Advanced Machine Learning Assignment 2

Neural Networks

(Arcot Balraj Tanmaiyee | tarcotba@kent.edu | 811321962)

Questions:

For the IMDB example that we discussed in class, do the following:
1. You used two hidden layers. Try using one or three hidden layers, and see how doing so affects validation and test accuracy.

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

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

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

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

ANS:

SUMMARY

	Test loss	Test Accuracy
Base Model	0.28	0.88
1HL	0.28	0.88
3HL	0.37	0.88
32 HU	0.28	0.88
64 HU	0.28	0.88
MSE Loss Fn	0.09	0.87
Tanh Act	0.28	0.88
L2 Regularization	0.34	0.88
Dropout Model	0.32	0.88

The dataset was used to test 8 distinct neural network models for categorizing reviews as either favourable or negative. The architecture of the models is different.

Base Model: ReLU activation and two 16-unit hidden layers.

1 Hidden Layer: One 16-unit hidden layer with ReLU activation.

3 Hidden Layer: ReLU activation and three hidden layers, each having 16 units.

32 Hidden Units: Two hidden layers (32 units each) with ReLU;

64 Hidden Units: Two hidden layers (64 units each) with ReLU; further expands volume.

MSE Loss Function: Uses MSE loss as a substitute of binary cross-entropy; explores loss function impact.

Tanh Activation Function: Employs tanh activation instead of ReLU; compares activation functions.

L2 Regularization: Applies L2 regularization to prevent overfitting and improves generalization.

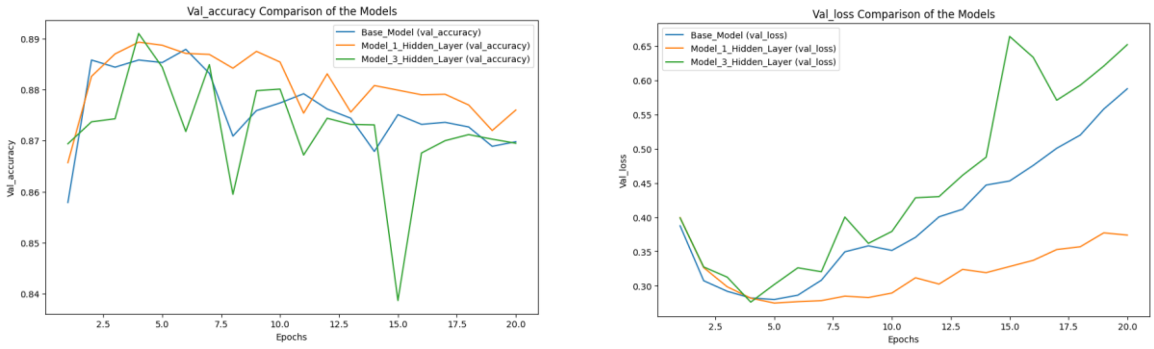
Dropout Model: Uses dropout to reduce overfitting and enhances model robustness.

Analysis

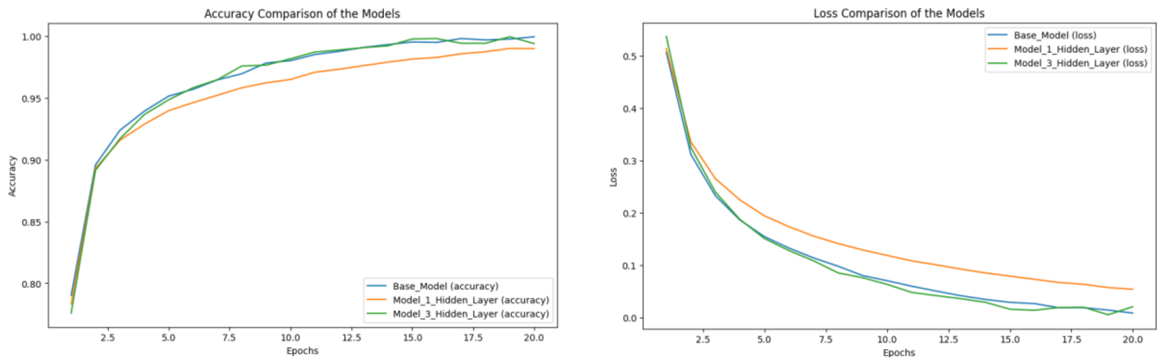
Ans 1: Comparison between **Base model, one hidden layer and 3 hidden layers**:

- With an 88% test accuracy, the Base Model and 1HL model performed the best. This implies that a more straightforward model with one or two hidden layers is adequate to identify the underlying patterns in the data for this specific purpose.
- With three hidden layers and a test accuracy of 86%, the 3HL model did slightly less well and had a greater test loss. This suggests that incorporating more layers into the model may not always result in improved performance and may even cause overfitting.

