Assignment 2 Report

Evaluation of Neural Network Performance

Name: Vikram kumar Bonagiri
Student ID: 811368127
Email: vbonagir@kent.edu

I. Introduction

This report presents the evaluation of multiple neural network configurations for sentiment analysis using the IMDB movie reviews dataset. The experiments aimed to explore how different model architectures, activation functions, loss functions, and regularization methods affect model performance. The dataset was preprocessed using multi-hot encoding to convert integer sequences into binary vectors, and models were trained for 20 epochs using an 80-20 training-validation split.

II. Performance of the Baseline Model

Baseline Model Configuration:

- 2 Hidden Layers with 32 Units Each
- Activation: ReLU for hidden layers, Sigmoid for output layer
- Loss Function: Binary Cross-Entropy
- Optimizer: RMSProp
- Epochs: 20, Batch Size: 512

Results:

- Training Accuracy: ~99.9%
- Validation Accuracy: ~86%
- **Observations:**
  The baseline model achieved high training accuracy but showed a steady rise in validation loss, indicating overfitting. Despite this, it reached solid

generalization with around 86% accuracy, serving as a good reference point for comparison.

III. Model Variations and Results

1. Change in Number of Hidden Layers

1.1 One Hidden Layer

- Best Validation Accuracy: ~87.9%
- Final Validation Accuracy: ~86.4%
- Validation Loss Trend: Gradually increases after early epochs.
- Observations:
  With fewer layers, the model was simpler and exhibited less overfitting. Accuracy plateaued early, showing the model reached its learning limit quickly.

1.2 Three Hidden Layers

- Best Validation Accuracy: ~87.3%
- Final Validation Accuracy: ~85.4%
- Validation Loss Trend: Increases significantly with epochs.
- Observations:
  Additional hidden layers increased model complexity and training accuracy but hurt validation performance due to overfitting.

Comparison Table:

Model	Best Val Accuracy	Overfitting	Validation Loss Trend
1 Hidden Layer	~87.9%	Less	Gradually Increases
2 Hidden Layers (Baseline)	~87.6%	Yes	Steady Increase
3 Hidden Layers	~87.3%	Severe	Sharp Increase

2. Change in Number of Units per Layer

2.1 32 Units Per Layer

- Best Validation Accuracy: ~87.6%
- Final Validation Accuracy: ~85.9%
- Validation Loss Trend: Increasing over epochs.
- Observations:
  A moderate number of neurons provided a balance between learning capacity and generalization.

2.2 64 Units Per Layer

- Best Validation Accuracy: ~86.6%
- Final Validation Accuracy: ~85.6%
- Validation Loss Trend: Sharp increase after mid-training.
- Observations:
  Increasing the number of neurons did not improve validation accuracy but led to greater overfitting.

Comparison Table:

Model	Best Val Accuracy	Final Val Accuracy	Validation Loss Trend	Overfitting
32 Units	87.6%	85.9%	Increases	Yes
64 Units	86.6%	85.6%	Sharp Increase	Yes (More)

3. Change in Loss Function (MSE vs. BCE)

Mean Squared Error (MSE) Loss

- Best Validation Accuracy: ~87.7%
- Final Validation Accuracy: ~86.4%

- Validation Loss Trend: Slightly more stable than BCE.
- Observations:
  Using MSE made the training more stable but did not significantly boost performance. BCE remained more suitable for binary classification.

Comparison Table:

Loss Function	Best Val Accuracy	Final Val Accuracy	Overfitting	Observation
Binary Cross-Entropy	87.6%	86.1%	Yes	Standard for binary tasks
Mean Squared Error	87.7%	86.4%	Slight	More stable but not better

4**. Change in Activation Function (tanh vs. ReLU)**

tanh Activation

- Best Validation Accuracy: ~86.8%
- Final Validation Accuracy: ~85.3%
- Validation Loss Trend: Gradual increase with epochs.
- Observations:
  The tanh activation performed slightly worse than ReLU. It converged slower and had a lower generalization capacity.

Comparison Table:

Activation	Best Val Accuracy	Final Val Accuracy	Overfitting	Observation
ReLU	87.6%	86.1%	Yes	Performs best
tanh	86.8%	85.3%	Yes	Slower convergence

5. Applying Regularization Techniques (Dropout 0.5)

With Dropout (0.5)

- Best Validation Accuracy: ~87.7%
- Final Validation Accuracy: ~86.3%
- Validation Loss Trend: Increases less sharply.
- Observations:
  Dropout helped reduce overfitting by preventing the network from memorizing training data. It improved stability and generalization across epochs.

Comparison Table:

Model	Best Val Accuracy	Final Val Accuracy	Validation Loss Trend	Overfitting
Baseline	87.6%	86.1%	Sharp Increase	Yes
With Dropout (0.5)	87.7%	86.3%	Moderate Increase	Reduced

IV. Summary and Conclusion

Overall Findings:

- Most models achieved between 85–88% validation accuracy.
- Deeper or wider networks led to higher overfitting without accuracy gains.
- Binary Cross Entropy loss remained the most effective for binary sentiment classification.
- ReLU activation consistently outperformed tanh.
- Dropout regularization significantly reduced overfitting and improved stability.

Final Recommendation:

The 2-layer neural network with 32 units per layer, ReLU activation, Binary Cross-Entropy loss, and Dropout (0.5) provided the best balance between accuracy and generalization.
It is the most reliable model for IMDB sentiment classification among all tested configurations.