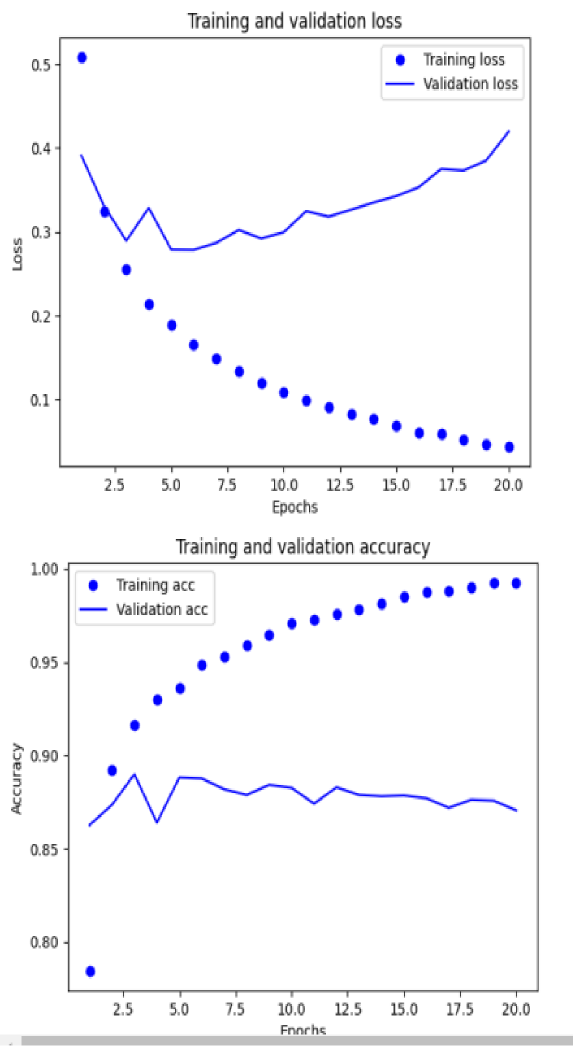
SUMMARY:

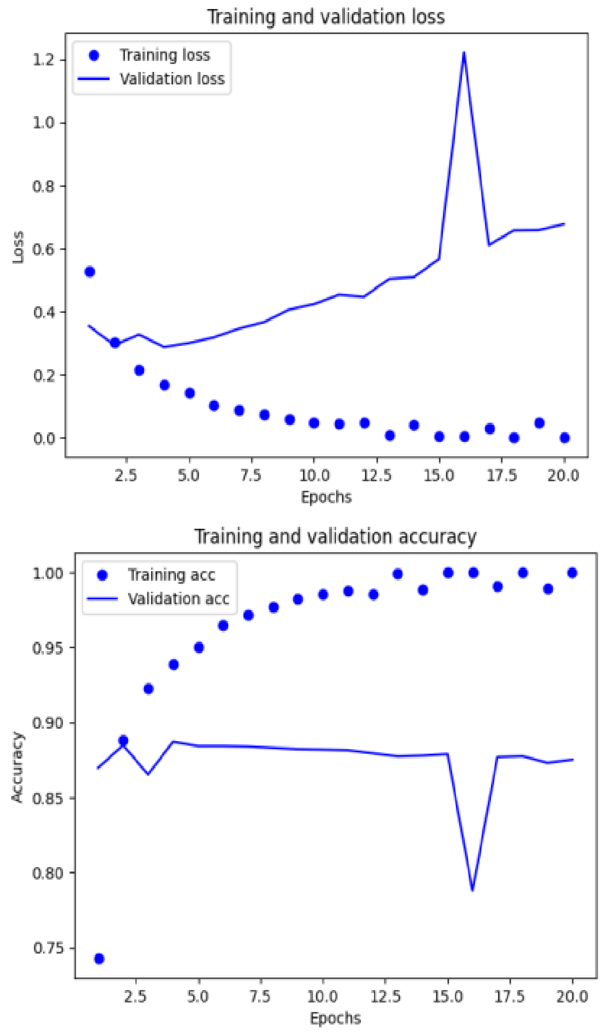
In this exercise, various neural network configurations on the IMDB dataset have been applied- modifying the number of hidden layers, hidden units, activation functions, and loss functions. However, despite using regularization techniques such as L2 regularization and Dropout, the model displayed signs of overfitting after several epochs, where the validation accuracy stagnated, and the loss increased.

1. You used two hidden layers. Try using one or three hidden layers and see how doing so affects validation and test accuracy.

