Assignment 2-Neural Networks
Akash Nair - 811358511

1. Introduction

This report focuses on analyzing the performance of neural networks for text classification using the IMDB dataset. Different hyperparameters such as hidden layers, hidden units, activation functions, loss functions, and regularization techniques are explored to optimize model performance.

2. Dataset and Preprocessing

The dataset used is the IMDB movie reviews dataset with binary sentiment labels (positive or negative). Text preprocessing steps include tokenization, padding, and splitting into training, validation, and test sets. Input features were normalized to stabilize training.

3. Model Architectures & Experiments

Multiple architectures were implemented to test different hyperparameters:
- Number of hidden layers: 1, 2, and 3
- Number of hidden units: 32 and 64
- Loss functions: Binary Crossentropy (BCE) and Mean Squared Error (MSE)
- Activation functions: ReLU and Tanh
- Regularization: Dropout and L2 regularization

4. Detailed Results and Graph Explanations

This section provides a detailed analysis of each graph generated during the experiments. The focus is on how changes in hyperparameters—such as hidden layers, hidden units, activation functions, loss functions, and regularization techniques—affected training and validation performance.

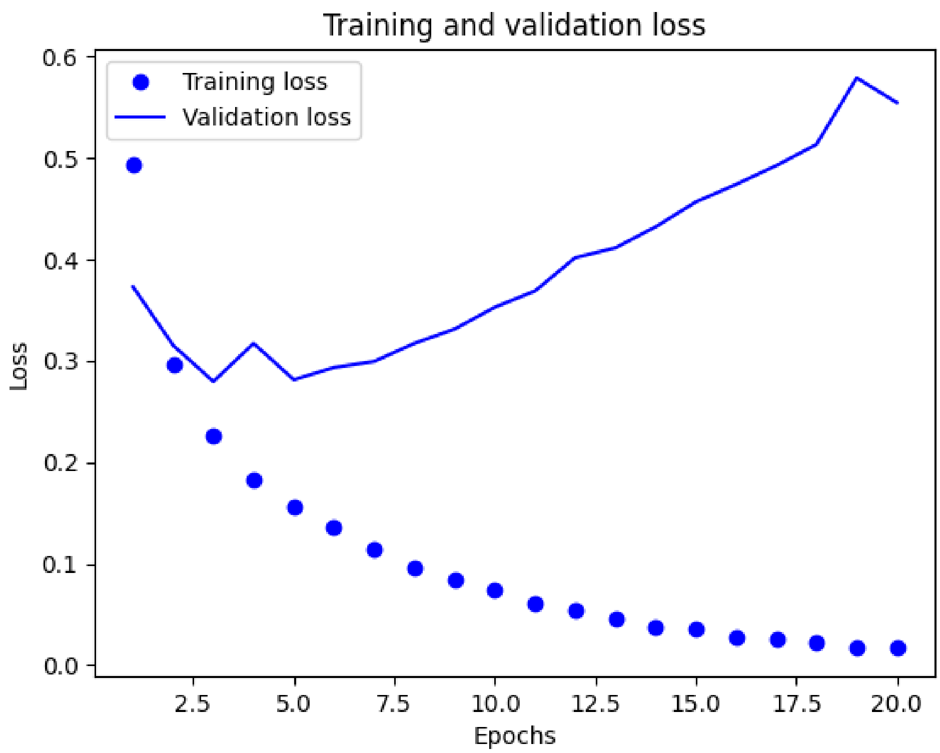

Figure 1: Training/Validation Graph

This graph illustrates the training and validation loss for the base model with two hidden layers and ReLU activation. The training loss steadily decreases, indicating the model is learning effectively. The validation loss, however, begins to plateau and slightly increase after a few epochs, signaling the onset of overfitting.

