Assignment 2 – Window-based Tagging

Daniel Bazar 314708181

Peleg shefi 316523638

Part 1

Architecture

- Both tasks implement the same Network: an MLP with
  - one hidden layer
  - tanh activation
- The network trained with a cross-entropy loss.
- We built a grid search function in order to find the best parameters combination. Then we kept the parameters which gave the best results for the DEV accuracy.
- Initizlizing the weights using Xavier.
- The optimizer is Adam.

Best parameters

- NER:
  - Hidden layer size: 130
  - Dropout probability: 0.3
  - Batch size: 128
  - Optimizer: Adam (Learning rate: 1e-4)
  - Epochs: 6
- POS:
  - Hidden layer size: 90
  - Dropout probability: 0.2
  - Batch size: 64
  - Optimizer: Adam (Learning rate: 5e-05)
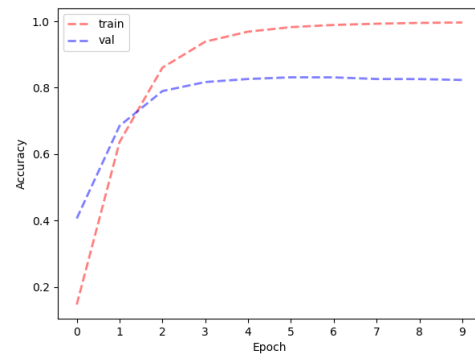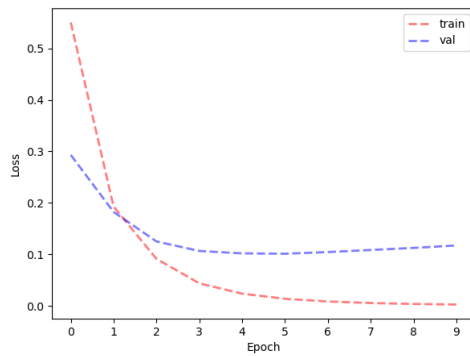 - Epochs: 8

Considerations

- We handle unseen words (words that are not in the vocab in this part it means not in the train file) by giving them an UNK tag. The UNK tag gets an Embedding vector as well.
- We padded the sentences with SOS (start of string) and EOS (end of string) at the beginning and end of the sentence. We tried using SOS2, SOS1, EOS1, EOS2 in order to capture the difference between them. It didn't improve our results, so weren't using it.
- As to the optimizer we tried SGD but got slower convergence. So, we tried doing an AdaBound optimizer which at the beginning acts like Adam and in later epochs more like SDG for accuracy. Yet it didn't yield any fruit, so we stayed with Adam.
- The initialization of Xavier is key! It improved our model significantly from around 78% on NER to the results written below. We experimented with other initialization which didn't perform well enough.

Results

- NER:
    - Loss validation: 0.101
    - Accuracy validation: 83.13%
- POS:
    - Loss validation: 0.135
    - Accuracy: 95.84%

Graphs

NER

POS