**Summary Report: IMDB Sentiment Analysis Using Neural Network Architectures**

**Shreya Thodupunuri**

**811301506**

**Introduction:**

The aim of this assignment was to enhance the performance of an existing neural network model on the IMDB sentiment analysis dataset. We tried out different model configurations, such as changing the number of hidden layers, units per layer, activation functions, loss functions, and regularization techniques. We wanted to see how each change impacted validation and test accuracy in order to find the best configuration.

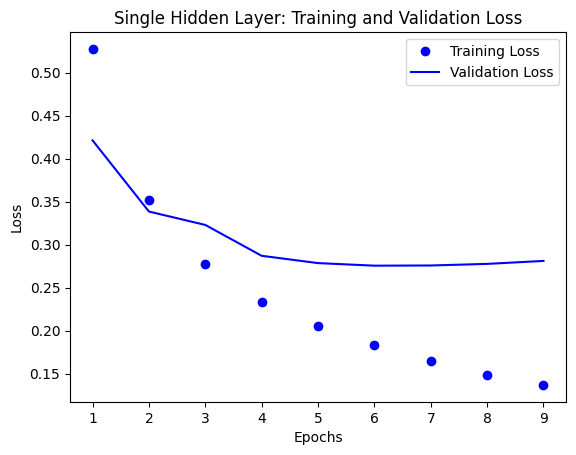
**Dataset:**

The dataset used for this assignment was the IMDB dataset, which consisted of 25,000 movie reviews for training and 25,000 reviews for testing. Each review was labeled as either positive or negative. The dataset was limited to the top 10,000 most frequently occurring words.

**Experimentation with Different Models**

**1. Single Layer vs. Multiple Layers**

**Single Hidden Layer Model**

A graph with red dots

Description automatically generated

* **Architecture:** 1 hidden layer with 16 units and ReLU activation.
* **Test Accuracy:** 88.09%
* **Loss:** 0.2976
* **Observation:** The model performed well with a simple architecture, effectively balancing training and generalization.

**Three Hidden Layers Model**

A graph with blue dots

Description automatically generatedA graph with red dots

Description automatically generated

* Architecture: 3 hidden layers with **16 units** each and **ReLU** activation.
* **Test Accuracy**: **87.22%**
* **Loss**: **0.3262**
* **Observation**: Increasing the number of layers did not improve the model's performance. The overfitting was caused by excessive capacity, leading to lower validation and test accuracy.

**2. Varying the Number of Units Per Layer**

**32 Units Per Layer**

* Architecture: 2 hidden layers with **32 units** each and **ReLU** activation.
* **Test Accuracy**: **88.32%**
* **Loss**: **0.2908**
* **Observation**: This model achieved the highest test accuracy, indicating that **32 units per layer** offered the optimal level of complexity for this dataset.

A graph with a line

Description automatically generated A graph with red dots

Description automatically generated

**64 Units Per Layer**

* Architecture: 2 hidden layers with **64 units** each and **ReLU** activation.
* **Test Accuracy**: **87.74%**
* **Loss**: **0.3023**
* **Observation**: Increasing the number of units to 64 did not improve the accuracy, indicating that additional complexity led to marginal overfitting.

A graph with a line

Description automatically generated A graph with red dots

Description automatically generated

**128 Units Per Layer**

* Architecture: 2 hidden layers with **128 units** each and **ReLU** activation.
* **Test Accuracy**: **83.43%**
* **Loss**: **0.4139**
* **Observation**: The **128 units per layer** model significantly overfitted the training data, leading to poor generalization and lower test accuracy.

A graph with blue lines

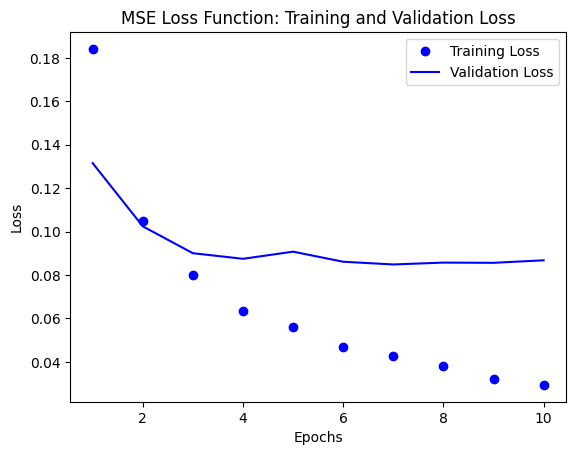
Description automatically generatedA graph with red lines