Graphs showing the validation loss and accuracy



Graphs showing the actual loss and accuracy

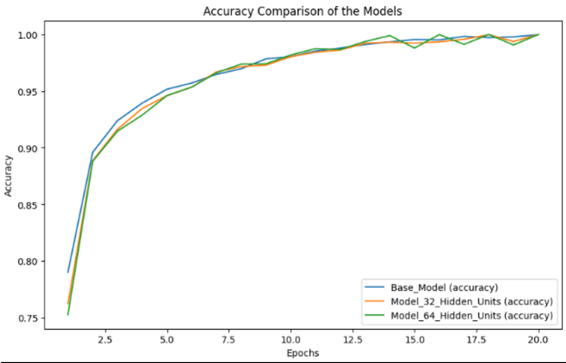


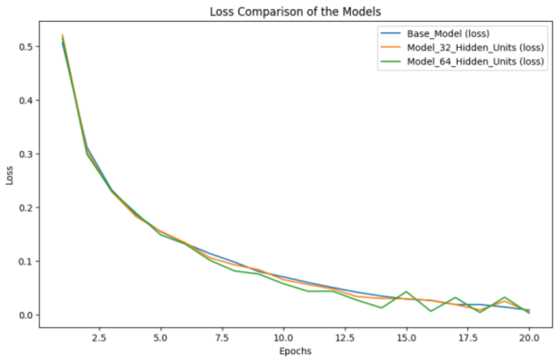
<u>Ans 2</u>: Comparison between **Base model, 32 hidden units and 64 hidden units**:

The test accuracy of the Base Model, 32 HU, and 64 HU models was about the same (88%), suggesting that performance on the given dataset was not considerably affected by the number of hidden units. This implies that the Base Model's 16 hidden units already had enough capacity to identify the pertinent patterns in the data, and that adding more capacity did not yield any new advantages.

The conclusion that model capacity was not a limiting factor in this scenario was further supported by the identical test loss (around 0.28), which was displayed by all three models.

Graphs showing the loss and accuracy of Base Model, 32 hidden units model and 64 hidden units model

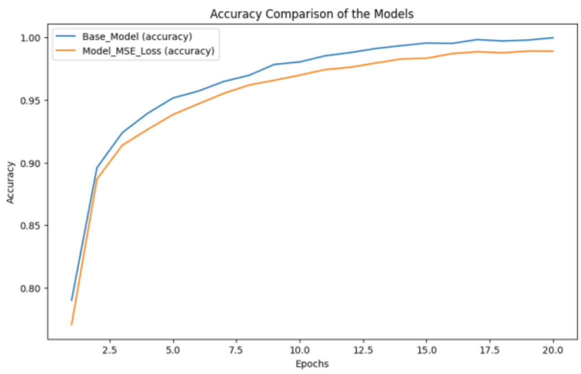


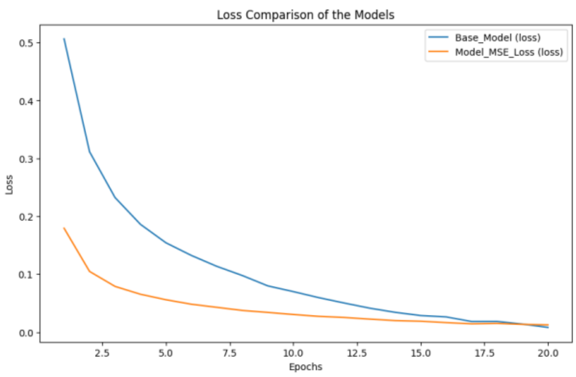


<u>Ans 3:</u> MSE loss function instead of binary_crossentropy

From the above summary table, it is evident that the MSE Loss Function model's test accuracy was marginally lower (87% vs. 88%), but it obtained a substantially lower test loss (0.09) than the Base Model (0.28). This suggests that while employing MSE loss produced a model that suited the training data more closely (lower loss), the overall accuracy on the test data did not significantly increase as a result of this tighter fit. Despite having a greater loss value, binary cross-entropy loss might be a little better option for this specific the task in order to achieve higher accuracy.

