ASSIGNMENT 2

Student's name: Tram Nguyen

ID: 811293917

SUMMARY

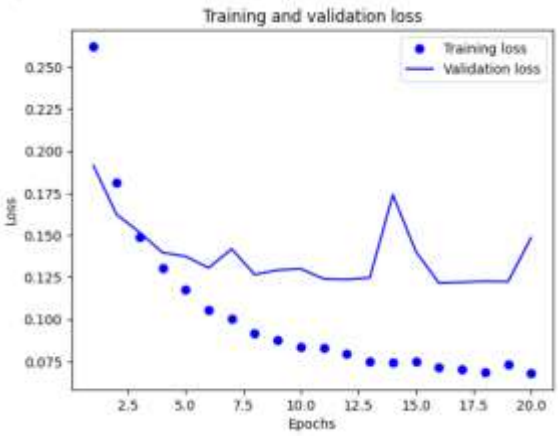
#IMDB MOVIE REVIEW CLASSIFICATION'S PERFORMANCE WITH 32 HIDDEN UNITS VERSUS 64 HIDDEN UNITS:

Model:

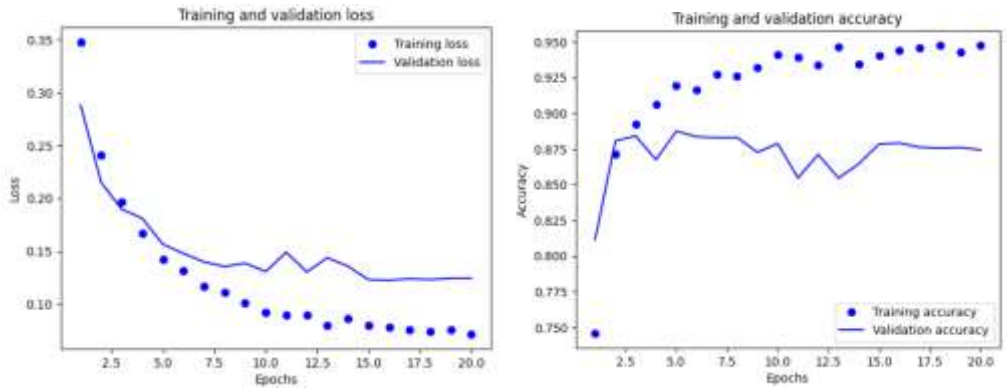
+ Hidden layers: 3
+ Hidden units: 32
+ Activation function: tanh
+ Dropout: 0.5
+ Loss function: Mean squared error (MSE)

Comparision of the model's performance between 32 hidden units and 64 units

+ Training and Validation loss:



- During the training period with **32 hidden units**, at the beginning, both training and validation loss initially decreased, indicating that the model was learning well from the data. Around epoch 3 (or approximately the 3rd epoch), validation loss started increasing while training loss continued to decrease. This is a sign of overfitting. After a few epochs (around the 3rd epoch here), the model starts learning too much about the details and noise of the training data, making it perform worse on unseen data (validation set). When validation loss increases, it means the model is no longer generalizing well.
- However, with **64 hidden units**:



Both training and validation dataset were decreasing. However, at the point epoch = 4, the loss value of validation started going up. The reason because with the more hidden units are provided, the algorithm has more opportunities to study more features of the dataset and need more time to optimize the accuracy. However, it can lead to Overfitting issues.

- On the other hand, Regularization (L2) and Dropout help reduce overfitting. Although validation loss increases after the 3rd epoch, the difference between validation loss and training loss is not too large because your model already has anti-overfitting mechanisms (like regularization and dropout), which prevent the model from overlearning the training data too much. Moreover, MSE (Mean Squared Error) may have made the model harder to optimize for a binary classification task, and using tanh activation might not be as effective as ReLU in deeper neural networks. We can improve it by replacing MSE to binary_crossentropy and changing the activation function from tanh to ReLU.