1 hidden layer-



3 hidden layers-



One Hidden Layer

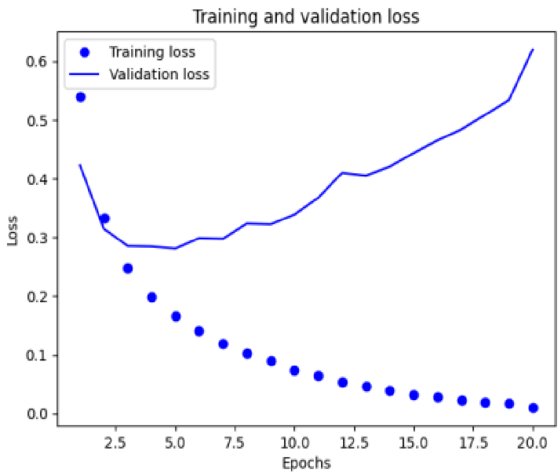
- Accuracy: The model achieved an accuracy of around 99.4% by the final epoch on the training set.
- Validation Accuracy: The validation accuracy peaked at around 88.9%.
- Loss: The training loss decreased steadily, while the validation loss fluctuated a bit, indicating potential overfitting.

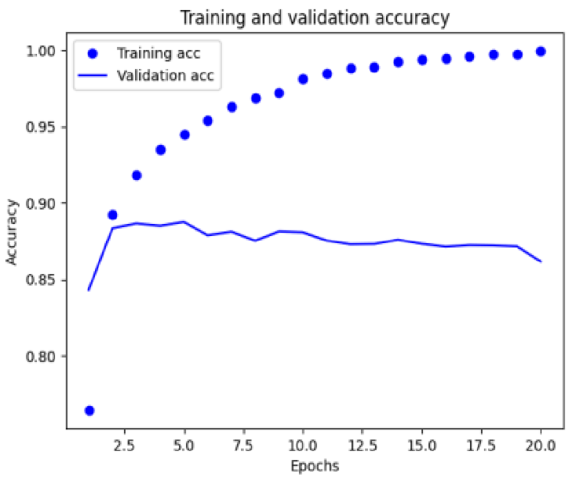
Three Hidden Layers

- Accuracy: The training accuracy reached 99.9%, showing a very high fit to the training data.
- Validation Accuracy: However, the validation accuracy was lower than that of the one hidden layer model, peaking at around 88.5%.
- Loss: The validation loss showed signs of instability, especially towards the end, which suggests the model may have overfit the training data.

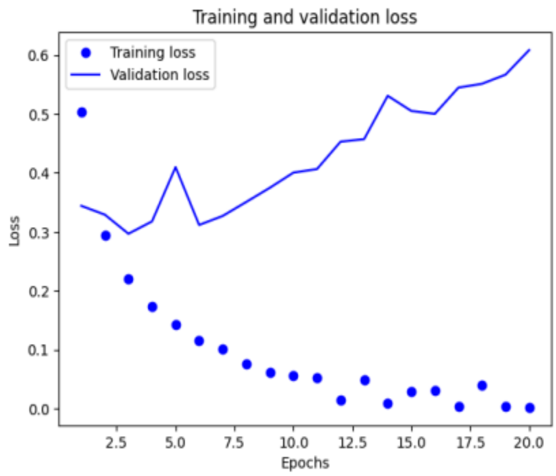
2. Try using layers with more hidden units or fewer hidden units: 32 units, 64 units, and so on.

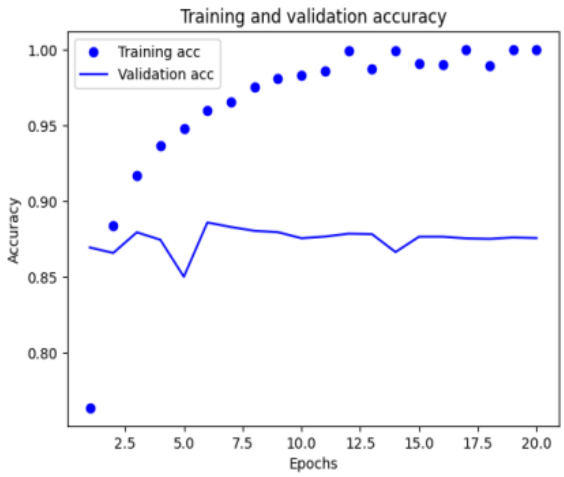
Fewer hidden units(32 units)-





More hidden units(64 units)-





Fewer Hidden Units (16 units)

- Training Accuracy: Achieved a training accuracy of 99.98%.
- Validation Accuracy: The peak validation accuracy was around 88.7%.
- Loss: The training loss decreased consistently, but the validation loss fluctuated, with a final value of around 0.6196.

