

## Assignment 4

### Experiment Setup

The experiments were conducted based on the IMDB example from Chapter 11, modifying the following parameters:

- **Review Length:** Reviews were truncated to 150 words.
- **Training Samples:** The number of training samples was varied to assess its impact on performance.
- **Validation Samples:** Validation was performed on a fixed set of 10,000 samples.
- **Vocabulary Size:** Only the top 10,000 words were considered.
- **Embedding Layers:** Two types of embedding layers were evaluated before the Bidirectional LSTM layer:
  - An Embedding Layer trained from scratch.
  - A Pre-trained Word Embedding (GloVe).

Approach		Test accuracy
Cutoff reviews after 150 words		0.823
Restrict training samples to 100		0.5
Validate on 10,000 samples		0.865
Consider only the top 10,000 words		0.870
Before the Bidirectional layer, a) an embedding layer		0.861
Before the Bidirectional layer, b) a pretrained word embedding.		0.877
training sample	100	0.594
	1000	0.743
	2000	0.750
	4000	0.810
	6000	0.834
	8000	0.832
	10000	0.855

### Analysis

1. **Impact of Embedding Type:** The pre-trained word embedding (GloVe) initially outperformed the embedding layer trained from scratch (0.877 vs 0.861). Pre-trained embeddings bring prior knowledge learned from a large corpus.
2. **Impact of Training Sample Size:** With only 100 training samples, the accuracy was poor (0.594). As the number of training samples increased, the accuracy improved significantly. However, the gains started to diminish beyond 6000 samples.
3. **Effect of Limiting Review Length and Vocabulary:** Cutting off reviews at 150 words (0.823) and considering only the top 10,000 words (0.870) provided a baseline performance and helps in reducing noise and focus on the most important features.

**Conclusion**

Pre-trained embeddings initially provide better performance, especially when training data is limited. As the training data increases, the performance gap between pre-trained embeddings and the embedding layer diminishes.

For prediction improvement, pre-trained embeddings are more suitable when data is limited. With sufficient data, training an embedding layer from scratch can achieve comparable results. The optimal number of training samples appears to be around 6000-8000, beyond which the gains in accuracy become marginal.