Description automatically generated

**3. Experimenting with Different Loss Functions**

**Mean Squared Error (MSE) Loss Function**

* Architecture: 2 hidden layers with **16 units** each and **ReLU** activation.
* **Test Accuracy**: **87.50%**
* **Loss**: **0.0941** (Note: MSE is not directly comparable with binary cross-entropy in magnitude)
* **Observation**: Using **MSE** instead of **binary cross-entropy** resulted in lower test accuracy. MSE is less suitable for classification problems, as it does not optimize probabilities effectively.

 A graph with red dots

Description automatically generated

**4. Activation Functions**

**Tanh Activation Function**

* Architecture: 2 hidden layers with **16 units** each and **tanh** activation.
* **Test Accuracy**: **88.35%**
* **Loss**: **0.2899**
* **Observation**: **Tanh** worked well and achieved a high test accuracy, comparable to ReLU. However, it may lead to slower convergence in larger models due to the vanishing gradient issue.

A graph with blue lines

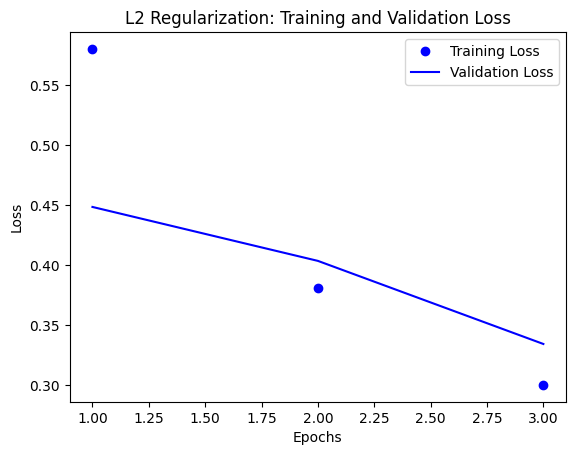
Description automatically generatedA graph with red dots

Description automatically generated

**5. Regularization Techniques**

**L2 Regularization**

* Architecture: 2 hidden layers with **16 units** each, **ReLU** activation, and **L2 regularization**.
* **Test Accuracy**: **88.30%**
* **Loss**: **0.3464**
* **Observation**: **L2 regularization** effectively controlled overfitting and resulted in a high test accuracy, closely aligned with the best-performing models.

A graph with red dots

Description automatically generated

**Dropout Regularization**:

* Architecture: 2 hidden layers with **16 units** each, **ReLU** activation, and **Dropout** with rate **0.5**.
* **Test Accuracy**: **86.98%**
* **Loss**: **0.3896**
* **Observation**: **Dropout** helped prevent overfitting but led to a slightly lower accuracy compared to L2 regularization. It likely reduced the model's effective capacity during training, which impacted its learning efficiency.

A graph with blue dots

Description automatically generated A graph with red dots

Description automatically generated

**Final Comparison of Model Performance**

A screenshot of a graph

Description automatically generated

**Key Insights:**

1. **Optimal Complexity:** The model achieved the best balance with 32 units per layer, reaching 88.32% accuracy, demonstrating optimal learning capacity and generalization.
2. **Overfitting in Complex Models:** Increasing complexity with three layers or 128 units per layer led to overfitting, resulting in lower test accuracy.
3. **Tanh vs. ReLU:** Tanh provides competitive accuracy compared to ReLU and demonstrates stable training behavior, making it a potential alternative for smaller models.
4. **Regularization Effectiveness:** Both L2 regularization and dropout helped prevent overfitting, but L2 regularization resulted in better generalization compared to dropout in this experiment
5. **Loss Function Choice:** Binary cross-entropy is more suitable than MSE for binary classification, leading to improved model convergence and accuracy.

**Conclusion**

The model with 32 units per layer, using binary cross-entropy loss and either ReLU or tanh activation, performed the best for the IMDB sentiment analysis task.

L2 regularization was effective in preventing overfitting and maintaining good generalization. Increasing model complexity beyond a certain level resulted in overfitting, highlighting the importance of selecting an optimal model architecture based on dataset size and complexity.