

Students: Nadav Spitzer, 302228275. Lior Frenkel, 204728315

Part 3

POS task –

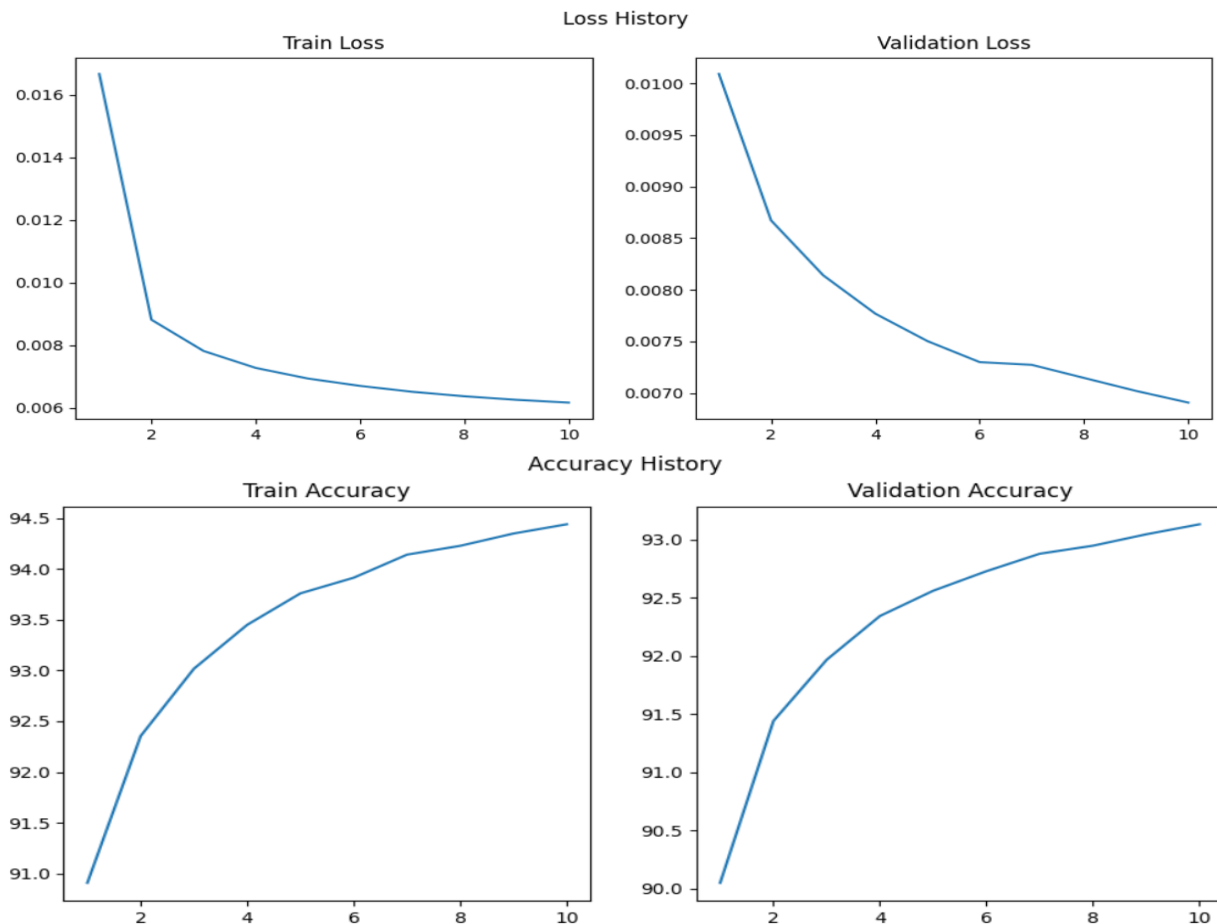
1. Parameters:

- a. Number of epochs: 10
- b. Learning rate: $1e-4$
- c. Batch size: 32
- d. Hidden layer size: 150
- e. Optimizer: Adam
- f. Dropout with 0.5 probability
- g. Weight decay: $1e-4$

2. Results:

- a. Train loss: 0.00616
- b. Train accuracy: 94.439%
- c. Validation loss: 0.007
- d. Validation accuracy: 93.132%