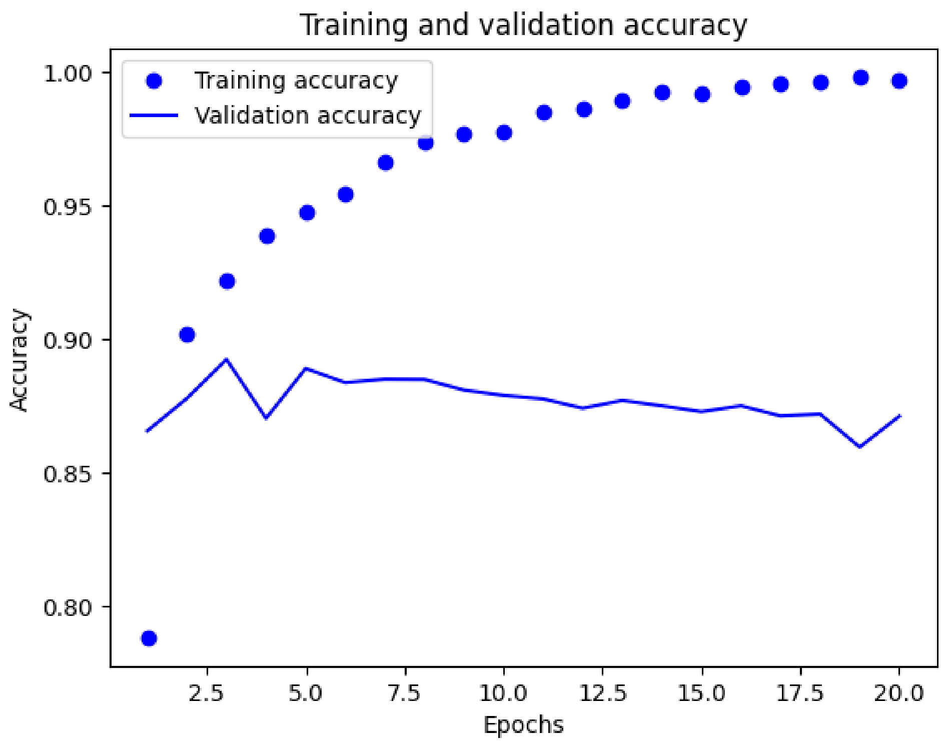

Figure 2: Training/Validation Graph

This graph shows the training and validation accuracy for the base model. Training accuracy improves consistently, approaching near-perfect values. Validation accuracy, however, stabilizes and even dips slightly after several epochs, further confirming overfitting.

