

Assignment 4 - Text and Sequence Data

Cheyenne Morrow

Summary Report

The trainable embedding model reached a final validation accuracy of about 0.59, and the pretrained GloVe model reached about 0.50. The accuracy graph showed that the trainable model kept improving with each training round, while the pretrained model stayed close to random accuracy. The results show that even with a small amount of data, the trainable embedding was able to learn the language style and common words in the IMDB reviews more effectively than the pretrained version. If more training samples were added, the trainable model would likely perform even better as it learns more about how words are used in this dataset.

Model	Accuracy
Trainable Embedding	0.59
Pretrained	0.50

The pretrained model did not perform as well in this case because its weights were based on general English text that did not fully match the words and tone found in movie reviews. The trainable embedding, on the other hand, learned patterns that were specific to the IMDB dataset, which helped it fit the task more closely. The model could be improved by adding dropout layers to prevent overfitting, increasing the number of LSTM units or layers, and testing more training samples to see how performance changes with more data.

The trainable embedding model gave stronger results after learning from the IMDB data. As more data becomes available, this method is likely to improve even more, while pretrained embeddings tend to be more helpful when only a small amount of labeled data is available.