Assignment 1 – Part 1

Submitters:

Ori Braverman 318917010

Elie nedjar 336140116

Model Hyper-parameters

- Optimaizer: Adam
- Learning rate: 0.001
- Size of hidden layer: 32

In the pos task:

- The batch size of the training was 32 for the training set and the dev set
- The number of epochs was 25 with early stopping.

In the ner task:

- The batch size of the training was 32 for the training set and the 128 for the dev set
- The number of epochs was 25 with early stopping.

Answers to the consideration part:

1. what do you do with a word that appears in the train set and not in the dev set? which word embeddings will you assign to it?

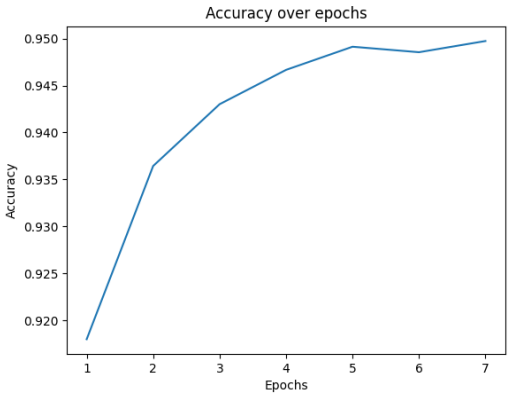
- Our strategy to handle out-of-vocabulary words in the dev data is by the '<UNK>' token. For any word not in the vocabulary (e.g. from the dev set but not in the train set), We assign it the embedding vector of '<UNK>'. To train this embedding vector, We converted all words with one occurrences in the training set as unknown words, assigning them the '<UNK>' embedding vector.

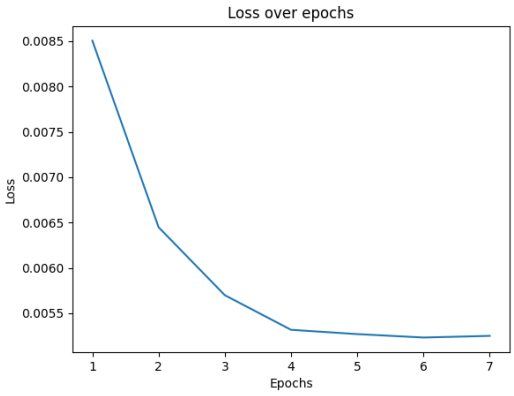
2. What vectors will you use for the words surrounding the first word in the sequence? And for the last word?

- For the words surrounding the first word in a sequence, we will use a special token '<PAD_START>' to represent the context to the left. For the last word, we will use '<PAD_END>' to represent the context to the right. These tokens will have embedding vectors that are learned during training.

Model Results

Pos: - Dev Accuracy: 0.9497 - Dev Loss: 0.0053





Ner: - Dev Accuracy: 0.7614 - Dev Loss: 0.0012