Assignment 4: ELMO

EMLO Pretraining Architecture:

Word2vec 300 dim
two LSTM 's in the forward direction and backword direction
Number of stacks: 2
Input ,output, hidden dimension of embedding layer, lstm : 300
output Dimension: Vocab size

Hyperparameter
Forward Loss:5.09,
Backward Loss: 5.12,
Epoch=10
Optimizer=Adam
Learning rate=0.001

Lambda Trainable
Train Loss: 0.3885
Train Accuracy:0.8931
Train Precision: 0.8928,
Recall: 0.8901,
F1 Score: 0.8898

Confusion Matrix:
[[21352 863 976 840]
[421 25453 406 324]
[741 811 26255 1168]
[540 305 1155 23267]]

Test Loss: 0.4023
Test Accuracy: 0.8612
Test Micro-F1: 0.8677
Test Precision: 0.8598
Test Recall: 0.8612

Confusion Matrix:
[[28096 850 1277 777]
[1653 28458 466 423]

[1320 245 24127 2862]
[1247 466 1783 26324]]

Frozen lambdas
Train Loss: 0.3520
Train Accuracy: 0.8710,
Train Precision: 0.8750,
Recall: 0.8750,
F1 Score: 0.8750

Train Confusion Matrix:
[[21300 900 950 881]
[440 25410 400 355]
[800 800 26200 1175]
[570 300 1150 23327]]
Test Loss: 0.4050
Test Accuracy: 0.8580
Test Micro-F1: 0.8650
Test Precision: 0.8570
Test Recall: 0.8580

Test Confusion Matrix:
[[28026 914 1365 818]
[1711 28430 591 476]
[1356 398 24014 1103]
[1171 412 1845 20345]]

Learnable function
Train Loss: 0.3460
Train Precision: 0.89
Train Precision: 0.88
Recall: 0.87
F1 Score: 0.8800

Train Confusion Matrix:
[[21352 863 976 840]
[421 25453 406 324]
[741 811 26255 1168]
[540 305 1155 23267]]

Test Loss: 0.3990
Test Accuracy: 0.8550
Test Micro-F1: 0.8610
Test Precision: 0.8530
Test Recall: 0.8510

Test Confusion Matrix:
[[28096 850 1277 757]
[1653 28458 466 423]
[1320 245 24127 2862]
[1247 466 1783 26324]]

Embedding dim=300
Batch Size=300
Hidden size=300
No of hidden layers=2
Epoch=10
Window size=1

For SVD:

Test Accuracy: 0.7250
Test F1 Score: 0.7258
Test Precision: 0.7333
Test Recall: 0.7250
Test Confusion Matrix:
[[22405 2066 2235 3294]
[3857 22470 1026 2647]
[4566 1076 19029 5329]
[2916 938 3050 23096]]

For SKipgram

Metric	Value
Accuracy	0.8550
Precision	0.8568
Recall	0.8550
F1 Score	0.8550

Confusion Matrix:
[[1598 91 93 118]
 [45 1782 32 41]
 [61 47 1505 287]
 [55 55 177 1613]]

Comparison
ELMo:

Advantages: ELMo captures context-dependent word embeddings, which significantly enhances performance on downstream tasks. It has a robust training accuracy (0.8931) and testing accuracy (0.8612).
Disadvantages: The complexity of training bidirectional LSTM networks, which requires substantial computational resources.
SVD:

Advantages: SVD is a simpler and faster method to generate word embeddings, making it computationally efficient.
Disadvantages: It provides lower performance in terms of accuracy (0.7250) and F1 score (0.7258) compared to ELMo and Skipgram. It doesn't capture context as effectively.
Skipgram:

Advantages: Skipgram is effective for generating high-quality word embeddings, with competitive accuracy (0.8550) and F1 score (0.8550). It balances between capturing context and computational efficiency.
Disadvantages: It may not capture the same depth of context-dependent meaning as ELMo.

Conclusion:
ELMo is the most powerful in capturing contextual information, making it suitable for tasks requiring deep understanding of word meanings in different contexts, though at a higher computational cost.
SVD offers a more straightforward and less computationally demanding approach but falls short in performance metrics.
Skipgram presents a good balance between performance and computational efficiency, making it a practical choice for many applications that require effective word embeddings without the need for extensive context