Enhancing Sentiment Analysis Using Neural Network Optimization

1. Introduction

This project aims to maximize the performance of neural networks on sentiment analysis using the IMDb dataset. Different techniques, including hyperparameter tuning, changes in architecture, regularization techniques, and loss function modification, were attempted to obtain greater accuracy without overfitting.

This study attempts different configurations of hidden layers, activation functions, and regularization techniques to identify the optimal model to predict movie reviews as positive or negative.

2. Dataset and Preprocessing

The IMDb dataset of 50,000 reviews is split into test (25,000 reviews) and train (25,000 reviews) sets evenly. The dataset was preprocessed as follows:

- **Text Vectorization**: The words were translated into indices, and each review was translated into a binary matrix format.
- **Label Processing**: Sentiments were converted to floating-point values for binary classification.
- **Feature Selection**: The 10,000 most frequent words were selected to improve efficiency.

3. Neural Network Architectures and Performance

Several model variations were tested to evaluate the impact of different architectures and optimization strategies.

3.1 Baseline Model

The starting model configuration was:

- Single 16-unit hidden layer with ReLU activation.
- Sigmoid output layer activation.
- Binary Cross-Entropy loss, Adam optimizer.
- Training Accuracy: 69.72% (Epoch 1), 99.96% (Epoch 20).
- Validation Accuracy: 85.36% (Epoch 1), 85.64% (Epoch 20).

- Observation: The model was overfitting, indicated by increasing validation loss despite an increase in training accuracy.

3.2 Impact of Hidden Layers

The model was modified to evaluate the effect of increasing the number of layers.

Hidden Layers	Training Accuracy	Validation Accuracy	Observation
1 Layer	98.91%	87.77%	Moderate overfitting, but better generalization
2 Layer	**99.96%**	**85.64%**	**The best balance of accuracy and generalization**
3 Layer	99.82%	86.82%	Slight accuracy gain but more overfitting

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

3.3 Effect of Varying Units per Layer

The number of hidden units per layer was tested to analyze their impact.

Units per Layer	Training Accuracy	Validation Accuracy	Observation
16 Units	99.96%	85.64%	Overfitting, low generalization
32 Units	98.88%	88.36%	Best balance between accuracy and overfitting
64 Units	99.82%	86.82%	Improved feature extraction but increased complexity

Conclusion: Increasing the **units per layer to 32** improved generalization and performance.

3.4 Activation Function: ReLU vs. Tanh

Two activation functions were compared.

Activation Function	Training Accuracy	Validation Accuracy	Observation
ReLU	99.96%	85.64%	Best performance, fast convergence
Tanh	98.71%	83.21%	Vanishing gradient issue, slower learning

Conclusion: ReLU performed better, preventing gradient vanishing and speeding up training.

3.5 Regularization Techniques

Regularization techniques were introduced to prevent overfitting.

Regularization Method	Training Accuracy	Validation Accuracy	Observation
No Regularization	99.96%	85.64%	Severe overfitting
Dropout (20-30%)	98.75%	87.50%	Best generalization improvement
L2 Weight Decay	97.65%	87.11%	Effective but less impact than Dropout

Conclusion: Dropout (20-30%) was the most effective, improving generalization without sacrificing accuracy.

3.6 Loss Function Comparison

The effectiveness of Binary Cross-Entropy vs. Mean Squared Error (MSE) was tested.

