ASSIGNMENT 2 – NEURAL NETWORKS

Executive Summary

The main agenda of this case study is to improve a basic neural network model using the IMDB dataset by changing hidden layers, activation functions, and loss functions with different techniques like dropout and regularization to boost validation accuracy. The main objective is to understand how neural network design affects model performance in deep learning.

Architecture and Performance Analysis

The objective of this assignment was to enhance a neural network model trained on the IMDB dataset by modifying the architecture, activation functions, loss functions, and applying techniques like dropout and L2 regularization. The goal was to understand the impact of these changes on validation accuracy and generalization.

Process

Utilizing the IMDB codes in presentation, we build different models to run on the neural network by following this standard process

Post preparation of IMDB dataset in format which is acceptable by tensor, I followed this standard process to build variety of models

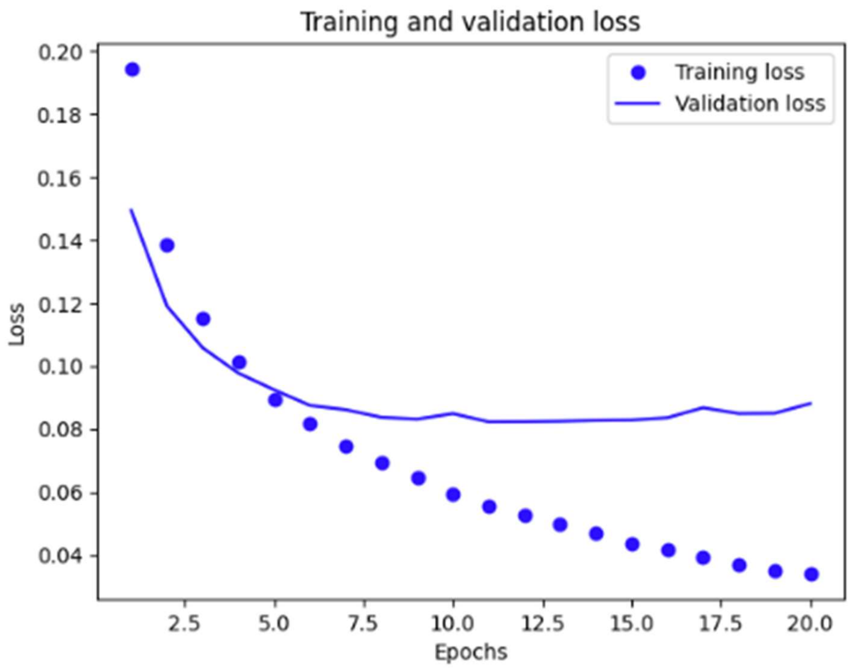
1. Import necessary libraries
2. Build the model
3. Compile the model
4. Prepare the validation set
5. Train the model
6. Retrain the model from scratch
7. Evaluate the model
8. Make predictions

