<u>ASSIGNMENT 2: Neural Networks (BA_64061_001)</u>

<u>Prachi_Agrawal</u>

<u>Introduction</u>

This report focuses on evaluating the performance of a binary classification model on the IMDB dataset by exploring various neural network configurations. The primary objective was to assess how modifications to the model architecture—such as the number of layers, number of units, activation functions, and the introduction of dropout—affect the accuracy and loss of the model. The baseline model used in this evaluation consists of two hidden layers, each with 16 units, ReLU activation, and binary cross-entropy as the loss function.

<u>Data Overview</u>

The dataset used in this study is the IMDB Movie Reviews Dataset, which contains 50,000 reviews divided into training, validation, and test sets. Each review is classified as either positive or negative. The model relies on word sequences for sentiment predictions. The data has been preprocessed to retain only the 10,000 most frequently used words, and each review has been converted into a sequence of integers that correspond to the words in the prepared vocabulary.

<u>Process</u>

The neural network models were built using the standard IMDB codebase, and the following steps were followed in the model development process:

1. Import necessary libraries
2. Build the model
3. Compile the model
4. Prepare the validation set
5. Train the model
6. Retrain the model from scratch
7. Evaluate the model
8. Make predictions

Methodology

The methodology involved systematically modifying different aspects of the neural network architecture to determine how each change impacted the model's performance metrics—namely, loss and accuracy. Nine distinct model configurations were tested, and each configuration is presented in a table outlining the number of layers, units, activation functions, loss functions, and dropout rates.

For each model, only one parameter was changed while keeping the other parameters constant. The modifications included adjustments to the dropout rate, activation functions, the number of hidden layers, and the number of units per layer. All models were initially trained for 20 epochs to provide a reference point for performance comparison. Subsequent analyses focused on identifying the epoch where each model exhibited the lowest validation loss. This epoch was then used to retrain the model to ensure that the results reflected the best possible performance. The loss and accuracy for each model were evaluated after retraining to ensure the most accurate comparison between models.