More Hidden Units (64 units)

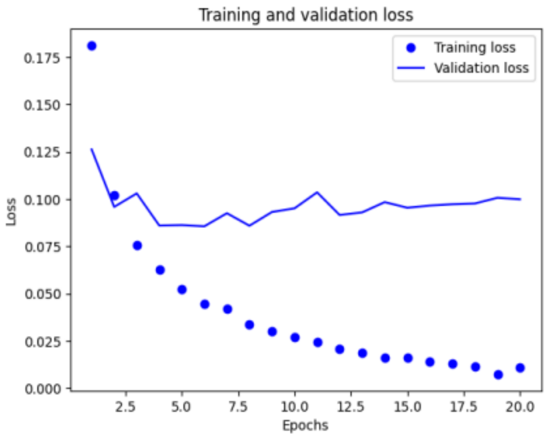
- Training Accuracy: Reached a training accuracy of 100% by the final epoch.
- Validation Accuracy: The validation accuracy peaked at around 88.6%, slightly lower than the model with fewer units.
- Loss: The validation loss showed some instability, with a final loss value of around 0.6087.

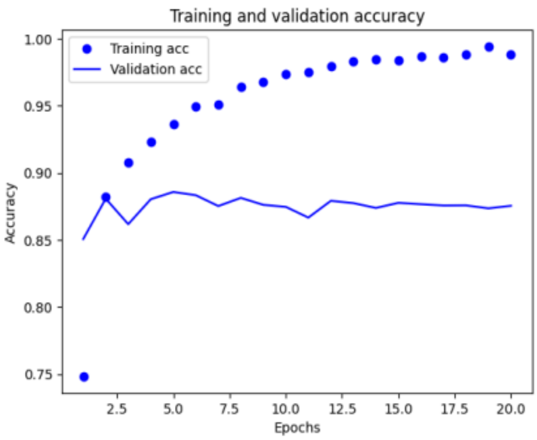
Observations

1. Overfitting: Both models showed signs of overfitting, indicated by the disparity between training and validation accuracy, especially the model with more units achieving perfect training accuracy.

2. Loss Stability: The model with fewer hidden units had a slightly more stable loss trajectory compared to the one with more units.

3. **Try using the mse loss function instead of binary_crossentropy.**





Training Results

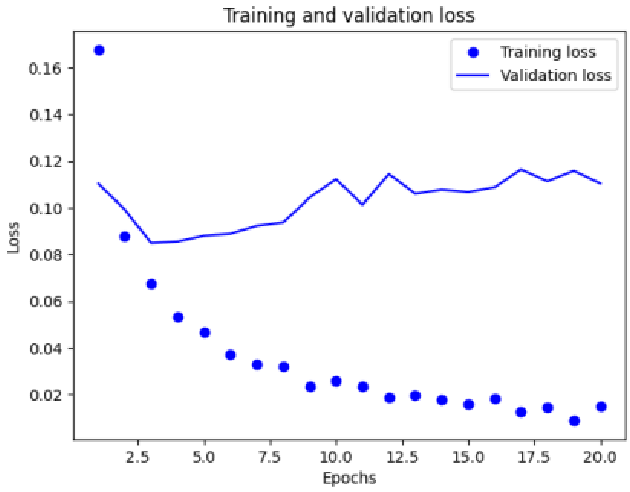
- Training Accuracy: Achieved a peak training accuracy of around 99.4%.
- Validation Accuracy: The peak validation accuracy was approximately 88.8%.
- Training Loss: The loss decreased steadily, reaching about 0.0135 at the end of training.
- Validation Loss: The validation loss fluctuated slightly, with a final value around 0.0999.

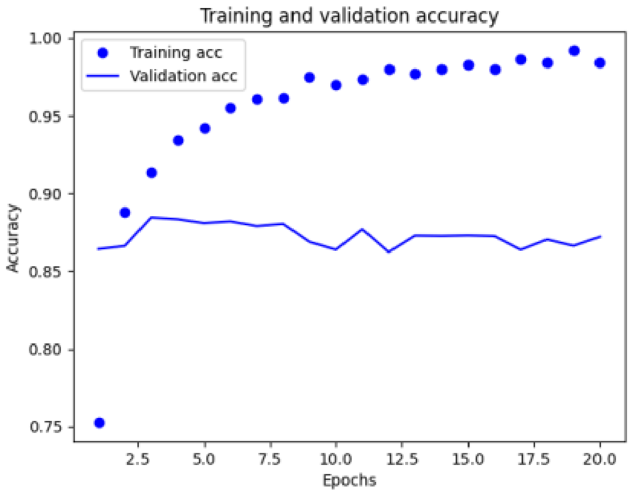
Observations

1. Performance with MSE: While MSE is typically used for regression tasks, it can be applied to binary classification. However, the performance in terms of accuracy might not be as optimal compared to using binary cross-entropy.
2. Validation Stability: The validation accuracy remained relatively stable, though it did not improve significantly over the epochs.

