Neural Networks Assignment_1 IMDB Dataset

Objective:

To build a neural network which can better generalize on the test set.

Hyper Tunning Parameters: 8 models with different layers, nodes and hyperparameters

#No.	Layers	Activation	Nodes	Regularization	Dropout	Optimizer	Loss
Model 1	2	relu	16	-	-	rmsprop	binary crossentropy
Model 2	1	tanh	64	-	-	rmsprop	mse
Model 3	3	relu	64	l2 (0.001)	-	rmsprop	binary crossentropy
Model 4	2	relu	64	-	0.5	rmsprop	binary crossentropy
Model 5	1	tanh	32	-	0.3	adam	mse
Model 6	2	relu	16	-	0.5	rmsprop	binary crossentropy
Model 7	2	relu	16	l2 (0.001)	-	rmsprop	binary crossentropy
Model 8	**2**	**relu**	**16**	**-**	**0.4**	**rmsprop**	**binary crossentropy**

#No.	Epochs	Batch Size	Optimal Epochs	Test Accuracy	Test Loss
Model 1	30	512	4	88.56%	28.78%
Model 2	50	512	3	88.37%	8.64%
Model 3	50	512	3	88.08%	42.28%
Model 4	50	256	3	88.52%	30.25%
Model 5	50	512	5	87.87%	8.85%
Model 6	30	512	7	88.71%	29.30%
Model 7	20	512	6	88.60%	33.88%
Model 8	**20**	**512**	**4**	**88.85%**	**27.29%**

Final Comments:

- Many of the models were built using the "relu" activation function. Relu is a simple and fast-to-compute activation function that improves the speed of the neural network. For the loss function, binary crossentropy was chosen as it is the most suitable for classification problems.
- It's worth noting that the models with the highest accuracy generally used relu as the activation function. Relu is a widely used non-linear activation function. Models 6 and 8, which had 2 hidden layers, also utilized relu as the activation function, with dropout added to control overfitting.
- Model 6 had a dropout rate of 0.5, meaning that only 8 out of 16 nodes were active in each layer. This resulted in an accuracy of 88.71%. When the dropout rate was reduced to 0.4, approximately 10 nodes were active in each layer, leading to an accuracy of 88.85%. This demonstrates the effectiveness of dropout in controlling overfitting and improving accuracy.
- On the other hand, regularization, another method to control overfitting, did not significantly increase the accuracy of the models. Models 3 and 7 were built with l2 regularization at a rate of 0.001, but they did not have a significant impact on the performance of the models.
- Model 2, which used 'tanh' as the activation function and 'mse' as the loss function, achieved a minimal loss of 8.64% and an accuracy of 88.37% on the test data. This suggests that it may be a better-generalized model, but it's important to note that 'mse' is not suitable for binary sentiment classification, which is the task of the IMDb dataset.
- Increasing the number of hidden layers from two to three did not have a significant impact on the model's performance. Models with two or single hidden layers tended to show higher accuracy on the test set.
- Similarly, setting a higher number of epochs initially to allow the model to overfit did not have a major impact. Most models reached their optimal performance within the initial epochs.

These observations provide insights into the choices and performance of the models built in the study. Consideration of activation functions, dropout, regularization, loss functions, and the number of hidden layers can help optimize model performance for future experiments.

Conclusion:

Based on the evaluation of the models, the best-performing model for generalization over the test set is **Model 8**. This model was built with two hidden layers, each consisting of 16 neurons. It utilized a dropout rate of 0.4% and used the rmsprop optimizer along with the binary cross-entropy loss function.

The performance of Model 8 on the test set resulted in an accuracy of 88.85% and a loss of 27.29%. These metrics indicate that Model 8 achieved a high level of accuracy in predicting the sentiment of the IMDb dataset.