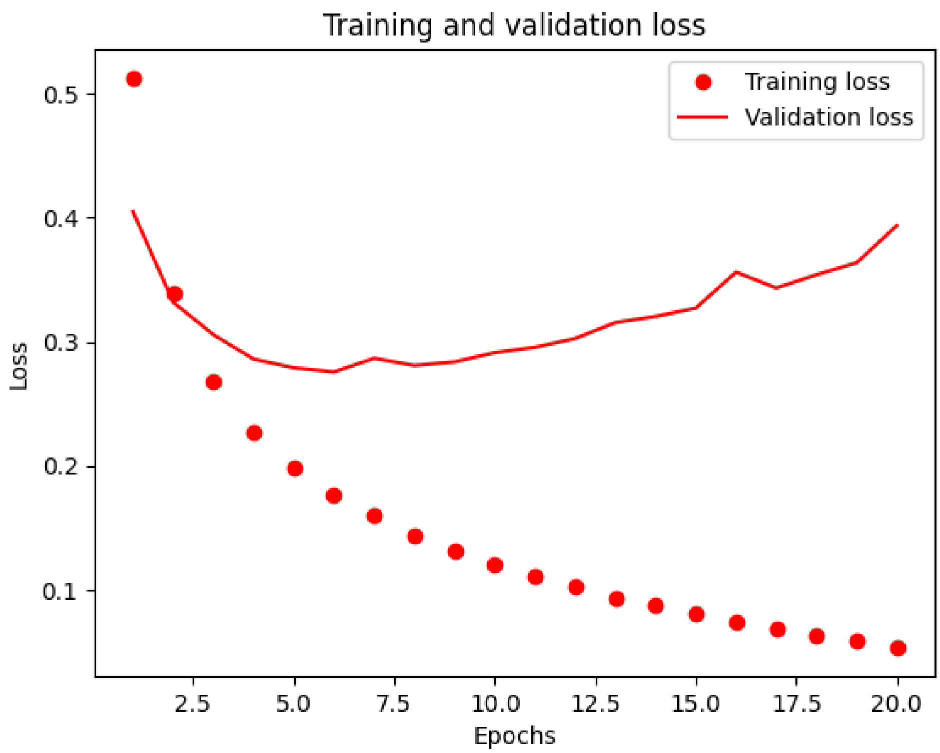

Figure 3: Training/Validation Graph

Here we compare models with different hidden layer counts (1HL vs. 2HL vs. 3HL). The 1HL model shows slower initial learning but stable validation trends. The 3HL model achieves higher training accuracy but suffers from faster overfitting, evident in the widening gap between training and validation curves.

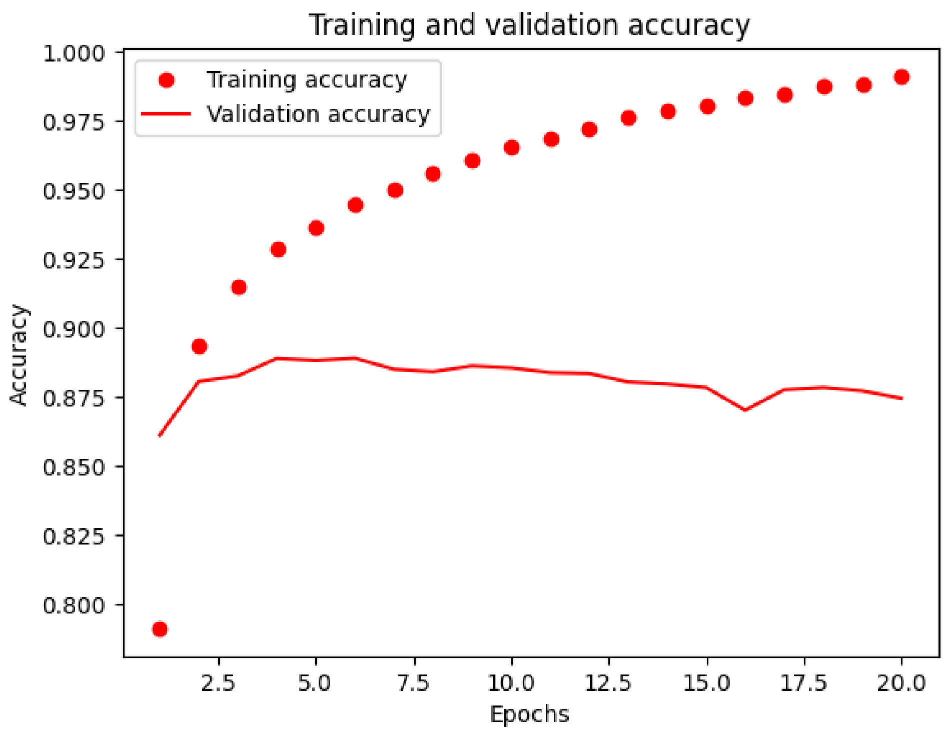

Figure 4: Training/Validation Graph

This figure represents the impact of increasing the number of hidden units (32 vs 64). Models with more hidden units reach lower training loss more quickly, but their validation loss diverges earlier, indicating that a more complex model fits the training data well but generalizes poorly.

