NLP HW4 Assignment:
Name: Payal Rashinkar
USC ID: 3885-1419-03

Task1:

Description:
In the initial phase of Task 1, I experimented with normalizing the input data by converting all characters to lowercase. This modification did not significantly alter the outcomes. Consequently, for Task 2, the decision was made to incorporate uppercase information, which proved advantageous. The model's performance—reflected in its accuracy, precision, recall, and F1 score—was optimized by adjusting various hyperparameters. These included the batch size during training, the learning rate, weight decay, and momentum in the SGD optimizer. Additionally, I employed techniques such as early stopping and learning rate schedulers, each of which required fine-tuning. The patience parameter of the scheduler, in particular, was carefully calibrated to enhance the model's results.

Eval output:

F1-Score: 80.20%

```
bash-3.2$ python3 eval.py -p dev1.out -g data/dev
processed 51577 tokens with 5942 phrases; found: 5380 phrases;
correct: 4540.
accuracy:  95.91%; precision:  84.39%; recall:  76.41%; FB1:  80.20
             LOC: precision:  88.89%; recall:  87.10%; FB1:  87.98
1800
            MISC: precision:  86.91%; recall:  77.77%; FB1:  82.08
825
             ORG: precision:  78.95%; recall:  68.23%; FB1:  73.20
1159
             PER: precision:  81.95%; recall:  71.01%; FB1:  76.09
1596
```

Task2:

Description:
To enhance the Bidirectional LSTM (BLSTM) model from Task 1, I integrated GloVe word embeddings, involving the initialization of our neural network's embeddings with GloVe's pre-trained vectors. However, given GloVe's case insensitivity and the significance of capitalization in Named Entity Recognition (NER), we introduced a binary feature to capture the capitalization status of each word's initial letter. This adjustment, alongside hyperparameter tuning similar to Task 1, significantly improved the model's precision,

recall, and F1 score, underscoring the importance of capitalization in accurately identifying named entities.

Eval output:
F1–Score: 89.35%

```
bash-3.2$ python3 eval.py -p dev2.out -g data/dev
processed 51577 tokens with 5942 phrases; found: 6127 phrases;
correct: 5392.
accuracy:  98.30%; precision:  88.00%; recall:  90.74%; FB1:  89.35
             LOC: precision:  92.00%; recall:  94.50%; FB1:  93.23
1887
            MISC: precision:  80.43%; recall:  85.14%; FB1:  82.72
976
             ORG: precision:  79.41%; recall:  83.97%; FB1:  81.62
1418
             PER: precision:  94.53%; recall:  94.73%; FB1:  94.63
1846
```