Assignment 1 – Part 3

Submitters:

Ori Braverman 318917010

Elie nedjar 336140116

Model Hyper-parameters

- Optimaizer: Adam
- Learning rate: 0.001
- Size of hidden layer: 128

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

How we handled the pre-trained vectors

- The pre-trained embedding vocabulary is in lower-case. So, when making a match between the words from the train/dev/test to the vocabulary, we first checked for the original word in case it is in the vocabulary and the lower case.
  This approach helps when dealing with words that have capital letters (e.g. start of a sequence).
- Words recognized as numbers by appropriate regex patterns are assigned tokens from the pre-trained vocabulary that represent numbers, such as 'DGDG', 'DG.DGDG', and 'NNNUMMM'.
- Words in the training, dev, or test sets that do not appear in the pre-trained vocabulary are assigned the unknown token 'UUUNKKK'. This token is part of the pre-trained vocabulary and represents out-of-vocabulary words.

Model Performance

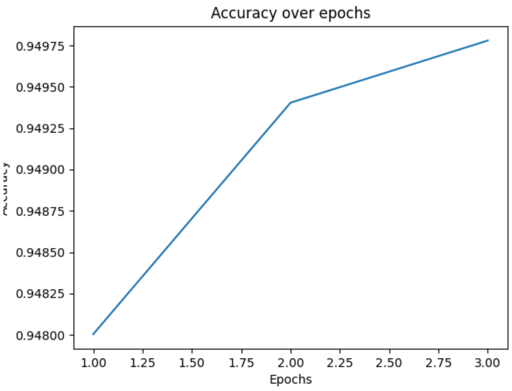
In compared to the previous tagger the results were slightly better.

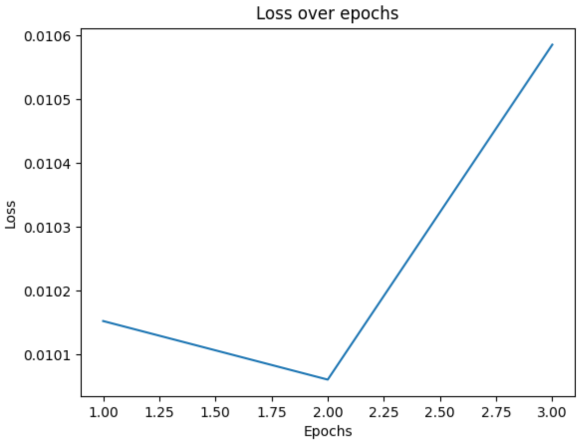
This similar performance could be attributed to the relatively small overlap between words in the pre-trained vocabulary and the training/development sets.

- If many words in the training and development sets are not present in the pre-trained vocabulary, the model will have to rely on a generic embedding.
- Context mismatch – if the pre-trained embeddings capture word meanings based on the contexts they were exposed to during the pre-training.

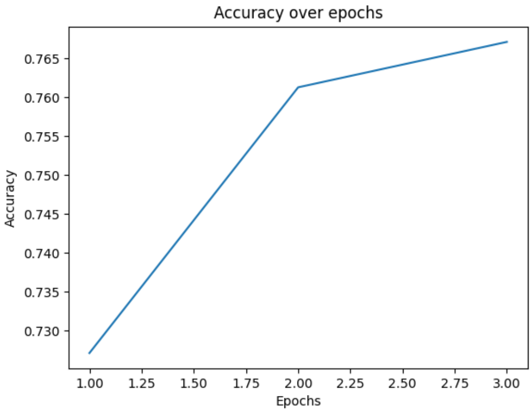
Model Results

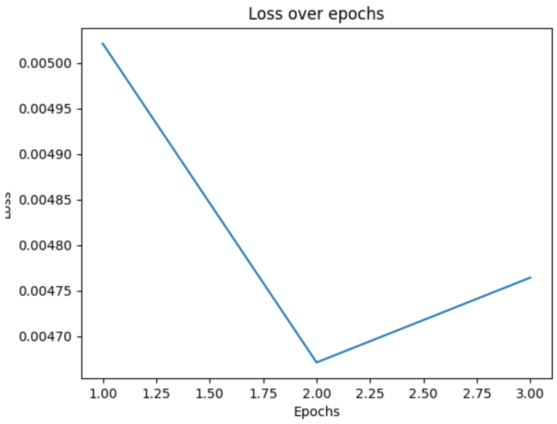
POS: - Dev Accuracy: 0.9498 - Dev Loss: 0.0106





NER: - Dev Accuracy: 0.7671 - Dev Loss: 0.0048


Accuracy over epochs
Accuracy
Epochs


Loss over epochs
Loss
Epochs