3. Overfitting: As seen in previous models, the model shows signs of overfitting, with the training accuracy significantly higher than the validation accuracy.

4. **Try using the tanh activation (an activation that was popular in the early days of neural networks) instead of relu.**





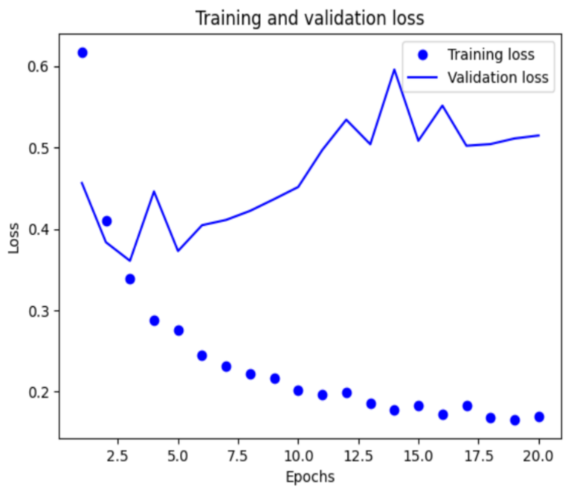
Training Results

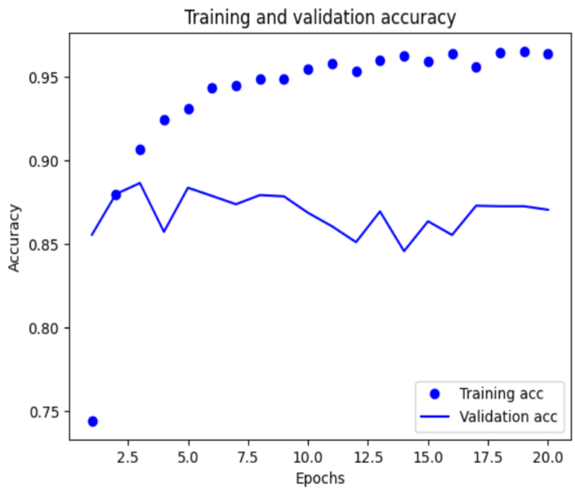
- Training Accuracy: Reached a peak accuracy of around 99.2%.
- Validation Accuracy: The best validation accuracy was about 88.1%.
- Training Loss: Training loss decreased consistently, reaching around 0.0229 by the end.
- Validation Loss: The final validation loss was approximately 0.1103.

Observations

1. Tanh vs. ReLU: The tanh activation function can help mitigate issues with dying neurons compared to ReLU, as it outputs values between -1 and 1. However, its performance can still be sensitive to the initialization and scale of the inputs.
2. Validation Stability: The validation accuracy showed minor fluctuations, similar to previous models, indicating a consistent performance but room for improvement.
3. Overfitting: While the training accuracy remains high, the validation accuracy does not follow suit as closely, suggesting some overfitting may still be occurring.

5. Use any technique we studied in class, and these include regularization, dropout, etc., to get your model to perform better on validation.





Training Results

- Training Accuracy: Peaked at around 97.3%.
- Validation Accuracy: Best validation accuracy reached about 88.0%.
- Training Loss: Decreased steadily, finishing at around 0.1583.
- Validation Loss: Ended at approximately 0.5147.

Observations

1. Dropout Effect: The inclusion of dropout appears to stabilize the validation accuracy compared to earlier models. However, it did not lead to a significant improvement in validation performance.
2. Validation Fluctuations: The validation loss shows some fluctuation, suggesting that the model is still experiencing some overfitting despite the dropout and regularization.
3. L2 Regularization: This has helped reduce the loss and maintain a reasonable validation accuracy, but further tuning (e.g., adjusting the regularization strength) might yield better results.

Overall, models with fewer and more hidden units performed similarly, with no drastic improvement from having more hidden units, though all suffered from overfitting. The MSE loss model showed slightly worse performance compared to binary cross-entropy, while the tanh activation model performed on par with the relu model.