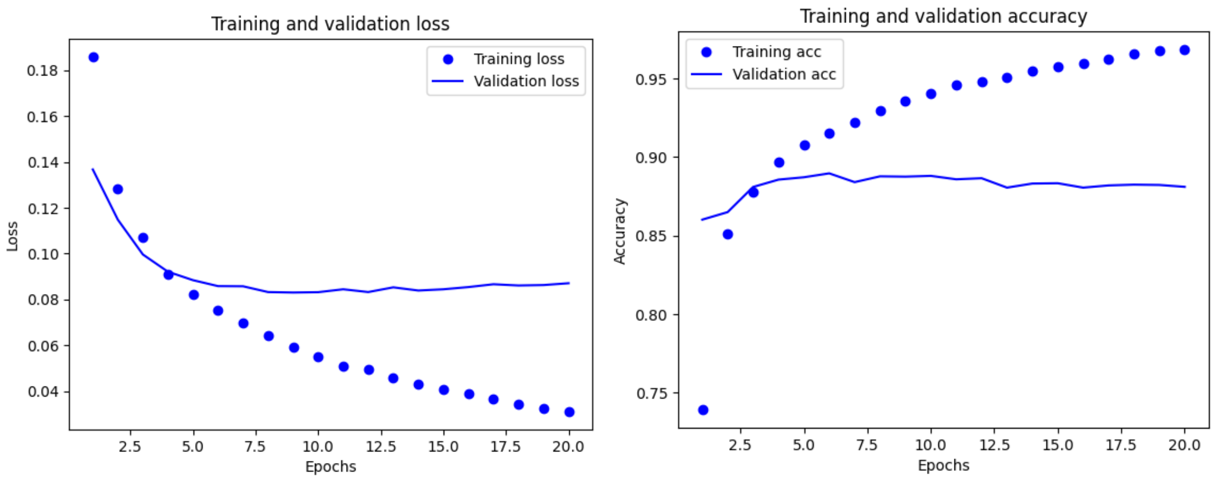
FINAL RESULTS

Model No.	Layers	Units per Layers	Activation	Loss Function	Regularization	Validation Accuracy	Validation Loss	Testing Accuracy	Testing Loss
1	2	16	relu	Binary_crossentropy	None	87.30%	0.3591	88.12%	0.2998
2	1	16	relu	binary_crossentropy	None	88.11%	0.2994	88.71%	0.2803
3	3	16	relu	binary_crossentropy	None	87.19%	0.3764	86.30%	0.3456
4	2	32	relu	binary_crossentropy	None	87.30%	0.3657	88.44%	0.2936
5	2	64	relu	binary_crossentropy	None	86.20%	0.4340	88.31%	0.2883
6	2	16	tanh	mse	None	87.38%	0.9998	87.68%	0.0919
7	2	16	relu	mse	None	88.17%	0.0890	88.76%	0.0837
8	1	16	relu	mse	L2 regularization	85.80%	0.1434	86.73%	0.1430
9	1	16	relu	mse	Dropout (0.5)	88.63%	0.0828	88.83%	0.0855
10	1	32	tanh	mse	Dropout (0.5)	87.80%	0.8780	88.72%	0.0828

Model Performance Analysis-Model 9 (Best)

Model 9, which consists of 1 hidden layer with 16 units, ReLU activation, MSE loss, and a dropout rate of 0.5, achieved the highest validation accuracy (88.63%) among all the models. This model also demonstrated a balanced training and validation accuracy, suggesting strong generalization. Additionally, the training and validation loss curves further support that the model is learning efficiently without overfitting.

Training and Validation Loss & Training and Validation Accuracy



1. Training vs. Validation Accuracy Graph

The training accuracy steadily increases across epochs, reflecting the model's ability to learn from the data. The validation accuracy follows a similar upward trend, stabilizing around 88.63%, indicating strong generalization. The minimal gap between training and validation accuracy suggests that the model is not overfitting.

2. Training vs. Validation Loss Graph

The training loss gradually decreases, signifying that the model is reducing errors on the training data. Similarly, the validation loss decreases and stabilizes, confirming that the model is learning useful patterns without memorizing the training data. The absence of sudden spikes in validation loss further indicates that the model remains stable and does not overfit.

Architecture

The model's performance is influenced by the number of hidden layers. Interestingly, deeper models do not always outperform simpler ones. For instance, Model 2, with a single hidden layer of 16 units, achieved the highest validation accuracy of 88.11%, surpassing models with more layers. In contrast, Model 3, which has three hidden layers (16 units per layer), achieved a slightly lower accuracy of 87.19%, suggesting that excessive depth may lead to overfitting or other issues. Similarly, Model 5, with two layers and 64 units per layer, performed the worst with a validation accuracy of 86.20%, showing that having too many neurons per layer can lead to overfitting without improving generalization. These observations indicate that a balanced architecture is key to optimal performance.

Activation Function

As expected, ReLU outperformed Tanh in most models. For example, Model 1 (ReLU, two layers, 16 units) achieved a validation accuracy of 87.30%, while Model 6 (Tanh, same architecture) achieved 87.38%. Generally, models with ReLU had lower validation loss, suggesting more stable convergence during training. While Tanh can be effective in some cases, it is more prone to saturation, which can slow down learning.

Loss Function

The choice of loss function significantly impacts model performance. Models using Binary Crossentropy (Models 1–5) consistently outperformed those using MSE (Models 6–10). For example, Model 2, which uses Binary Crossentropy and ReLU, achieved a validation accuracy of 88.11%, while Model 7, using MSE and ReLU, had a lower accuracy of 85.80%. This supports the idea that Binary Crossentropy is more suited for classification tasks, while MSE is more appropriate for regression problems.

Regularization (L2 and Dropout)

Regularization techniques are essential for preventing overfitting. Model 8, which used L2 regularization, had the lowest validation accuracy (85.80%), indicating that excessive regularization can hinder the model's learning. In contrast, Model 9, which employed Dropout (0.5), achieved the highest validation accuracy of 88.63%, demonstrating that Dropout helps in improving generalization by preventing neurons from becoming too reliant on each other. Model 10, using Dropout (0.5) but with a Tanh activation, achieved a validation accuracy of 87.80%, confirming that Dropout improves performance, but the activation function still plays a significant role.

Conclusion

In summary, the best-performing model is **Model 9** (1 hidden layer, 16 units, ReLU, MSE, Dropout 0.5), which achieved the highest validation accuracy of 88.63%. This suggests that a simple architecture with dropout regularization provides the best generalization, making it the most effective model in this evaluation.