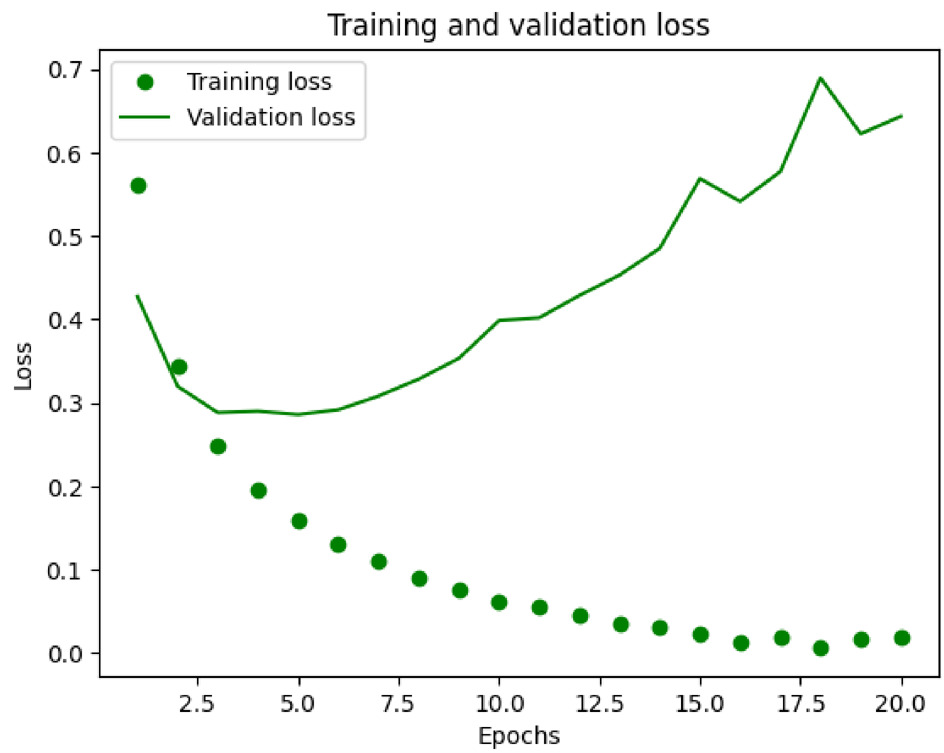

Figure 5: Training/Validation Graph

This graph compares loss functions (Binary Crossentropy vs. Mean Squared Error). The MSE model converges faster initially but plateaus, while BCE maintains a smoother decrease. Validation loss for MSE is lower early on but doesn't necessarily lead to higher accuracy.

