Assignment 1

Question 1:

Initially, a model with two hidden layers with **16 units** and **3 epochs** is considered. Later it is modified to use one, three, and four hidden layers, respectively. Here are the validation and test accuracy results for each of them:

Layers	Loss	Accuracy	Validation_loss	Validation_accuracy	T_Loss	T_Accuracy
1	0.2002	0.9334	0.2804	0.8842	0.3052	0.8432
2	0.1847	0.9361	0.2787	0.8873	0.2943	0.8521
3	0.1757	0.9378	0.2827	0.889	0.3133	0.8333
4	0.1764	0.9374	0.3047	0.8781	0.3435	0.8124

From these results, it can be observed that the original model with two hidden layers performed the best, but the difference in performance was not significant.

Question 2:

The initial model is modified to use different numbers of hidden units **(16, 32, 64, 128)** in the layers, and here are the validation and test accuracy results:

Units	Loss	Accuracy	Validation_loss	Validation_accuracy	T_Loss	T_Accuracy
16	0.19	0.9321	0.2795	0.8873	0.2943	0.8521
32	0.1736	0.938	0.2898	0.885	0.3022	0.8421
64	0.1724	0.9365	0.3586	0.8628	0.3221	0.8351
128	0.1691	0.9359	0.288	0.8855	0.3115	0.8433

From these results, it is observed that using 16 hidden units (same as the original model) resulted in the lowest validation loss which can be implied as the best performance. Using more units resulted in slightly lower performance, but the difference was not significant.

Question 3:

The model is modified to use the mean squared error **(MSE)** loss function instead of binary cross-entropy. Here 2 layers, 16 units, and 3 epochs are considered. The validation and test accuracy results:

Loss function	Loss	Accuracy	Validation_loss	Validation_accuracy	T_Loss	T_Accuracy
mse	0.0416	0.9492	0.1076	0.8598	0.1124	0.8433
binary crossentropy	0.0973	0.966	0.391	0.8719	0.2943	0.8521

From these results, it is seen that the original model with binary cross-entropy loss performed slightly better than using the MSE loss function.

Question 4:

Now the model is modified to use the **tanh** activation function instead of **relu**. Here we considered **2 layers**, **16 units**, and **3 epochs**. The validation and test accuracy results:

Activation Function	Loss	Accuracy	Validation_loss	Validation_accuracy	T_Loss	T_Accuracy
tanh	0.1618	0.9405	0.3061	0.8769	0.3244	0.8343
relu	0.19	0.9321	0.2795	0.8873	0.2943	0.8521

From these results, it can be interpreted that using the relu activation function in the original model resulted in better performance than using tanh.

Question 5:

Here L2 regularization is added to the original model, with a regularization factor of 0.001, and here are the validation and test accuracy results:

L2 regularization: The validation accuracy of 88.7%, test accuracy of 85.4%

No regularization (original model): The validation accuracy of 88.8%, test accuracy of 85.2%

From these results, it can be interpreted that adding L2 regularization slightly improved performance compared to the original model with no regularization.

Conclusion

In conclusion, several modifications to the original IMDB neural network model are tried and found that the original model with two hidden layers, 16 hidden units, binary cross-entropy loss, and relu activation function performed the best.

However, some of the modifications resulted in only minor changes to performance, and other modifications, such as using tanh activation, resulted in decreased performance. It is also found that adding L2 regularization slightly improved performance.

Overall to achieve the best model, 2 hidden layers, 16 hidden units, binary cross entropy loss function, relu activation function, and L2 regularization factor 0.001 is used to obtain the test accuracy of 85.2%.