**Neural Network Model Performance Enhancement for IMDB Dataset**

**Assignment 2**

**Sai Sarath Sarma Peri**

**811345337**

[**speri2@kent.edu**](mailto:speri2@kent.edu)

1. **Goal of the assignment**

The goal here is to maximize the performance of a neural network model on the task of IMDB sentiment analysis. This is achieved by experimenting with different configurations by tuning parameters such as number of hidden layers, hidden units, activation functions, and loss functions. Regularization methods such as dropout are also employed to prevent overfitting.

1. **IMDB Dataset**

IMDB dataset, with positive or negative tags in the movie reviews, is used in this task. The dataset has 50,000 reviews, out of which 25,000 is for training and 25,000 for testing. The model is required to assign sentiment label to the reviews.

1. **Test Setup**

The basic feed-forward architecture upon which the neural network model is built has the following features:

* Input layer: The binary multi-hot encoding of the 10,000 most frequently occurring terms in the reviews, represented as a 10,000-dimensional vector.
* Hidden layers: The experiment determines how many units and layers are used.
* Output layer: One unit that predicts whether reviews will be good or negative using a sigmoid activation function.

**Experiment Results:**

**Effectiveness of number of hidden layers:**

The goal is to assess how well models with 1, 2, 3, 4, and 5 hidden layers perform.

A graph of different colored lines

AI-generated content may be incorrect.

A graph of different colored lines

AI-generated content may be incorrect.

•The single hidden layer causes the model to converge faster, and the validation accuracy remains close to the training accuracy, indicating good generalization.

• Adding more hidden layers made the model more complicated but only brought about incremental gains (or even losses) in accuracy, possibly due to overfitting.

**Changing the number of units**

To examine the effects on model performance of varying the amount of hidden units per layer (32, 64, 128, 264).

A graph of different colored lines

AI-generated content may be incorrect.

A graph of different colored lines

AI-generated content may be incorrect.

• As there are more hidden units, the accuracy improves consistently.

• The 256 hidden unit model produced smoother learning curves, indicating stronger and more stable training compared to the smaller unit models.

**Changing the number of hidden layers**

To evaluate the models' performance against the baseline binary\_crossentropy using the MSE loss function.

A graph of a graph

AI-generated content may be incorrect.A graph of loss functions

AI-generated content may be incorrect.

The plot of comparison of MSE and binary crossentropy shows that models trained with binary crossentropy converge faster and are more accurate, as would be expected because binary crossentropy is more appropriate for classification problems.

**Using tanh function instead of ReLU**

A graph of different colored lines

AI-generated content may be incorrect.

A graph of a graph showing different colored lines

AI-generated content may be incorrect.

* ReLU activation function is always superior with a test accuracy of 0.87148 compared to Tanh's 0.85176. Tanh, although applied in earlier neural networks, is prone to the vanishing gradient problem, particularly in deeper networks, causing slower and less efficient training. ReLU does not suffer from this issue, and it enables faster convergence and improved accuracy.

**Adding the dropout layer**

To assess how using dropout regularization can lessen overfitting.

A graph of different colored lines

Description automatically generated

A graph of different colored lines

Description automatically generated

A dropout rate of 0.7 gives the highest validation accuracy, which means that it is a good compromise between model complexity and generalization. While dropout helps avoid overfitting, very high rates may hinder learning; hence, 0.7 seems to be the optimal here. Overall, all the dropout rates perform similarly, which shows the stability of the model to regularization.

**Evaluation of final test results**

* Final Model Test Accuracy: 0.8619
* Final Model Test Loss 0.4936

**Best parameters for the model**

* Hidden layers: 1 Layer
* Hidden units: 256 Units
* Loss function: Binary Crossentropy
* Activation function: ReLU
* Regularization: Dropout Rate of 0.7

**Conclusion**

The best fit model of the neural network is a single hidden layer of 256 units, which has the ideal balance between complexity and performance. It utilizes the Binary Crossentropy loss function, which produced the highest validation and test accuracies. The ReLU activation function also enhanced the model's performance, contributing to a high validation accuracy. Additionally, the dropout rate of 0.7 was applied for regularization, which effectively mitigated overfitting without compromising the accuracy metrics. Overall, this configuration resulted in a test loss of 0.4936 and a test accuracy of 0.8619, which was the highest-performing model for the dataset.