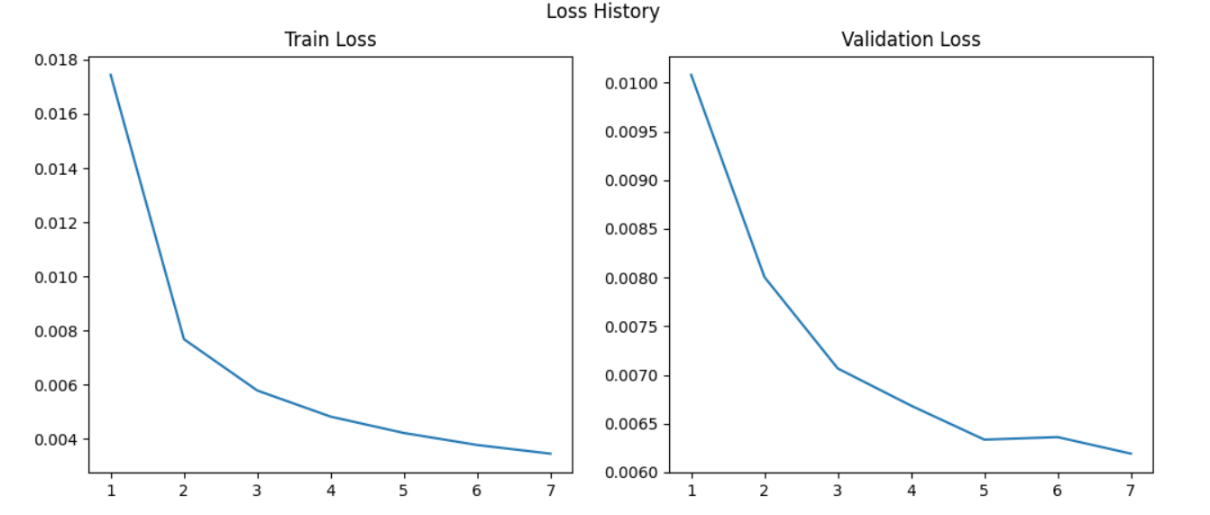
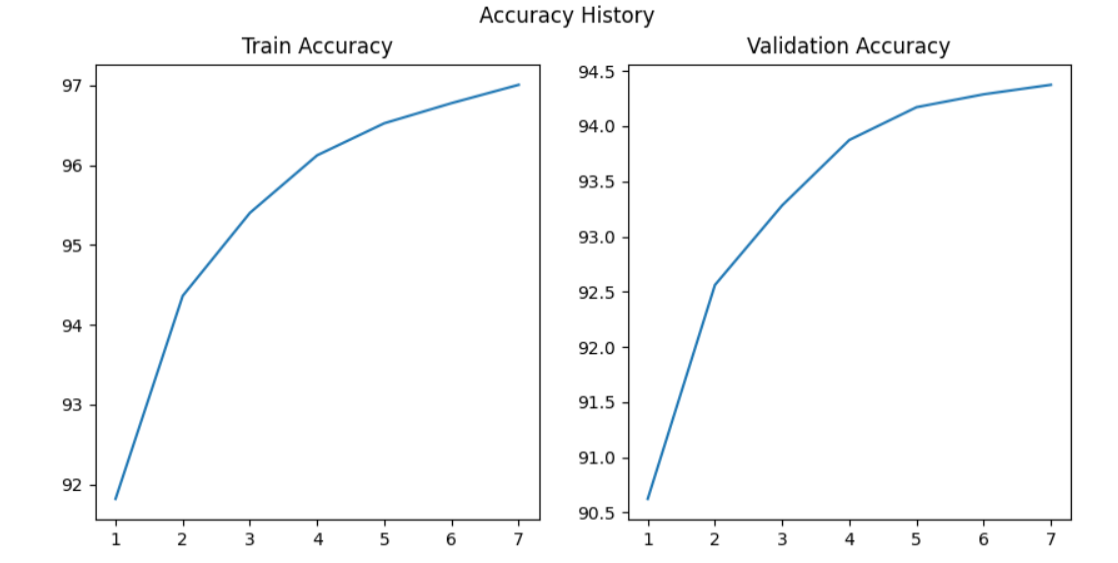
**Students:** Nadav Spitzer, 302228275. Lior Frenkel, 204728315