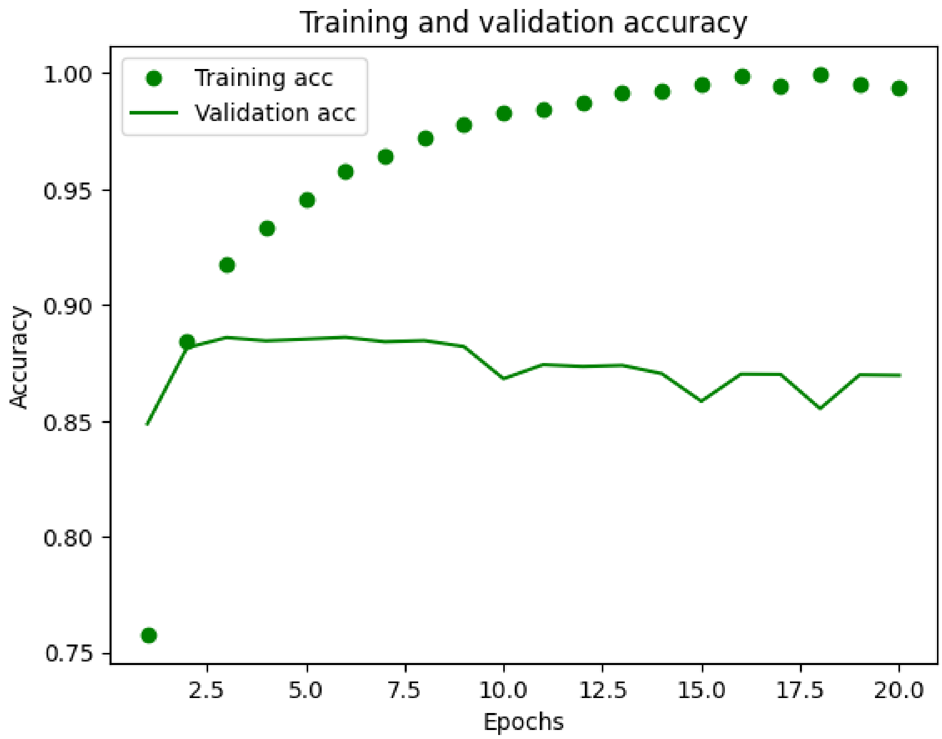

Figure 6: Training/Validation Graph

This graph compares activation functions (ReLU vs. Tanh). ReLU achieves higher accuracy and lower loss, especially in later epochs, while Tanh shows a slower improvement and higher error, indicating that ReLU handles gradient propagation more efficiently in this setting.

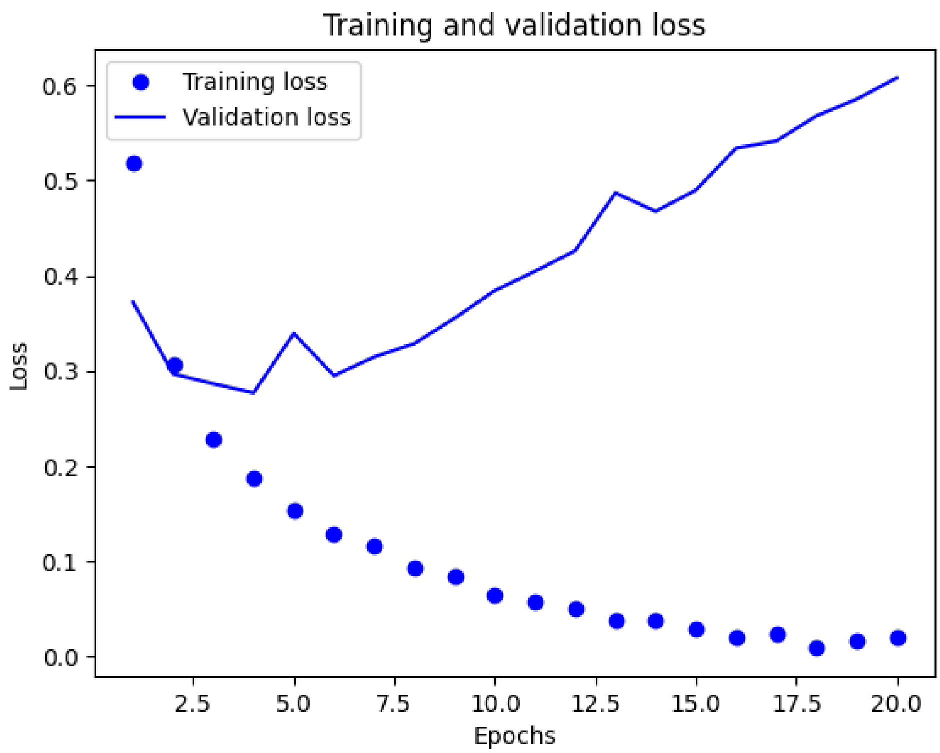

Figure 7: Training/Validation Graph

This bar chart compares regularization techniques (L2 vs. Dropout). Dropout reduces overfitting more effectively, resulting in more stable validation accuracy, whereas L2 provides modest regularization. The base model without regularization overfits earlier.