Graphs showing the loss and accuracy of MSE and Base Models

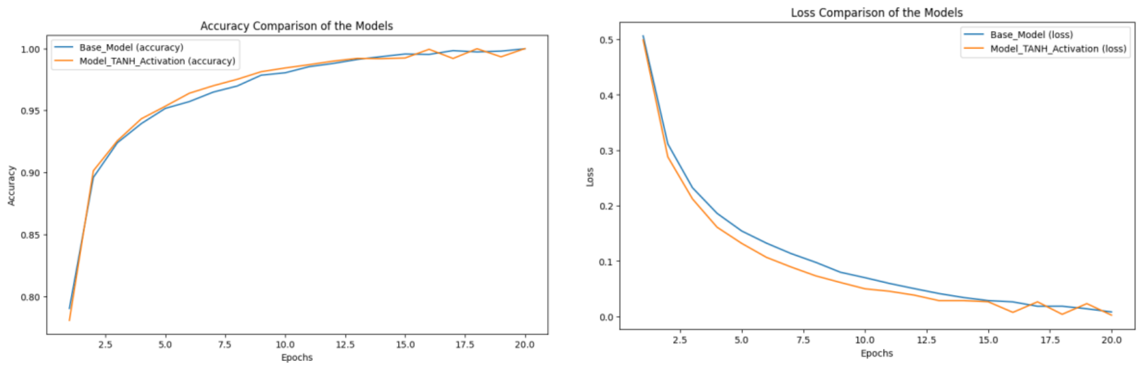




<u>Ans 4:</u> Tanh Activation instead of ReLU

From the summary table when the base model and Tanh activstion model are compared bith Tanh Act model (using tanh activation) and the Base Model (using ReLU activation) have the same test accuracy (88%) and test loss (0.28). This implies that performance is not greatly affected by the choice between the ReLU and tanh activation functions for this given set of data. Both activation functions are capable of accurately predicting unseen movie reviews and efficiently learning the correlations within the data.

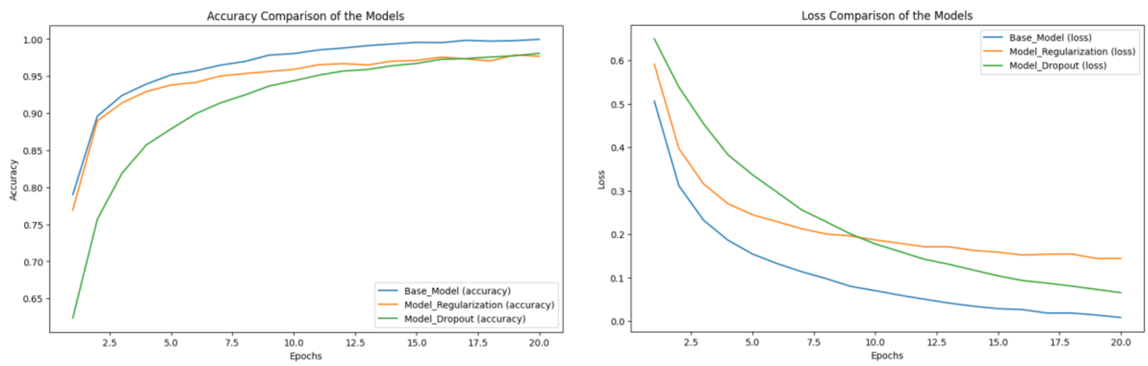
Graphs showing the loss and accuracy of Tanh and Base Models



<u>Ans 5:</u> Comparison between **Base model, L2 regularization and Dropout models.**

While keeping the same test accuracy of 88%, the overfitting mitigation strategies L2 Regularization and Dropout produced a greater test loss than the Base Model (0.34 and 0.32, respectively, against 0.28). This implies that the models' complexity was effectively limited by these regularization techniques, which also stopped them from fitting the training data too closely. However, as evidenced by the higher loss, this restriction also resulted in a marginally less accurate fit to the test data. The total accuracy was constant in spite of this trade-off, suggesting that overfitting may not have been a significant problem for the Base Model on this dataset.

Graphs showing the loss and accuracy Base model, L2 regularization and Dropout models



Conclusion:

The findings show that the Base Model, a simple two-layered network with 16 hidden units, performed well on IMDB sentiment analysis, with an accuracy of 88%. There were no significant advantages from changes in the design, hidden units, or activation function. Although MSE loss resulted in a smaller test loss, its accuracy was slightly reduced than that of the Base Model, and regularization methods did not increase accuracy even if they increased loss.

Graphs comparing all the models