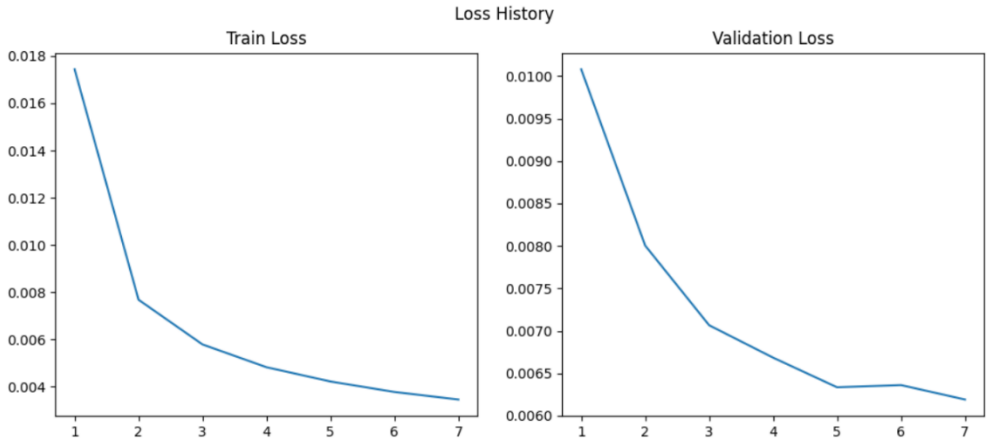
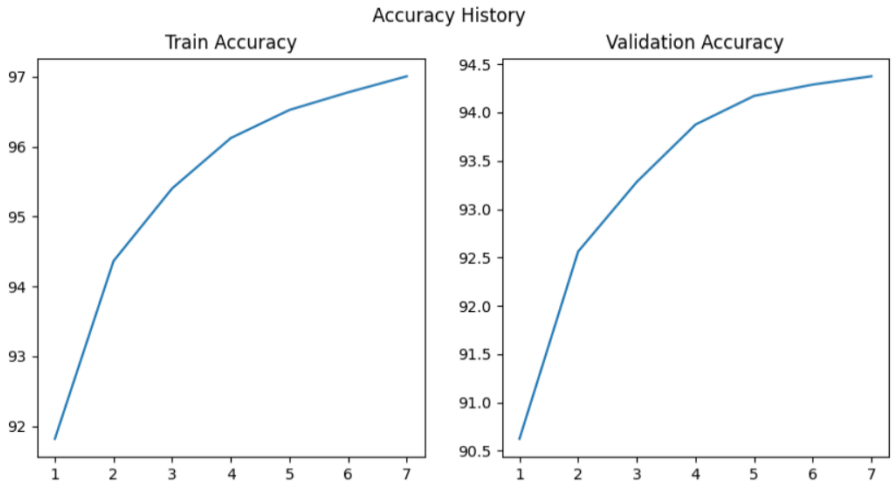
Students: Nadav Spitzer, 302228275. Lior Frenkel, 204728315

Part 1

POS task –

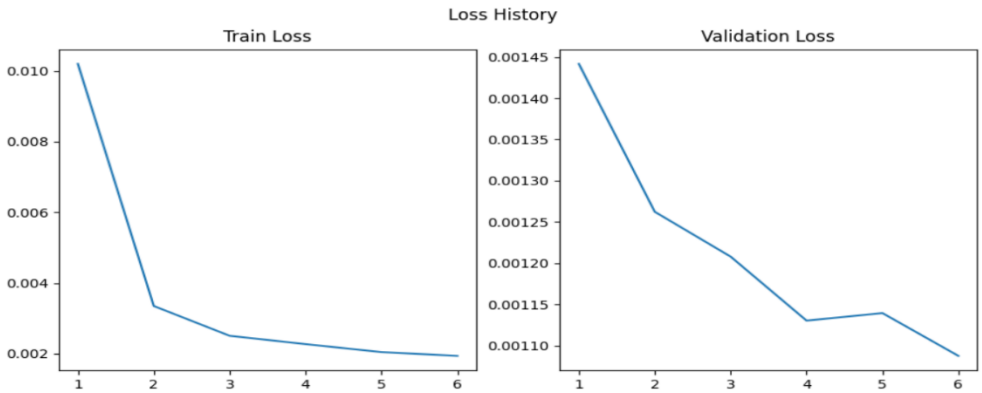
1. Parameters:
    a. Number of epochs: 7
    b. Learning rate: 1e-3
    c. Batch size: 32
    d. Hidden layer size: 150
    e. Optimizer: Adam
    f. Dropout with 0.5 probability
2. Results:
    a. Train loss: 0.00346
    b. Train accuracy: 97.007%
    c. Validation loss: 0.00619
3. Validation accuracy: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

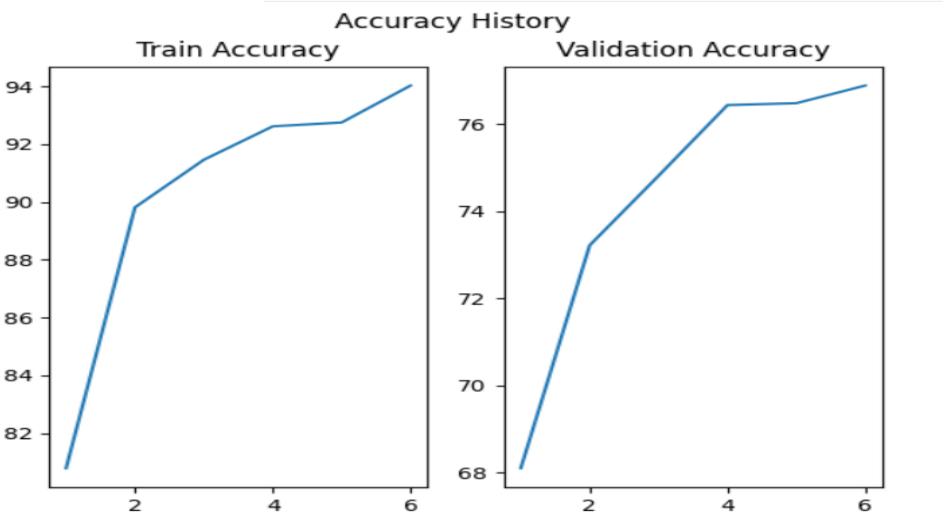




NER task –

1. Parameters:
    a. Number of epochs: 6
    b. Learning rate: 1e-3
    c. Train Batch size: 32
    d. Dev Batch size: 128
    e. Hidden layer size: 100
    f. Optimizer: Adam
    g. Dropout with 0.5 probability
    h. Weight decay: 1e-4
2. Results:
    a. Train loss: 0.001938
    b. Train accuracy: 94.024%
    c. Validation loss: 0.00108
    d. 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.