

AML ASSIGNMENT - 4

REPORT

INTRODUCTION:

The primary objective is to understand completely how different parameters influence the model's performance and to identify the versions that perform the best in various scenarios. Limiting the size of training samples, taking into consideration of the top 10,000 words, and evaluating the performance of a model using an embedding layer with that of a model using a pre-trained word embedding are the few modifications made.

Scratch Model's Results:

MODELS	TRAINING SAMPLE SIZE	LOSS	ACCURACY
1	-	0.32	0.84
2	100	0.71	0.52

Pre-Trained Model's Results:

MODELS	TRAINING SAMPLE SIZE	LOSS	ACCURACY
1	100	3.26	0.51
2	15000	4.92	0.52
3	30000	0.90	0.55

Embedding layer (1) and Embedding layer and conv1D(2,3) Model's Results:

MODELS	TRAINING SAMPLE SIZE	LOSS	ACCURACY
1	1000	0.66	0.59
2	15000	0.39	0.85
3	30000	0.45	0.82

RESULTS:

- By Comparing both the Scratch Model's we got the Accuracy to 84% for Model 1 whereas Model 2 got just 52% Accuracy and no modifications were made on these models.
- It indicates that based on the text of the reviews, it could properly classify them as either positive or negative. The model's performance, however, drastically declined when the evaluations of the train were limited to 100 samples, resulting in a test accuracy of only 52%.
- The models using an embedding layer and a pre-trained word embedding both resulted the same test accuracy when just taking into consideration the top 10,000 words.
- The training sample size was altered to see whether the embedding layer outran the pre-trained word embedding. With 1,000 training sample, the embedding layer performed better, acquired a test accuracy of 0.59 whereas the pre-trained word embedding at training sample 100, which only achieved a test accuracy of 0.51.
- We employed conv1D along with the embedding layers and enlarging the training sample size to 15000 and 30000 because an increased training sample size had very little effect on accuracy. It had a very huge impact on test accuracy that was improved to 85% and 82%.
- The models with embedding layer and Conv1d outperformed the pre-trained word embedding test accuracies when using 15000 and 30000 training sample size.

Conclusion :

The model 2 that used embedding layer and Conv1d had the best accuracy of 85%. So finally, these findings indicate that the model's performance may be significantly influenced by the particular parameters utilized, such as the quantity of training data, the word embedding, the maximum review duration and more. Generally, when working with smaller datasets, adopting an embedding layer could be more efficient.