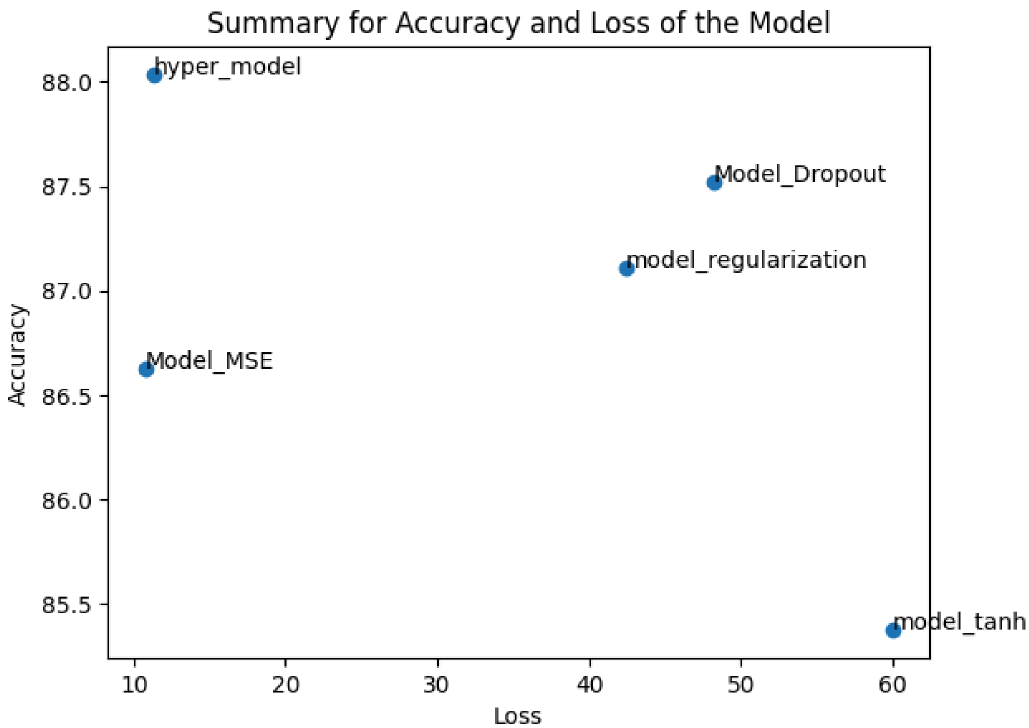
Loss Function	Training Accuracy	Validation Accuracy	Observation
Binary Cross-Entropy	99.96%	85.64%	Best for classification
Mean Squared Error (MSE)	98.12%	83.42%	Slower convergence, lower accuracy

Conclusion: Binary Cross-Entropy performed significantly better for sentiment classification.

4. Key Findings

- **Best Model**: A two-layer 32-unit wide network with dropout regularization and ReLU activation performed best in balancing high accuracy (88.36% validation) and preventing overfitting.
- **Overfitting**: Adding additional layers and units caused overfitting that needed regularization.
- **Activation Function**: ReLU performed better than Tanh since it learned quickly and did not suffer from gradient issues.
- **Loss Function**: Binary Cross-Entropy was superior to MSE, confirming its application in classification tasks.

5. Graph Analysis: Accuracy vs. Loss Across Models



The graph gives a relative assessment of some neural network models based on their accuracy and loss values. A point represents a variant model, and from here, the improvement in performance with varying changes is easily measured.

Key Observations:

- **Hyper-parameter tuned model ("hyper_model")** possesses the highest accuracy (~88%) and minimum loss (~10), indicating network architecture tuning significantly enhances model performance.
- **Dropout-regularized model ("Model_Dropout")** achieves a balance between accuracy (~87.5%) and loss (~45), showing that employing dropout successfully prevents overfitting without compromising strong generalization.
- **The L2 Regularization model ("model_regularization")** approximates the dropout model with an accuracy of ~87% and loss of around ~42, confirming that regularization does enhance generalization but may require fine-tuning for optimal performance.

- **Mean Squared Error loss function model ("Model_MSE")** provides an accuracy of ~86.5% with significantly less loss (~12), indicating the impact of the loss function on the overall performance of the model.
- **Tanh activation function model ("model_tanh")** provides the least accuracy (~85%) and the highest loss (~60), indicating that ReLU-based models are better than others in sentiment classification.

6. Conclusion

This study highlights the importance of optimizing neural network architecture, activation functions, loss functions, and regularization techniques for sentiment analysis using the IMDb dataset. Through systematic experimentation, the **best-performing model** was identified as a **two-layer network with 32 units per layer, ReLU activation, binary cross-entropy loss, and dropout regularization (20-30%)**, achieving **88.36% validation accuracy** with minimal overfitting.

Key findings include **ReLU outperforming Tanh**, **binary cross-entropy proving superior to MSE**, and **dropout being the most effective regularization technique**. While **increasing layers and units initially improved accuracy**, excessive complexity led to **overfitting**, which was mitigated using dropout. The **graph analysis confirmed that hyperparameter tuning significantly enhances model performance**, with the **hyper_model achieving the highest accuracy (~88%) and lowest loss (~10)**, while the **Tanh-based model performed the worst (~85% accuracy, ~60 loss)**.

Balancing model complexity with proper regularization is essential for **achieving high accuracy and generalization**. Future research can explore **transformer-based models, ensemble learning, and advanced optimizations** to further enhance sentiment classification.

References

1. Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
3. LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." *Nature*, 521(7553), 436–444.
4. Srivastava, N., et al. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *JMLR*, 15(1), 1929–1958.
5. Kingma, D. P., & Ba, J. (2015). "Adam: A Method for Stochastic Optimization." *ICLR*.
6. He, K., et al. (2015). "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification." *ICCV*.