3. Graphs:



NER task –

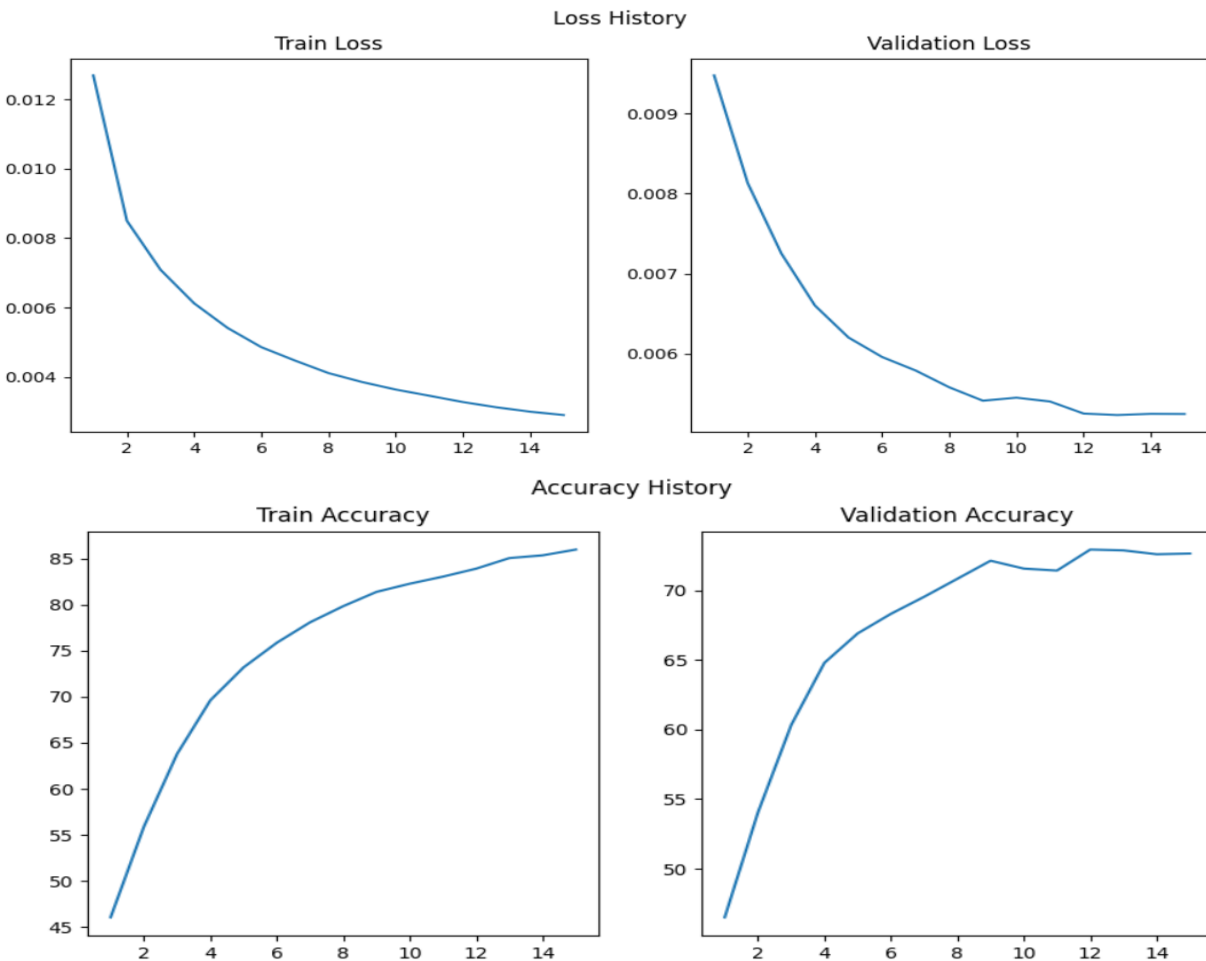
4. Parameters:

- a. Number of epochs: 15
- b. Learning rate: $1e-4$
- c. Train Batch size: 32
- d. Dev Batch size: 32
- e. Hidden layer size: 150
- f. Optimizer: Adam
- g. Dropout with 0.5 probability
- h. Weight decay: $1e-4$

5. Results:

- a. Train loss: 0.003
- b. Train accuracy: 85.982%
- c. Validation loss: 0.00525
- d. Validation accuracy: 72.636%

6. Graphs:



Considerations: Each sentence in the data (sequence of words between blank rows) were padded with special words for start and end ('<s>', '</s>') in order to achieve window of size 5 when the required word is in the middle.

Words that were seen in the DEV or TEST files but not in the TRAIN were given a special word of 'UUUNKKK' and a tag of UNSEEN.

Results:

The scores didn't improve but even got worse by 1%-3%. This could be caused by the fact that the weights are already trained for a specific task that could be quite different from ours.