# **Assignment 3 Text and Sequence**

Shivani Haridas Pitla

**Objective**: To gain a deeper understanding of how various parameters affect the performance of the model, and to determine which variations work best under different conditions. Specifically, the modifications include limiting the number of training samples, considering only the top 10,000 words, and comparing the performance of a model using an embedding layer to one that uses a pre-trained word embedding.

**Training sample sizes and their results**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Training sample size | Scratch or pre-trained | Test loss | Test accuracy |
| 1 | - | Scratch | 0.35 | 0.86 |
| 2 | 100 | scratch | 0.69 | 0.50 |
| 3 | 100 | Pre-trained | 3.3 | 0.50 |
| 4 | 1000 | Embedding layer | 0.68 | 0.56 |
| 5 | 15000 | Embedding layer and conv1D | 0.41 | 0.81 |
| 6 | 30000 | Embedding layer and conv1D | 0.43 | 0.80 |
| 7 | 15000 | Pre-trained | 4.8 | 0.50 |
| 8 | 30000 | Pre-trained | 0.95 | 0.52 |

1. Comparing the model 1 and 2 the first model, where no modifications were made, the model achieved a test accuracy of 86%, indicating that it was able to accurately classify reviews as either positive or negative based on their text content. However, when the reviews of the train were restricted to 100 samples, the performance of the model dropped significantly, achieving a test accuracy of just 50%
2. With cutoff reviews after 150 words, restricting the training sample to 100, validating on 10000 samples, and considering only the top 10,000 words both the models with an embedding layer and a pre-trained word embedding achieved almost the same test accuracy of 50%
3. To determine at what point the embedding layer outperforms the pre-trained word embedding, the number of training samples was varied. When using 1,000 training samples, the embedding layer performed better, achieving a test accuracy of 0.56, compared to test accuracy of just 0.50 with the pre-trained word embedding at training sample 100.
4. Next, as just an increase in training sample size had a very slight impact on accuracy, we have used conv1D along with the embedding layers while also increasing the training sample size to 15000 and 30000. Which resulted in better test accuracy that is 81% and 80% a
5. When using 15000 and 30000 training samples the models with embedding layer and Conv1d outperformed the pre-trained word embedding test accuracies.

Overall, these results suggest that the performance of the model can be highly dependent on the specific parameters used, including the number of training samples, the word embedding used, the maximum review length, etc. In general, using an embedding layer may be more effective when working with smaller datasets.