ADVANCED MACHINE LEARNING _ASSIGNMENT – 4

Training Models with Varying Sample Sizes:

The values are configured as follows:

Review cutoff is 150 words.

100 training samples are available.

10,000 words is the number of validated samples.

The models' test accuracy and loss are summarized in the table below. The models were trained with varying training sample sizes ranging from 100 to 10,000.

Sample size	One hot encoded sequence		Embedded		Embedded masked		Pre trained	
	Test loss	Test accuracy	Test loss	Test accuracy	Test loss	Test accuracy	Test loss	Test accuracy
100	0.6218	0.6629	0.6708	0.5857	0.6586	0.608	0.6787	0.6134
500	0.697	0.565	0.7138	0.6067	0.7351	0.6192	0.6193	0.6696
2000	0.6596	0.5992	0.7226	0.7108	0.8353	0.7032	0.5391	0.7248
5000	0.4891	0.7961	0.5375	0.7924	0.765	0.7645	0.5137	0.7836
10000	0.4380	0.801	0.4455	0.798	0.4349	0.811	0.4573	0.783

100 train samples and 10,000 validations:

- **Initial Setup:**

1. For this assignment, the IMDB review dataset has been imported.

2. The model's first setup involved gathering 100 training samples, each review having a maximum word length of 150, and a total of 10,000 words were used as input.

3. In addition, 10,000 validation samples of both positive and negative reviews are used to validate this model.

4. Since the classification model had an optimizer named Adam, the loss function known as "binary cross-entropy" was employed.

• Trained Models:

1. Using accuracy as a performance criterion, four models were trained, verified, and tested using the original configuration.

2. A hot-encoded sequence model has demonstrated 0.801 test accuracy and 0.4380 test loss.

3. The test accuracy and loss for the embedded model without masking were 0.4455 and 0.798, respectively.

4. The test accuracy and loss of an embedded model with masking were 0.4349 and 0.811, respectively.

5. A pre-trained model called Global Vectors for Word Representation (GloVe) produced test accuracy and loss values of 0.783 and 0.4573, respectively.

The investigation's findings indicated that, when it came to sentiment analysis, RNNs with embedded layers outperformed alternative word embedding strategies, like one-hot encoded sequences. Test accuracy and test loss were consistently higher with the embedded layer-based models than with other methods.

Additionally, several embedded layer types such as conventional embedded and masked embedded layers are compared. When compared to masked embedded layers, the normal embedded layer-based models performed marginally better in terms of test accuracy. Although the masking technique enables the model to ignore padding tokens and concentrate only on the word embeddings themselves, resulting in more meaningful representations and improved performance, it is evident from this model implementation that masking has no effect on the provided IMDb dataset.

Conclusion:

- For all cutoff reviews and training sample sizes, the validation accuracy of the embedding layer model is greater than the test accuracy. This implies that there's a chance the model is overfitting the training set.
- For certain cutoff reviews and training sample sizes, the validation accuracy of the pre-trained model is higher than the test accuracy, but lower for other situations. This indicates the model's performance is less consistent than that of the embedding layer model.
- The results defied the widespread belief that pre-trained embeddings improve model performance, showing that the simple embedding layer model outperformed the

pretrained model. Overall, it's critical to remember that the pre-trained model in this case is not ideal for the given job and hasn't had its embeddings properly adjusted during training. In essence, improving the embeddings could result in higher performance.

- Finally, since these results are based on a small number of training samples and a constrained set of hyperparameters, we need to take caution when applying conclusion from them. It's feasible that further training data or alternative hyperparameters will produce different results.