Assignment-04:

(Analysing text and sequence data using RNNs)

**Recurrent Neural Network(RNN):**

**Introduction**:

This report aims to investigate the effects of different hyperparameters on the sentiment analysis model's performance through modifications to the existing IMDB dataset**.**

**Methodology**:

As per the requirements in the problem, the conditions are:

1. Cutoff reviews after 150 words.  
2. Restrict training samples to 100.  
3. Validate on 10,000 samples.  
4. Consider only the top 10,000 words.

Here we considered a basic model with embedded layer and a pretrained word embedding model. Glove embedding is the technique used here for pretrained word embedded models.

For both the basic embedded model and the pretrained embedded models different sample sizes such as 100, 200 and 500 are implemented and the results obtained are as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of the model** | **Training sample size** | **Training accuracy** | **Validation accuracy** | **Training loss** | **Validation loss** |
| Embedded\_model | 100 | 1.0000 | 0.5500 | 0.5378 | 0.6899 |
| Embedded\_model2 | 200 | 1.0000 | 0.6300 | 0.1652 | 0.6439 |
| Embedded\_model3 | 500 | 0.0970 | 0.6400 | 0.4765 | 0.6709 |
| Pretrained model1 | 100 | 1.0000 | 0.5029 | 0.0223 | 1.7660 |
| Pretrained model2 | 200 | 0.9950 | 0.5868 | 0.0398 | 0.7199 |
| Pretrained model3 | 500 | 1.0000 | 0.6075 | 0.0288 | 0.7738 |

From the above results we can say that as we increase the training sample size, the validation accuracy of the model increases and thus a better performance is obtained for both the simple embedded models and the pretrained models.

When the LSTM is used the results obtained are as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of the model** | **Training sample size** | **Training accuracy** | **Validation accuracy** | **Training loss** | **Validation loss** |
| Simple LSTM | 6000 | 0.9688 | 0.7517 | 0.1016 | 0.5276 |

**Conclusion**:

From all the results obtained, we can say that by increasing the training sample size the performance of the model increases in general for both the embedded and pretrained models.

The LSTM model outperforms all the other models and produces best results. With a larger dataset it might perform even better.