

NLP Assignment 1 - POS Tagging

Harshavardhan P - 2021111003

> All evaluations were done on RNN model with 2 linear layers for classification, and run for 15 epochs. Further runs noticed overfitting for the model, and hence the model was stopped at 15 epochs.

Hyperparameters

ELMo

- `embedding_size` = 200
- `hidden_size` = 100 - Always has to be half of the embedding size to account for the concatenation of the forward and backward hidden states
- Trained on all the sentences, but all the sentences were trimmed to 90th percentile of the sentence lengths, and padded to the maximum length of the trimmed sentences
- `epochs` = 15
- `batch_size` = 128
- `optimizer` = Adam with `lr` = 0.001
- `loss` = CrossEntropyLoss

RNN

- `embedding_size` = 200
- `hidden_size` = 256
- `activation` = relu
- `epochs` = 15
- `batch_size` = 128
- `optimizer` = Adam with `lr` = 0.001
- `loss` = CrossEntropyLoss

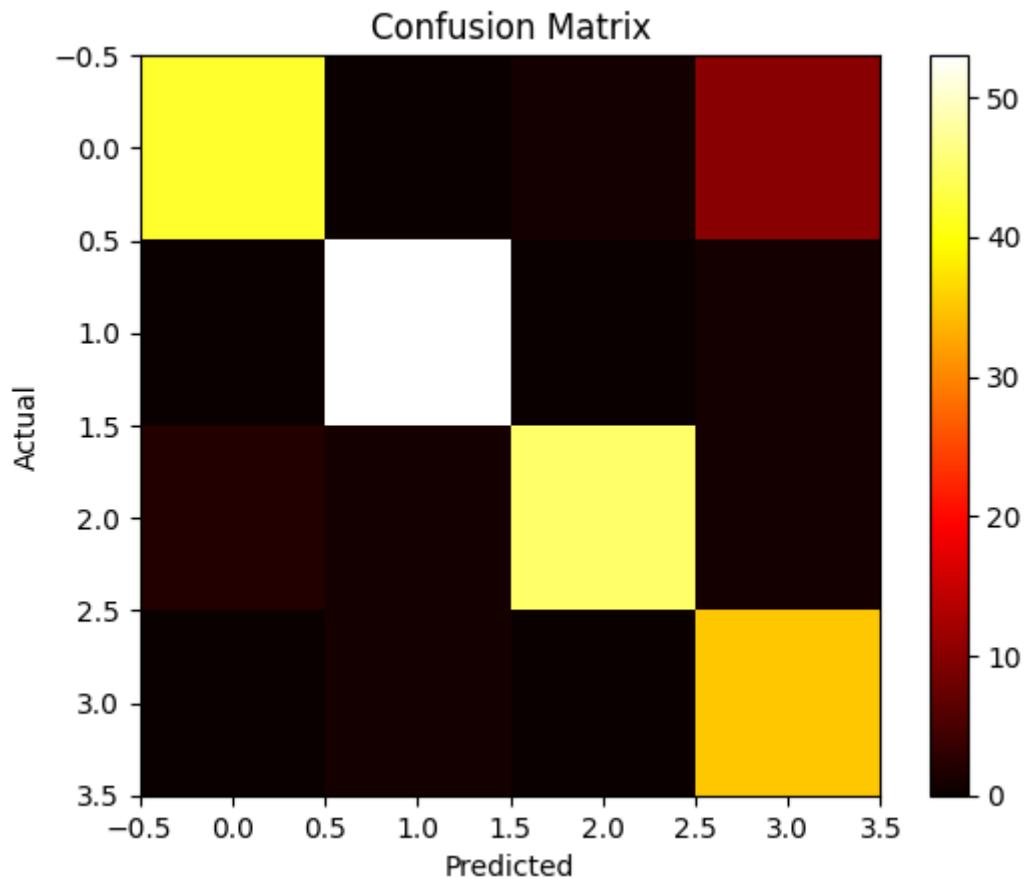
Graphs and Evaluation Metrics

ELMo

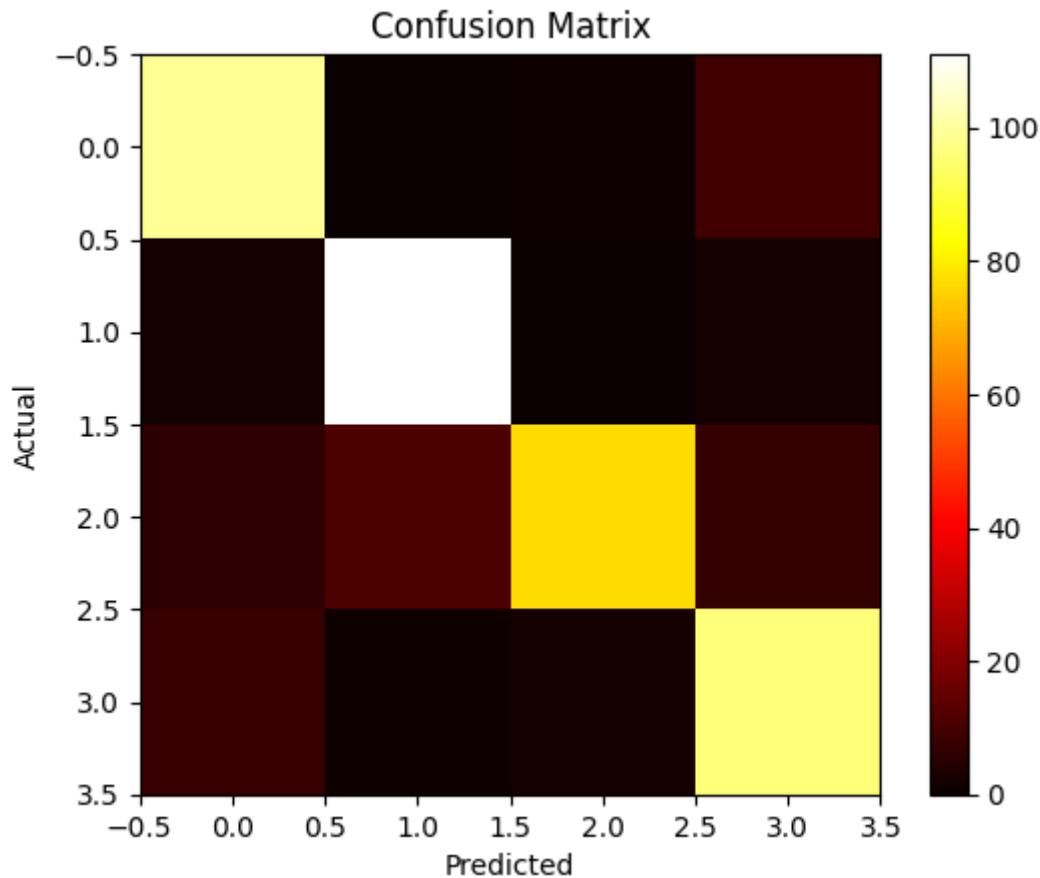
RNN Classifier on ELMo Embeddings

- Const Is:
 - Training Accuracy: 0.9583333333333334
 - Training Precision: 0.9595352564102564
 - Training Recall: 0.9583333333333334
 - Training F1 Score: 0.9585551236044655
 - Test Test Accuracy: 0.8634259259259259

- Test Precision: 0.8709821109760044
- Test Recall: 0.8634259259259259
- Test F1 Score: 0.8636638747632817
- Training Confusion Matrix:



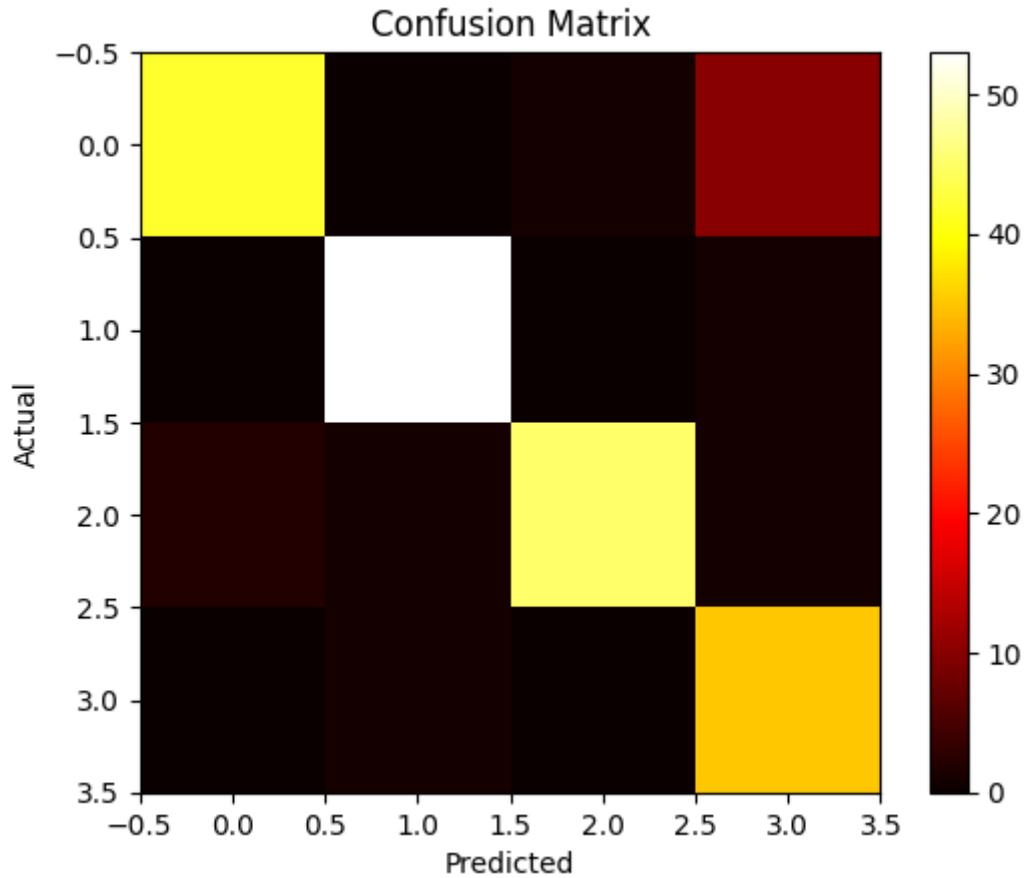
- Test Confusion Matrix:



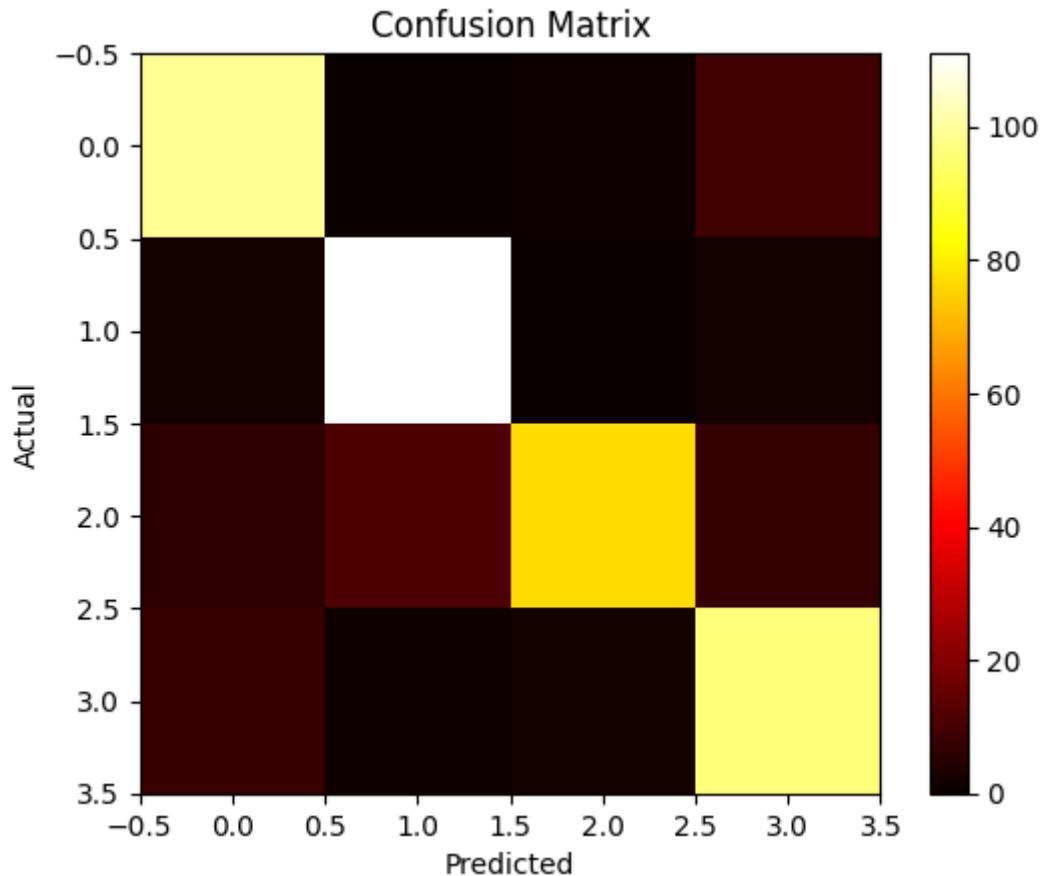
- Learnable Is:

- Training Accuracy: 0.9114583333333334
- Training Precision: 0.92380503111597
- Training Recall: 0.9114583333333334
- Training F1 Score: 0.9124644119744595
- Test Test Accuracy: 0.8865740740740741
- Test Precision: 0.8910486888228436
- Test Recall: 0.8865740740740741
- Test F1 Score: 0.8854400451951334

- Training Confusion Matrix:

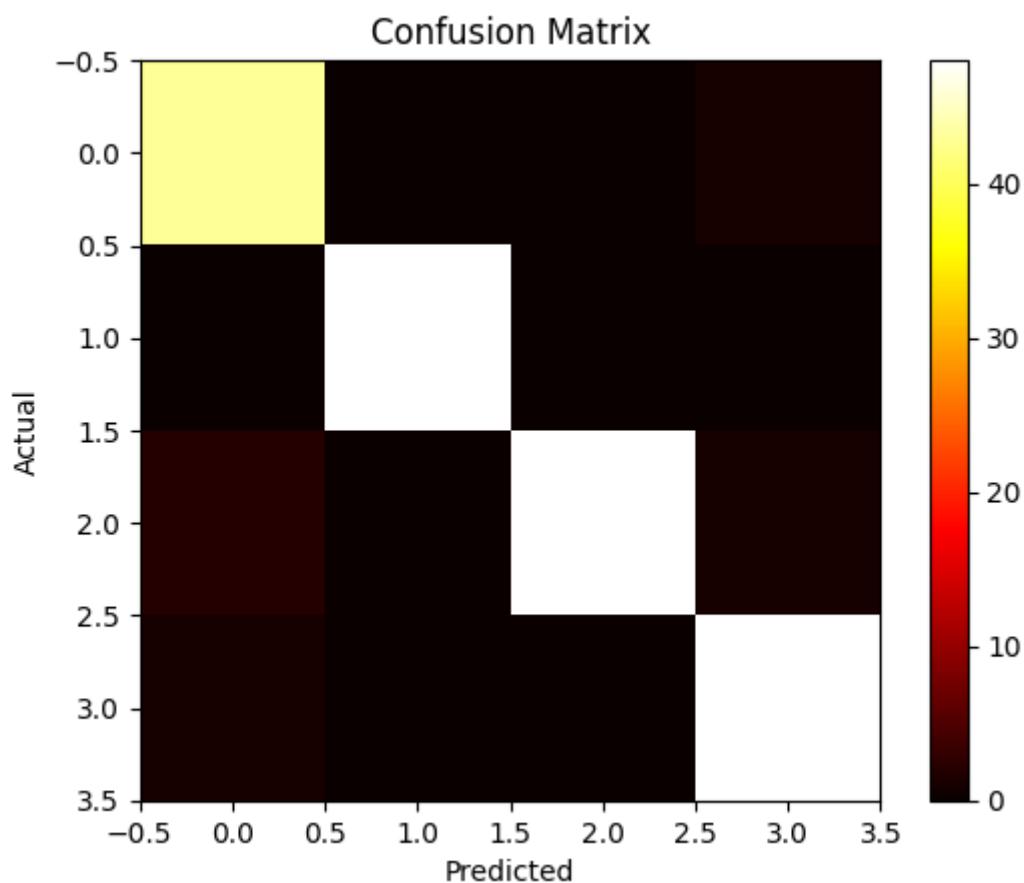


- Test Confusion Matrix:

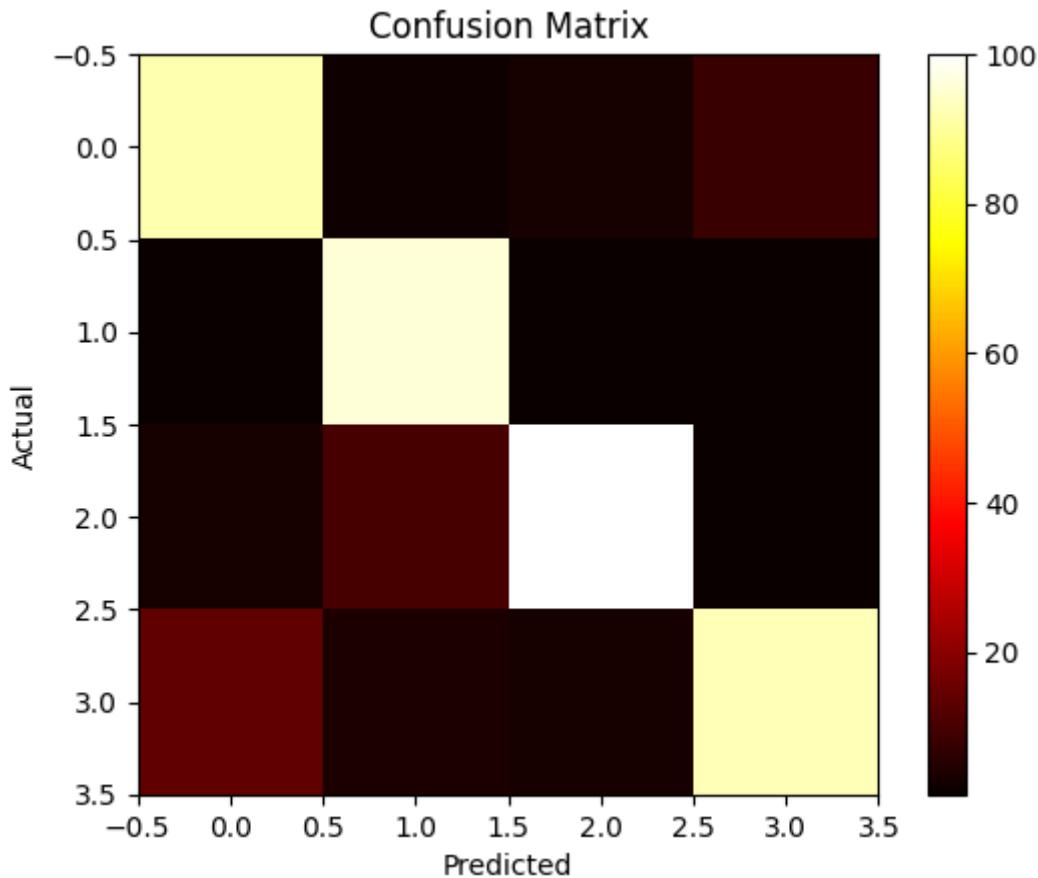


- Learnable Is with 2 linear layers:

- Training Accuracy: 0.9739583333333334
- Training Precision: 0.9748460144927535
- Training Recall: 0.9739583333333334
- Training F1 Score: 0.9740319865319865
- Test Accuracy: 0.9019444444444444
- Test Precision: 0.9146051379283391
- Test Recall: 0.9219444444444444
- Test F1 Score: 0.9115450919025843
- Training Confusion Matrix:



- Test Confusion Matrix:



Analysis and Observations

Trained Models

Trained models can be found here: [Google Drive](#) in this google drive folder.

Analysis of performance of ELMo on LM and Classifier tasks

- ELMo pretraining for the purpose of language modeling gave very good results with 52% accuracy which is very high for a language modeling task.
- The classifier task also gave good results with 91% accuracy which is very high for a sentence classifier task.

Comparison of ELMo with word2vec-skipgram and SVD

- ELMo gave similar results to word2vec-skipgram and SVD in classification tasks.
- As SVD was little bad compared to word2vec-skipgram, ELMo was better than SVD in classification tasks, but little behind word2vec-skipgram, or sometimes similar.
- Best model in my case was word2vec-skipgram, but ELMo was also very good in terms of performance.

Refer to previous assignment report for metrics on the other models (can be found with the models).