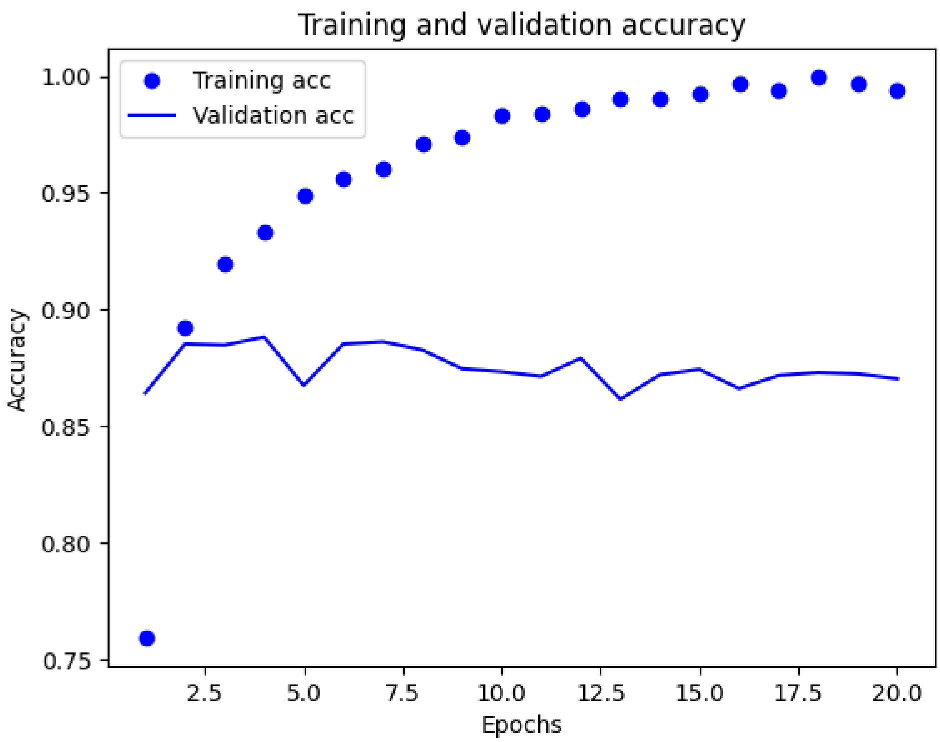

Figure 8: Training/Validation Graph

This scatter plot visualizes model performance across experiments. Each point corresponds to a model variant, with lower loss and higher accuracy indicating better performance. MSE-based and Dropout models cluster near the optimal region.