**Part 1**

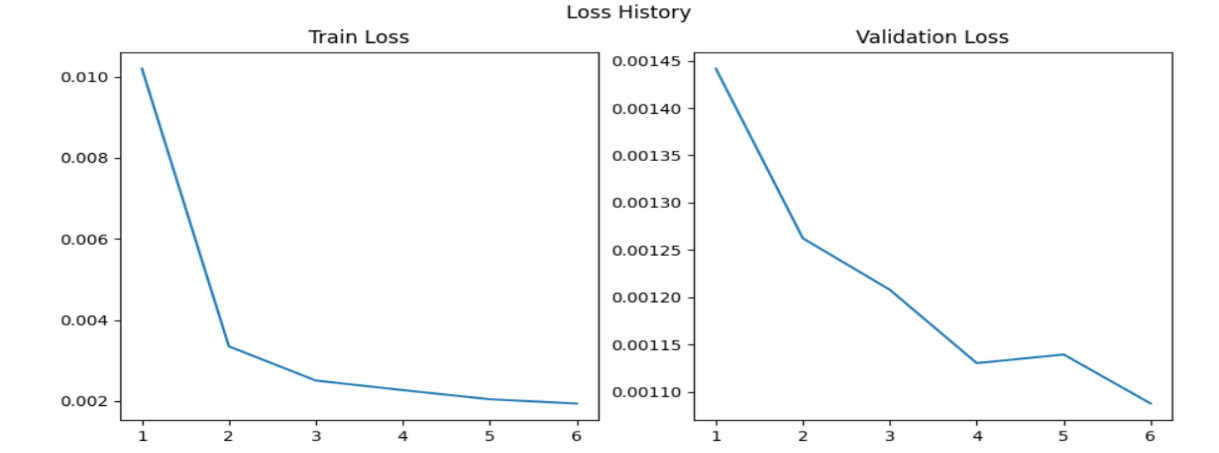
**POS task –**

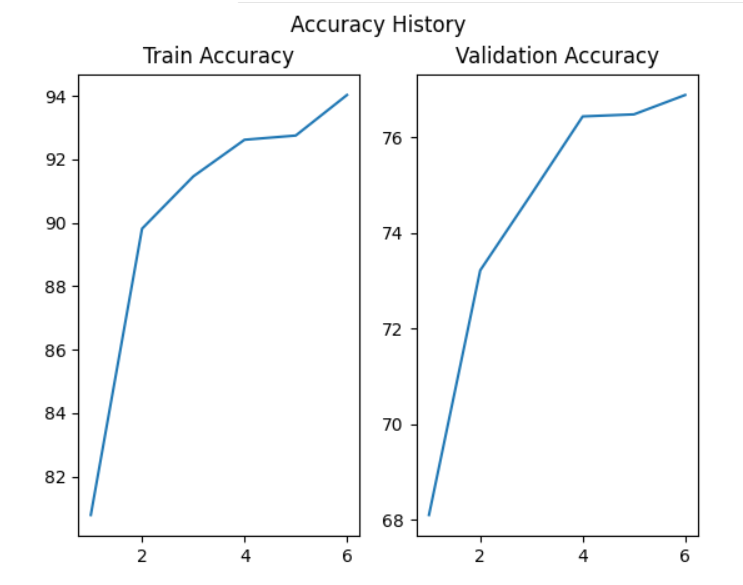
1. Parameters:
   1. Number of epochs: 7
   2. Learning rate: 1e-3
   3. Batch size: 32
   4. Hidden layer size: 150
   5. Optimizer: Adam
   6. Dropout with 0.5 probability
2. Results:
   1. Train loss: 0.00346
   2. Train accuracy: 97.007%
   3. Validation loss: 0.00619
3. Validation accuracy: 94.375%
4. Graphs:





**NER task –**

1. Parameters:
   1. Number of epochs: 6
   2. Learning rate: 1e-3
   3. Train Batch size: 32
   4. Dev Batch size: 128
   5. Hidden layer size: 100
   6. Optimizer: Adam
   7. Dropout with 0.5 probability
   8. Weight decay: 1e-4
2. Results:
   1. Train loss: 0.001938
   2. Train accuracy: 94.024%
   3. Validation loss: 0.00108
   4. Validation accuracy: 76.884%
3. Graphs:



**Considerations:** Each sentence in the data (sequence of words between blank rows) were padded with special words for start and end (‘<S>’, ‘<E>’) 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 ‘<U>’ and a tag of UNSEEN.