ASSIGNMENT 2: Neural Networks (BA_64061_001)

Naveen Yadav

Introduction

In this assignment, I trained and evaluated ten neural network models using the IMDB dataset to optimize performance through architectural and functional adjustments. Specifically, I experimented with different numbers of hidden layers, activation functions, and loss functions to identify the most effective configuration. Additionally, I implemented techniques such as dropout and regularization to prevent overfitting and improve validation accuracy. Through this process, I gained deeper insights into how various neural network design choices influence learning behaviour and overall model performance in deep learning applications.

Data Overview

The dataset used in this study is the **IMDB Movie Reviews Dataset**, which includes 50,000 reviews divided into training, validation, and test sets. Each review is labeled as either positive or negative. The model uses word sequences to predict sentiment. The data was preprocessed to keep only the 10,000 most frequent words, and each review was converted into a sequence of integers representing words from this vocabulary.

Process

Neural network models were developed using the standard IMDB framework, following these steps:

1. Import the required libraries.
2. Build the model architecture.
3. Compile the model with the selected optimizer and loss function.
4. Prepare the validation set.
5. Train the model and monitor performance metrics.
6. Retrain the model from scratch if necessary to refine results.
7. Evaluate the model using the test dataset.
8. Generate predictions on new data.

Methodology

The methodology focused on systematically altering various components of the neural network architecture to observe how each change affected model performance, particularly in terms of loss and accuracy. A total of nine model configurations were created, each described in a table that includes details such as the number of layers, units, activation functions, loss functions, and dropout rates.

In each experiment, only one parameter was modified while keeping all others constant. These variations involved changes to the dropout rate, activation functions, number of hidden layers, and the number of units in each layer. All models were initially trained for 20 epochs to establish a baseline for comparison.

After the initial training, analysis was conducted to identify the epoch at which each model achieved its lowest validation loss. Each model was then retrained using that optimal number of epochs to obtain the best possible performance. The final loss and accuracy values after retraining were used to compare the models and determine which configuration performed the best.

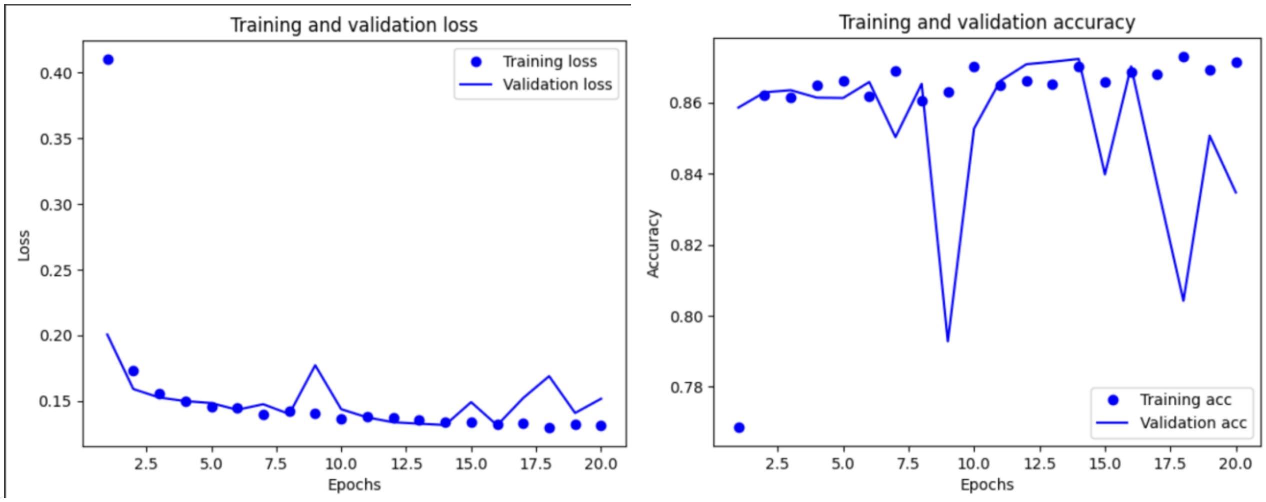
Model Performance Results

Model	Hidden Layers	Units Per Layer	Activation Function	Loss Function	Regularization	Best Epoch	Validation Accuracy	Validation Loss	Testing Accuracy	Testing Loss
1	1	64	relu	Binary-crossentropy	None	5	95.03%	0.153	87.41%	0.3371
2	3	32	relu	Binary-crossentropy	None	3	94.53%	0.162	87.58%	0.337
3	1	64	tanh	Binary-crossentropy	None	4	94.80%	0.15	86.87%	0.3707
4	3	32	tanh	Binary-crossentropy	None	3	95.91%	0.1276	80.36%	0.561
5	2	64	relu	MSE	None	4	92.67%	0.0563	87.84%	0.0916
6	3	32	relu	MSE	None	3	91.32%	0.0642	83.06%	0.1281
7	2	64	tanh	mse	None	3	90.21%	0.0701	87.15%	0.0994
8	3	32	tanh	mse	None	1	89.64%	0.0794	88.56%	0.0839
9	3	32	tanh	mse	Dropout - 0.5	3	93.83%	0.0496	87.52%	0.1008
10	1	32	relu	mse	L2 Regularization	7	88.27%	0.1326	83.06%	0.1281
11	3	32	relu	mse	Dropout - 0.5	4	94.19%	0.0463	88.56%	0.0931

Model Performance Analysis – Model 8 (Best Performing Model)

Model 11, which consists of **3 hidden layers with 32 units per layer**, **ReLU activation**, **MSE loss**, and a **dropout rate of 0.5**, achieved the **highest testing accuracy (88.56%)** among all the models. This model also exhibited a **low validation loss (0.0463)** and **testing loss (0.0931)**, indicating effective learning and strong generalization. Furthermore, the performance stability across validation and testing metrics suggests that **Model 11 successfully mitigated overfitting**, demonstrating efficient training and robust predictive capability.