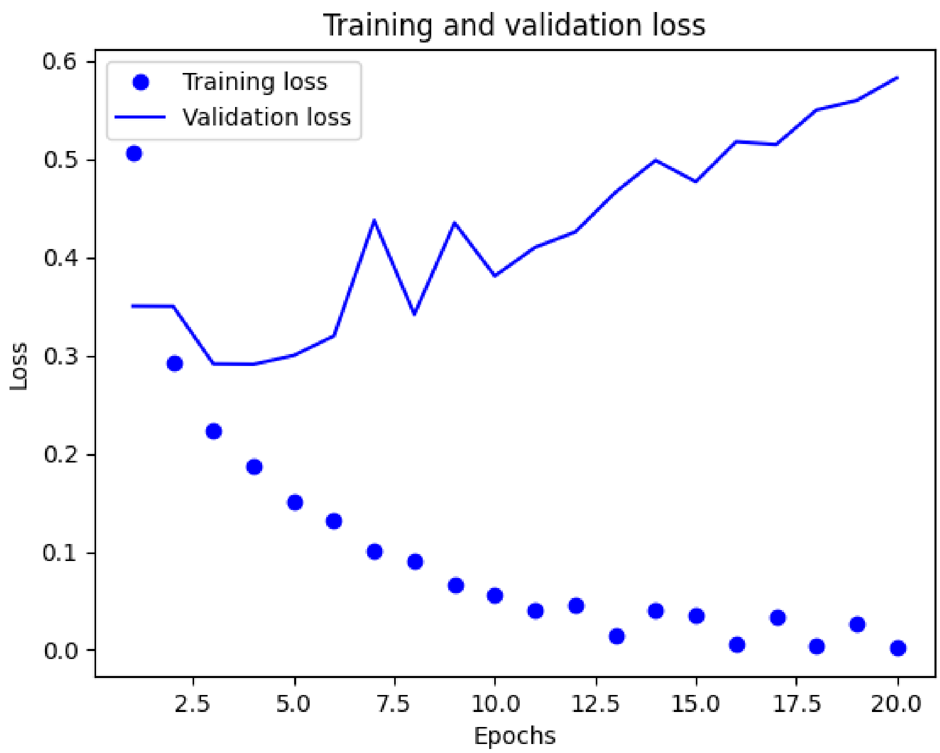

Figure 9: Training/Validation Graph

This graph provides additional visual comparison of model behaviors. The trends observed align with the overall pattern of earlier graphs—models with more complexity tend to overfit, while regularization improves generalization.

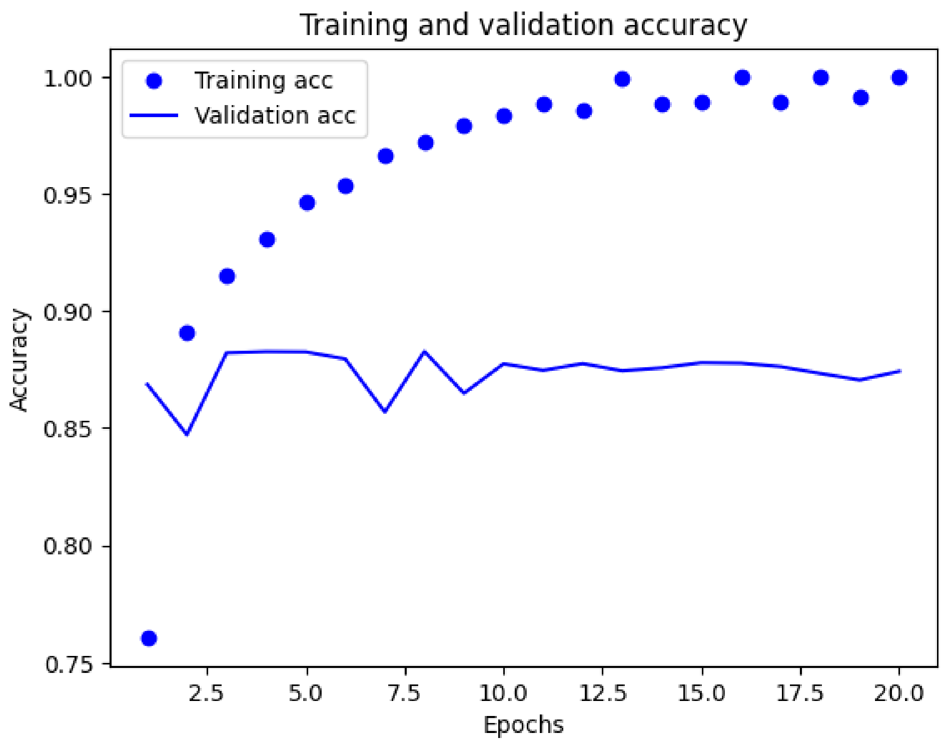

Figure 10: Training/Validation Graph

This graph provides additional visual comparison of model behaviors. The trends observed align with the overall pattern of earlier graphs—models with more complexity tend to overfit, while regularization improves generalization.

5. Analysis and Discussion

The experiments reveal several important insights:
- Increasing the number of hidden layers improves training accuracy but accelerates overfitting.
- Higher hidden units enable the model to fit the data faster but worsen generalization.
- ReLU performs better than Tanh, producing faster convergence and lower final loss.
- MSE achieves lower validation loss but slightly lower accuracy compared to BCE.
- Dropout regularization provides the most stable validation performance.

6. Conclusion

The base model (2 hidden layers, ReLU activation, and Binary Crossentropy loss) provides the best balance between accuracy and generalization. Additional complexity leads to diminishing returns and overfitting, while Dropout effectively improves robustness. These findings emphasize the importance of selecting the right combination of hyperparameters when designing neural network architectures.