

## ASSIGNMENT - 4

### TEXT AND SEQUENCE

#### Introduction:

In order to locate reviews and ultimately provide a binary prediction that distinguishes between positive and negative reviews, we are training the RNN model on the IMDB data set. RNNs, a type of neural network model intended to process data sequentially in a single direction, have a repeated connection that supports this process.

| Review words | Max length   | Training size | Validation size | Embedding accuracy |
|--------------|--------------|---------------|-----------------|--------------------|
| 150          | 10000        | 100           | 10000           | 75.23              |
| 150          | 10000        | 1000          | 10000           | 85.81              |
| <b>150</b>   | <b>10000</b> | <b>10000</b>  | <b>10000</b>    | <b>87.32</b>       |
| 150          | 20000        | 15000         | 10000           | 81.24              |
| 200          | 20000        | 20000         | 10000           | 86.16              |

#### Different samples:

- The network architecture utilized to run the model in the previously mentioned samples was extremely basic, consisting only of one embedding layer, and the model was trained on a limited amount of data. There were not many variances in accuracy across the data set. However, the model's best accuracy for 150 top words was 87.32.
- At 200 samples, the sample reaches its maximum accuracy; after that, it changed, but only a little. The model's performance will change if the hyperparameters are changed. which include drop-out rate, dense layers, and learning rate.

#### Accuracy of the pre-trained model:

| Review words | Max length | Training size | Validation size | Pre-trained accuracy |
|--------------|------------|---------------|-----------------|----------------------|
| 150          | 10000      | 100           | 10000           | 80.1                 |
| 150          | 10000      | 1000          | 10000           | 81.27                |
| 150          | 10000      | 20000         | 10000           | 80.1                 |

The layers and vector dimensions of the pre-trained network model. Because the training of the model is unable to recognize fresh data, the model's performance seems to be low on unknown data, which affects the higher performance on the test set.

#### Conclusion:

All models were included in the final model. The performance of the embedding layer model rose with sample size, reaching an accuracy of 88.16 for the fully trained model and 80.1 for the pretrained one. We may infer that convolution layers, learning late, and dense layers all have an impact on performance. The model's performance was 8.32 from all sample sets in the embedding layer with top reviews of 150 samples.

