CLASSIFICATION OF IMDB DATASET – USING NEURAL NETWORKS

We have used the IMDB dataset to create a simple Neural Network model. The IMDB dataset consists of 50000 movie reviews. The main motive of the assignment was to classify the movie reviews as positive and negative using a neural network or a Binary classification problem. The dataset is divided into 25000 reviews for training and 25000 reviews for testing.

In this particular assignment, we will see how the model performs when we use different hyperparameters and techniques. We will mainly look at the effects of:

1. Using different numbers hidden layers
2. Using different numbers hidden units
3. The choice of the loss function
4. The choice of activation function
5. The use of regularization techniques such as dropout and L2 regularization, early stopping, and Adam optimizer.

We will discuss the various factors that can affect the performance of a neural network and provides a foundation for further experimentation on the neural network models.

Let’s discuss the performance of the model in the following approaches:

1. Effect of Hidden layers: -

A hidden layer is a layer in between input and output layers, where output is produced through an activation function when the artificial neurons take in a set of weighted inputs. We used 2 hidden layers initially and then we used 1 and 3 hidden layers and the results were as follows: -

HIDDEN LAYERS	VALIDATION ACCURACY	TEST ACCURACY	VALIDATION LOSS	TEST LOSS
1	0.8773	0.9897	0.3786	0.0547
2	0.8702	0.9957	0.5665	0.0214
3	0.8676	0.9920	0.6719	0.0286

From the above results, we can infer that the model with one hidden layer performed well over the others. Due to the complexity of the data, adding more hidden layers will cause overfitting, which reduces the model’s performance. The best results were produced by 1 hidden layer because the model was able to capture relevant features in the data.

2. Effect of hidden units:-

We have used hidden units of 32,64,128 and 256. The results were as follows: -

HIDDEN UNITS	VALIDATION ACCURACY	TEST ACCURACY	VALIDATION LOSS	TEST LOSS
32	0.8652	0.9800	0.4387	0.0685
64	0.8659	0.9830	0.4445	0.0590
128	0.8704	0.9928	0.4424	0.0358
256	0.8757	0.9944	0.4288	0.0225

From the above results, it can be inferred that increasing the number of hidden units will allow the model to capture more complex features in the data which helps in better performance. This is clearly depicting in the results which showed the model with 256 hidden units performed the best.

3. Using MSE loss function instead of Binary Cross entropy: -

Loss function	Validation accuracy	Test accuracy	Validation loss	Test loss
MSE	0.8739	0.9916	0.1000	0.0114
Binary Cross Entropy	0.8702	0.9957	0.5665	0.0214

When we used the MSE loss function instead of Binary Cross entropy and compared the results, we found out that the model with the binary cross entropy loss function performed slightly better. This is because this is a binary classification problem, whereas, MSE is used in regression problems where the output is continuous.

4. Using Tanh activation instead of Relu:-

Activation function	Validation accuracy	Test accuracy	Validation loss	Test loss
Tanh	0.8653	0.9997	0.7362	0.0035
Relu	0.8702	0.9957	0.5665	0.0214

From the above results, we can conclude that the RELU activation performed slightly better than the Tanh activation function. Relu is a commonly used activation function because of its ability to handle non-linearity well and how it eliminates the problem of vanishing gradient problem.

5. Effect of using different techniques: -

Techniques	Validation accuracy	Test accuracy	Validation loss	Test Loss
Base Model	0.8702	0.9957	0.5665	0.0214
Adam optimizer	0.8492	1.0000	0.9677	0.0039
L1 Regularization	0.8656	0.8954	0.4990	0.4388
Dropout	0.8758	0.9652	0.5647	0.0987
Early Stopping	0.8844	0.9340	0.2910	0.1819

In this, we have used different techniques to know the model performance. The results showed that dropout and early stopping improved the model's performance the most. Adam is usually the best optimizer but it did not perform well on this dataset. This might be because Adam might have not generalized well to the test set. We might have not optimized the learning rate of Adam properly. The L1 regularization also did not perform well. However, we should bear in mind that the regularization techniques should be selected based on the problem and we should check as many techniques as possible to get the optimal results and enhance the model's performance.

Conclusion: -

We have used different approaches to a binary classification problem to understand the impact of different hyperparameters and techniques on the performance of a neural network model. Our main motive was to classify movie reviews as positive or negative and from the approaches we have used in above, we can conclude the following: -

1. A particular neural network's performance is affected by the number of hidden layers in it. When we add more hidden layers to the mode, it increases the complexity of the model but also increases the chances of overfitting. In this assignment, we have found that the one hidden layer produced the best results. Therefore, it is suggested to start with one hidden layer and only if necessary, increase the number of hidden layers.
2. A model's performance also depends on the number of hidden units in the neural network. When we increase the hidden units in the neural network, it allows the model to capture more complex features in the data, which improves the model's performance. We have seen from the above that the model with 256 hidden units performed the best with the highest validation accuracy and test accuracy. Hence, it is suggested to experiment with the number of hidden units to know the optimal number.
3. For binary classifications, using the binary cross entropy loss function is the best option as from the above results, we saw that the binary cross entropy loss function performed better than the mean squared error loss function. Hence, the choice of loss function affects the model's performance.
4. Using the Tanh activation function improved the model's performance on the test set but overall, it did not perform well over Relu activation function. Hence, it is suggested to use different activation functions to find the right activation function.
5. Using different techniques like Adam, Regularization, Dropout, and early stopping will help to improve the model's accuracy.
6. The Adam optimizer achieved a perfect test accuracy despite achieving lower accuracy than the base model.

7. The L1 regularization technique did not perform well on the test data despite achieving a good enough validation accuracy.
8. The dropout technique achieved a high validation accuracy but a lower test accuracy.
9. The early stopping technique achieved a high validation accuracy and a good test accuracy.

To conclude, this project shows that there are many factors that affect the model's performance and it is very important to try out different techniques to optimize the model's performance for a particular problem.