Introduction to NLP

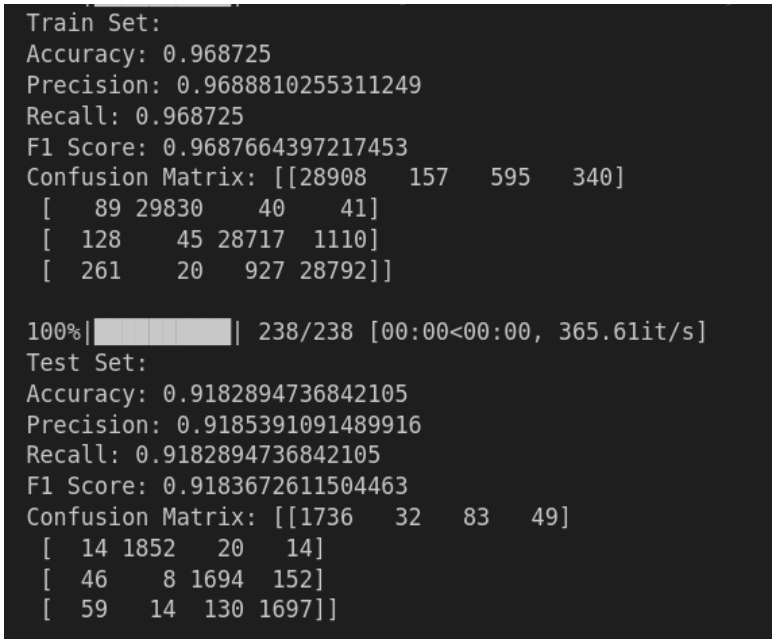
Assignment-4

Submitted by - Ishit Bansal (2021101083)

Report

Hyperparameter Tuning

- Trainable λs



- Frozen λs

```
Train Set:
Accuracy: 0.947025
Precision: 0.9471167941478705
Recall: 0.947025
F1 Score: 0.9470310359340889
Confusion Matrix: [[28362   318   827   493]
 [   66 29771   121    42]
 [  434   108 27690  1768]
 [  483   122  1575 27820]]

100%|██████████| 238/238 [00:00<00:00, 372.70it/s]
Test Set:
Accuracy: 0.9140789473684211
Precision: 0.9142387814401838
Recall: 0.9140789473684211
F1 Score: 0.9140920125807337
Confusion Matrix: [[1739   38   73   50]
 [  12 1853   27    8]
 [  43   17 1677  163]
 [  54   19  149 1678]]
```

- Learnable Function

```
Train Set:
Accuracy: 0.939475
Precision: 0.940556274108918
Recall: 0.939475
F1 Score: 0.9395294676040294
Confusion Matrix: [[27887   344   724  1045]
 [  165 29585    87   163]
 [  549   196 26771  2484]
 [  285   115  1106 28494]]

100%|██████████| 238/238 [00:00<00:00, 300.17it/s]
Test Set:
Accuracy: 0.9017105263157895
Precision: 0.9027609059432867
Recall: 0.9017105263157895
F1 Score: 0.9016840605747216
Confusion Matrix: [[1695   43   73   89]
 [  25 1835   15   25]
 [  60   31 1600  209]
 [  46   17  114 1723]]
```

Analysis

In the pre training phase, the ELMo model is trained on the given train dataset using next word prediction and previous word prediction tasks. During training, the model learns to generate word embeddings that capture contextual information by considering the surrounding words in the input text. This contextual information allows ELMo embeddings to capture the complexities in word meaning and syntactic structure.

In the downstream task of classification, we got accuracies around 89 and 90 for SVD and Word2Vec(using SkipGram) embeddings. However, using ELMo embeddings we got accuracies around 91 which is more.
ELMo embeddings capture rich contextual information, allowing the classifier to better understand the meaning of words in context. This leads to higher accuracy, precision, recall, and F1 score compared to Word2Vec and SVD embeddings

Hyperparameter Tuning

We observe that we get more accuracy using Trainable λs as compared to Frozen λs and Learnable Function.

The trainable λs technique allows the model to adaptively adjust the importance of each layer's representation based on the specific characteristics of the downstream task. This adaptability enables the model to better capture the complexities of the data, leading to improved performance.
Using frozen λs, we are not able to train the weights of the embeddings which limits the model’s ability to adapt to the downstream task.
Using a learnable function increases the model's complexity and leads to overfitting.