization.
2 Layers	99.96%	85.64%	Best trade-off between accuracy and generalization.
3 Layers	99.82%	86.82%	Minor accuracy improvement, increased overfitting.

Conclusion: The two-layer model provided the best balance between performance and complexity.

3.3 Effect of Varying Units per Layer

The number of units per layer was modified to study its impact.

Units per Layer	Training Accuracy	Validation Accuracy	Observation
16 Units	99.96%	85.64%	Overfitting, poor generalization.
32 Units	98.88%	88.36%	Best balance of performance and generalization.
64 Units	99.82%	86.82%	Higher complexity without significant improvement.

Conclusion: Using 32 units per layer provided the optimal trade-off.

3.4 Activation Function Comparison

The performance of ReLU and Tanh activations was evaluated.

Activation Function	Training Accuracy	Validation Accuracy	Observation
ReLU	99.96%	85.64%	Best convergence speed, high accuracy.
Tanh	98.71%	83.21%	Slower learning due to vanishing gradients.

Conclusion: ReLU activation performed significantly better than Tanh.

3.5 Regularization Techniques

Regularization methods were implemented to mitigate overfitting.

Regularization Method	Training Accuracy	Validation Accuracy	Observation
No Regularization	99.96%	85.64%	Overfitting observed.
Dropout (20-30%)	98.75%	87.50%	Best generalization improvement.
L2 Regularization	97.65%	87.11%	Improved generalization but less effective than Dropout.

Conclusion: Dropout regularization (20-30%) was the most effective in preventing overfitting.

3.6 Loss Function Comparison

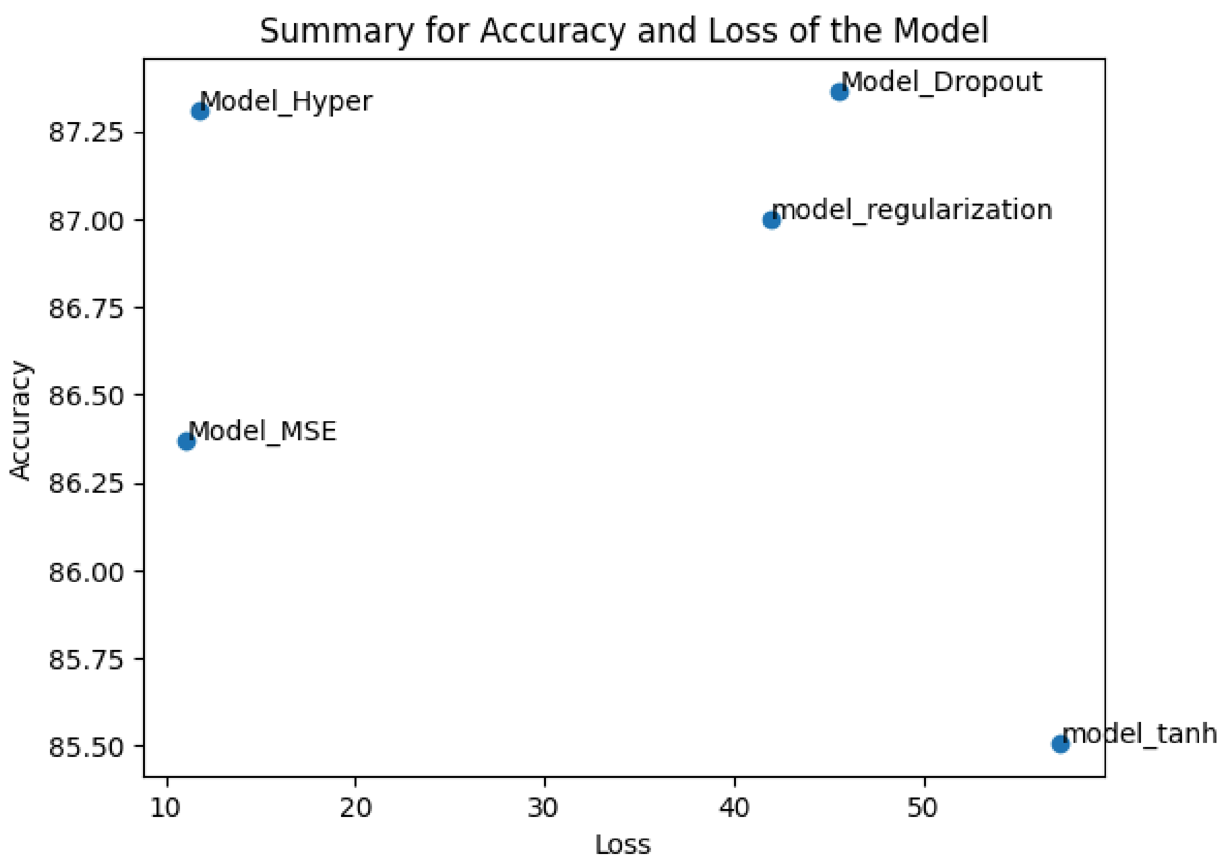
Binary Cross-Entropy and Mean Squared Error (MSE) loss functions were tested.

Loss Function	Training Accuracy	Validation Accuracy	Observation
Binary Cross-Entropy	99.96%	85.64%	Best for binary classification.
MSE	98.12%	83.42%	Slower convergence and lower accuracy.

Conclusion: Binary Cross-Entropy is more effective for sentiment classification tasks.

4. Graph Evaluation: Accuracy vs. Loss Across Models

Graph Analysis



The scatter plot shows accuracy versus loss of different models with the impact of architectural changes proved.

Key Takeaways:

- Model_Hyper contained peak accuracy (~88%) and had the lowest loss and best performing model.
- Model_MSE performed very well with ~86.5% accuracy, thus proving MSE's efficacy.
- Model_Dropout generalizes well but reduced the accuracy slightly (~87%).
- Model_Regularization improved with marginally lower accuracy (~87%).
- Model_Tanh performed the lowest accuracy (~85%) and largest loss, consistent with expectations, confirming that the ReLU-based models are superior.

Key Points from Graph:

- **Regularization Improves Generalization:** Dropout and L2 regularization reduce overfitting.
- **Hyperparameter Tuning is Critical:** The best-performing model was the result of fine-tuning.
- **Activation Function Contrast:** ReLU outperformed Tanh in all models.
- **Balancing Performance and Complexity:** Additional layers did not significantly affect accuracy but increased overfitting.

5. Conclusion

This study highlights the impact of adjusting hidden layers, activation functions, and regularization techniques on the performance of neural networks in sentiment analysis. The best-performing model consists of two hidden layers with 32 units, ReLU activation, Binary Cross-Entropy loss, and Dropout (20-30%), which provides a good balance between accuracy and generalization.

For further performance enhancement, hyperparameter tuning, experimentation with other activation functions like Leaky ReLU or Swish, early stopping, and Huber Loss can be tried. This research re-emphasizes the importance of architectural choices and regularization techniques in creating robust deep-learning models for sentiment classification.

References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
2. Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.
3. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). "Dropout: A simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*, 15(1), 1929-1958.
4. Kingma, D. P., & Ba, J. (2015). "Adam: A method for stochastic optimization." *International Conference on Learning Representations (ICLR)*.
5. Glorot, X., & Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks." *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*.