**Neural Networks**

**Introduction**

Neural Network is a Machine Learning algorithm. It is designed to analyse patterns in data and predict the outcomes from the interconnected nodes organised in the form of layers. The Neural Network consists of an input layer, an output layer, and one or more hidden layers in between. Nodes of a layer are connected to the nodes of the adjacent layers. During the training phase, nodes receive input data from the input layers in the form of tensors and perform operations then process to next layer until a final output is produced in the output layer. In this process, random weights are assigned to the nodes which will be adjusted during the iteration to minimise the difference between predicted and actual value.

**IMDB Dataset:**

IMDB dataset has 50,000 reviews divided into train and test datasets having an equal proportion of positive and negative reviews. IMDB model has been created with two layers of 16 units each having the loss function “binominal cross entropy” and activation function as “Relu”.

**Analysis:**

|  |  |
| --- | --- |
| **Model** | **Accuracy** |
| Actual | 93.28 |
| Increased Layers | 93.66 |
| Decreased Layers | 92.61 |
| Increased Units =32 | 93.59 |
| Increased units =64 | 93.43 |
| MSE loss function | 92.94 |
| Tanh activation function | 93.47 |

From the table it is observed that Accuracy depends on the parameters such as number of layers and units used, the loss function and activation function used in the network.

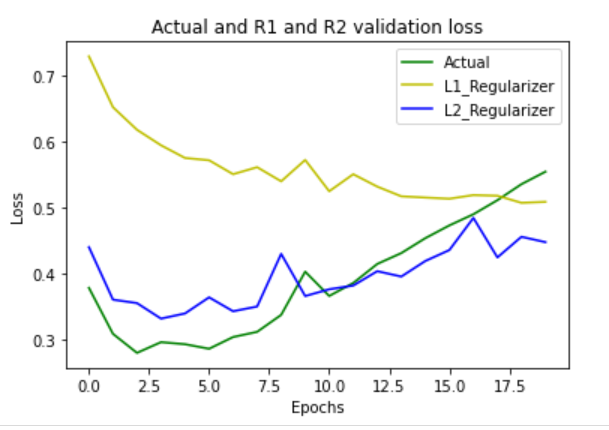
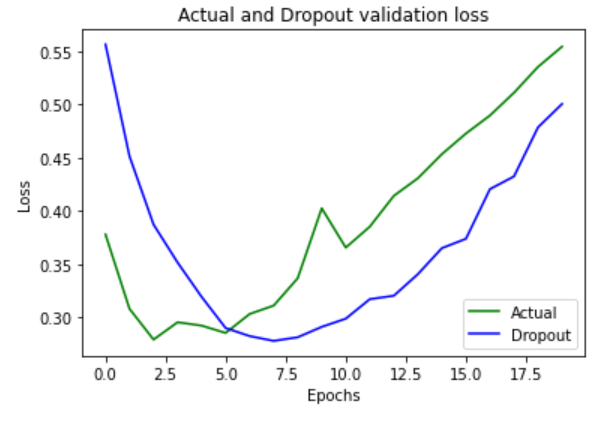
Increase in the number of layers improves the accuracy of the model due to the its ability to learn the complex pattern increases with a greater number of nodes. However, validation loss / overfitting increases as the model learns more training data and performs badly on the unseen data. Whereas Decrease in the number of layers decreases the accuracy of the model and reduces the overfitting as the model learns only the specific data and does not focus on the complex patterns. Hence fewer layers are easier and faster for small datasets and more layers for large datasets.

Increasing the number of units in the network increases the model to learn more data. This increases the model to identify complex patterns leading to better accuracy. Overfitting increases due to model’s low performance in the test data as the model learns more training data.

As IMDB dataset is a classification model, Binary cross entropy as a loss function is the best suited for this model. MSE loss function is mostly used for regression model. Using MSE loss function for the classification model does not address the objective, where predicted error is not measured as the difference between true and predicted values but as the number of misclassified samples. Hence accuracy decreases.

For a classification model, Both Relu and tanh activation function can be used depending on the Architecture and problem type . Relu is computationally efficient. It’s non saturating nature will not allow the large values to approach to its maximum or minimum values and only a selected neurons are activated makes the model less prone to overfitting. On the other hand, tanh activation function has symmetry to the origin which is useful for the data having positive and negative values and has continues 1st and 2nd derivatives.

Regularisation in the neural network is used to improve generalisation which allows it to perform well in both training and test data and prevent overfitting which is the common problem with the machine learning models. Below graphs shows that with the increase in the number of epochs, model learns more training data resulting in the more validation loss. Regularisation techniques help to decrease the validation loss as the model captures only the required data.

It encourages the weights to be small due to the addition of the penalty term to the loss function which helps in the reduction of overfitting. L1 regularisation adds the sum of the absolute values of the weights to the loss function whereas L2 regularisation adds the sum of squares of the weights.

In the dropout technique selected neurons are dropped out during the training. This will let the other neurons to learn only the specific features that helps in the reduction of overfitting in the network.