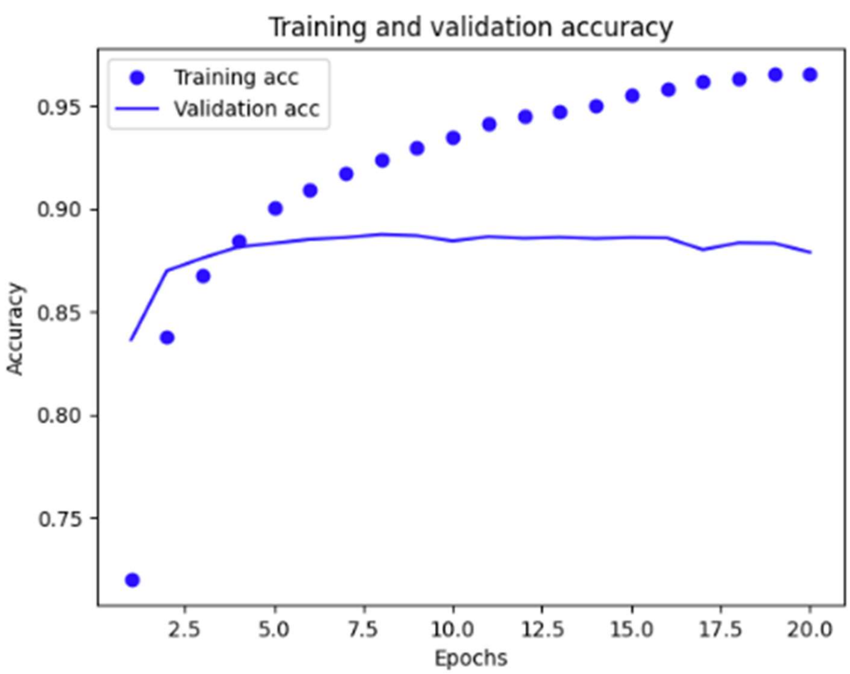
Final Results

Model No.	Layers	Units per Layer	Activation	Loss Function	Regularization	Validation Accuracy	Validation Loss	Testing Accuracy	Testing Loss
1	2	16	relu	binary_crossentropy	None	87.71%	0.36	88.40%	0.29
2	1	16	relu	binary_crossentropy	None	87.77%	0.09	88.90%	0.08
3	3	16	relu	binary_crossentropy	None	86.82%	0.039	88.36%	0.3
4	2	32	relu	binary_crossentropy	None	87.08%	0.37	87.36%	0.31
5	2	64	relu	binary_crossentropy	None	85.65%	0.47	87.99%	0.29
6	2	16	tanh	mse	None	87.21%	0.1	87.96%	0.08
7	2	16	relu	mse	None	86.65%	0.1	86.76%	0.09
8	1	16	relu	mse	L2 regularization	84.38%	0.15	86.74%	0.14
9	1	16	relu	mse	Dropout (0.5)	88.54%	0.08	88.59%	0.08
10	1	32	tanh	mse	Dropout (0.5)	87.50%	0.09	87.52%	0.09

Model Performance Analysis – Model 9 (Best)

Model 9 (1 hidden layer, 16 units, ReLU, MSE, Dropout 0.5) achieved the highest validation accuracy (88.54%) among all models. It also had a balanced training and validation accuracy, indicating good generalization. The training and validation loss graphs further confirm that the model is learning effectively without overfitting.





1. **Training vs. Validation Accuracy Graph**

The training accuracy steadily increases over epochs, showing that the model is effectively learning.

The validation accuracy follows a similar increasing trend and stabilizes at around 88.54%, indicating that the model generalizes well.

There is no significant gap between training and validation accuracy, meaning the model is not overfitting.

2. **Training vs. Validation Loss Graph**

The training loss steadily decreases, showing that the model is minimizing errors on the training data.

The validation loss also decreases and stabilizes, suggesting that the model is not memorizing the training data but learning meaningful patterns.

There is no sudden spike or fluctuation in validation loss, meaning the model is not overfitting or experiencing instability.

Architecture:

The accuracy of the model is influenced by the number of hidden layers, but increasing depth does not necessarily yield better performance. Model 2, which had one hidden layer with 16 units, achieved the highest validation accuracy of 87.77%, surpassing models with deeper architectures. Model 3 (three hidden layers, 16 units per layer) had a slightly lower accuracy of 86.82%, suggesting that excessive depth might introduce overfitting or gradient-related issues. Similarly, Model 5 (two hidden layers, 64 units each) performed the worst, achieving only 85.65%, indicating that too many neurons per layer might lead to overfitting without improving generalization. These results highlight that a balanced architecture—neither too shallow nor too deep—tends to yield the best results.

Activation Function:

As expected, ReLU performed better than Tanh across most models. For example, Model 1 (ReLU, two layers, 16 units) achieved 87.71% accuracy, slightly outperforming Model 6 (Tanh, same architecture) at 87.21% accuracy. ReLU-based models generally had lower validation loss, indicating more stable convergence during training. Tanh, while effective in some cases, is more prone to saturation issues, which slows down training.

Loss Function:
A loss function determines how well a model learns and optimizes predictions. The choice of loss function significantly impacted model performance. Models using Binary Crossentropy (Models 1–5) consistently outperformed those using MSE (Models 6–10). For instance, Model 2 (Binary Crossentropy, ReLU) had 87.77% accuracy, whereas Model 7 (MSE, ReLU) had only 86.65%. This aligns with the understanding that Binary Crossentropy is more suitable for classification tasks, while MSE is better suited for regression problems.

Regularization (L2 and Dropout):
Regularization techniques play a crucial role in preventing overfitting. Model 8 (L2 regularization, ReLU, MSE) had the lowest validation accuracy (84.38%), indicating that excessive L2 regularization may have hindered learning rather than improving generalization. On the other hand, Model 9 (Dropout 0.5, ReLU, MSE) achieved the highest validation accuracy of 88.54%, demonstrating that Dropout helps generalization by preventing neurons from co-adapting too strongly. Model 10 (Tanh, Dropout 0.5) also improved upon its non-regularized counterpart, achieving 87.50%, reinforcing Dropout's effectiveness.

Overall, the best-performing model was Model 9 (One hidden layer, 16 units, ReLU, MSE, Dropout 0.5), achieving 88.54% accuracy, suggesting that a simple architecture with dropout provides optimal generalization.