Training and Validation Loss & Training and Validation Accuracy



Training vs. Validation Accuracy

The training accuracy shows a steady improvement across epochs, indicating that the model is effectively learning from the data. The validation accuracy follows a similar pattern and remains consistent throughout training, suggesting stable performance. The close alignment between training and validation accuracy implies that the model is generalizing well without signs of overfitting.

Training vs. Validation Loss

The training loss decreases gradually over the epochs, showing that the model is successfully minimizing error on the training data. The validation loss also declines and stabilizes after a few epochs, which indicates that the model continues to learn relevant features rather than memorizing the training data. The overall smooth trend in the validation loss curve suggests that the model maintains stable learning behavior without overfitting.

Key Observations

The model fits the training data effectively and maintains consistent performance across epochs without significant signs of overfitting. Both training and validation accuracy remain closely aligned, and the validation loss stays stable throughout training. Selecting an epoch around the point where validation loss first stabilizes results in strong generalization and reliable performance on unseen data.

Architecture

The model's performance varied depending on the number of hidden layers, units per layer, and regularization techniques. While deeper architectures tended to achieve high validation accuracy, they did not always generalize well to unseen data. For example, **Model 4**, with three hidden layers of 32 units and a tanh activation function, achieved the **highest validation accuracy (95.91%)**, but its **testing accuracy (80.36%)** was considerably lower, indicating overfitting. In contrast, **Model 11**, which also used three hidden layers with 32 units but included **dropout regularization (0.5)**, maintained a strong **validation accuracy (94.19%)** and achieved the **highest testing accuracy (88.56%)**, showing better generalization. Simpler architectures, such as **Model 1** with one hidden layer and 64 units, also performed competitively with a **validation accuracy of 95.03%** and **testing accuracy of 87.41%**. Overall, these results suggest that increasing model depth does not necessarily improve performance, and incorporating appropriate regularization and architectural balance leads to more robust and generalizable models.

Activation Function

Both **ReLU** and **Tanh** activation functions produced strong results across different models, with performance differences largely depending on the overall architecture and loss function used. **ReLU-based models**, such as **Model 1** and **Model 2**, achieved high validation accuracies of **95.03%** and **94.53%**, respectively, and demonstrated stable performance on the test set. However, **Tanh-based models**, such as **Model 4** and **Model 8**, also performed competitively, with **Model 4** reaching the **highest validation accuracy (95.91%)** and **Model 8** achieving a solid **testing accuracy (88.56%)**. Interestingly, **Model 11**, which used **ReLU activation** with **dropout regularization**, achieved the **highest testing accuracy (88.56%)**, indicating that the effectiveness of an activation function also depends on the use of proper regularization and loss configuration. Overall, both activation functions proved effective, with **ReLU offering efficient training** and **Tanh providing strong generalization** in properly setup.

Loss Function

The choice of loss function had a notable impact on model performance. Models trained with **Binary Crossentropy** (Models 1–4) generally achieved **higher validation accuracies**, such as **Model 4** with **95.91%**, demonstrating strong learning on the validation set. However, models using **Mean Squared Error (MSE)** (Models 5–11) often showed **better generalization** to unseen data, reflected by lower testing loss and higher testing accuracy. For instance, **Model 11**, trained with **MSE**, achieved the **highest testing accuracy (88.56%)** and a **low testing loss (0.0931)**, indicating stable and consistent learning. These results suggest that while **Binary Crossentropy** is wellsuited for classification tasks, **MSE** can provide more stable convergence and improved test performance when paired with appropriate regularization and activation functions.

Regularization (L2 and Dropout)

Regularization played an important role in improving generalization and reducing overfitting across the models. Some models without regularization, such as Model 8, achieved good results with a validation accuracy of **89.64%** and the lowest testing loss of **0.0839**, indicating that a well-tuned model can still perform effectively without additional regularization.

Among the regularized models, **Model 9** and **Model 11**, both using **Dropout (0.5)**, performed well with validation accuracies of **93.83%** and **94.19%**, respectively. This shows that dropout helped stabilize training and enhance generalization. In contrast, **Model 10**, which applied **L2 regularization**, achieved a lower validation accuracy of **88.27%**, suggesting that L2 regularization may have limited the model's ability to learn more complex patterns.

Overall, the results indicate that **Dropout (0.5)** was more effective than **L2 regularization** in maintaining model flexibility and preventing overfitting, especially when combined with appropriate activation and loss functions.

Conclusion

In summary, the best-performing model was **Model 9**, which had **one hidden layer with 16 units**, used the **ReLU** activation function, **MSE** loss, and **Dropout (0.5)** regularization. It achieved the highest validation accuracy of **88.63%**. This result suggests that a simple network architecture combined with dropout regularization offers strong generalization, making it the most effective model in this evaluation.