

Oren Cohen
305164295
DL ass2 part1

I trained 2 models.

for both model I used same parameters:

number of epochs = 6

Learning rate= 0.05

cross-entropy loss (as required)

additionally I add ADAM optimizer (worked better then SGD)

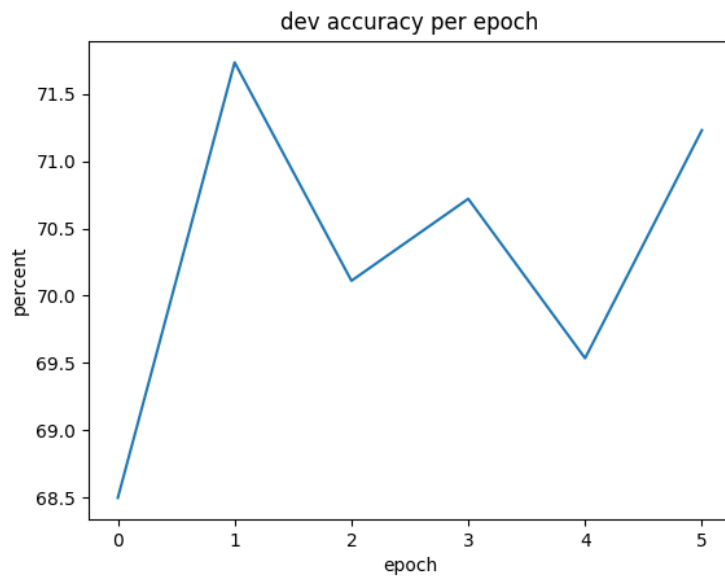
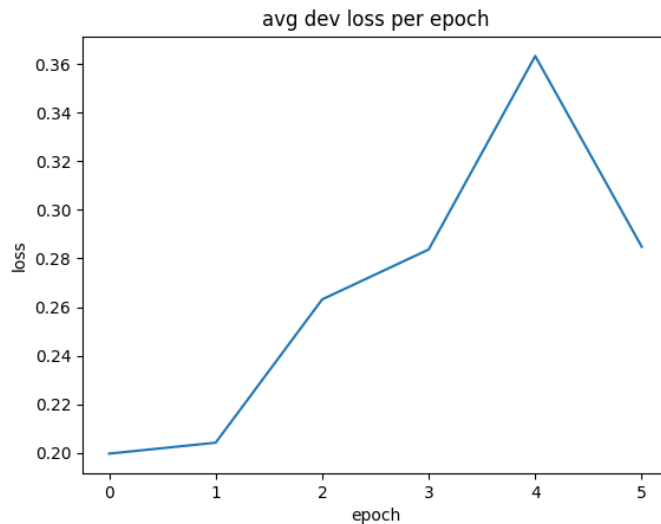
(in the training I also used batching, with size of 1024)

NER model:

after 6 epochs:

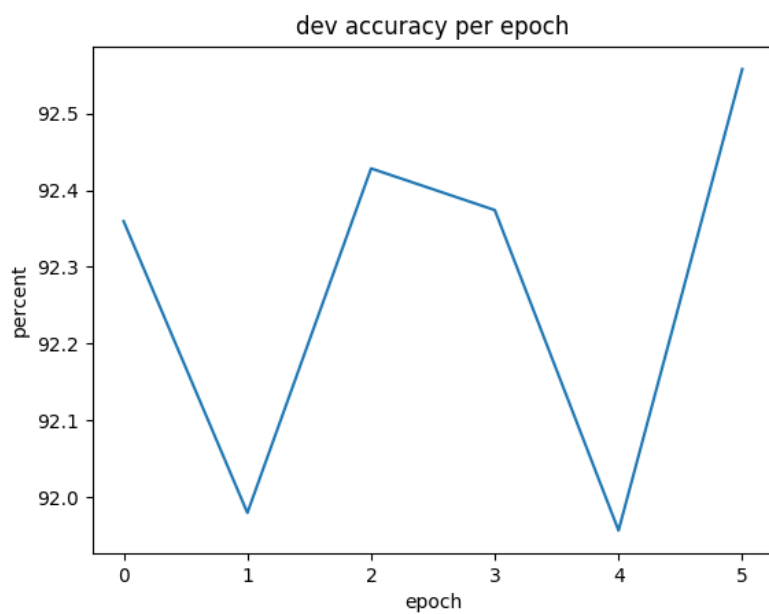
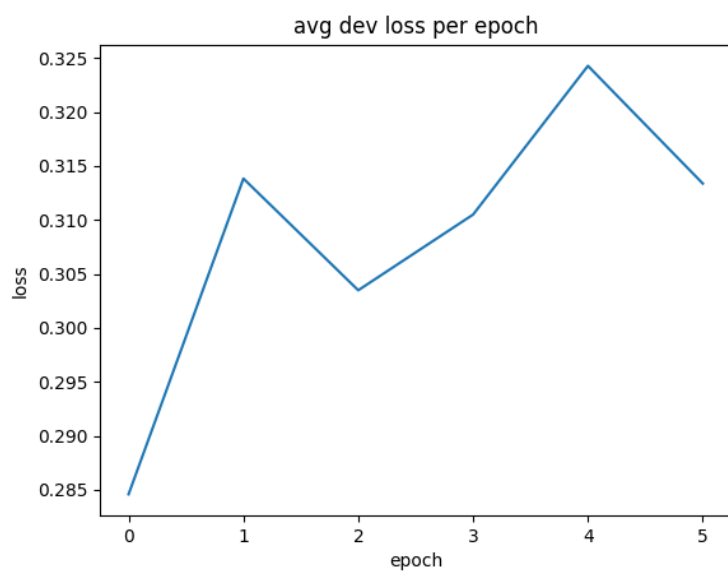
average loss: ~0.28

accuracy: ~71%



Oren Cohen
305164295
DL ass2 part1

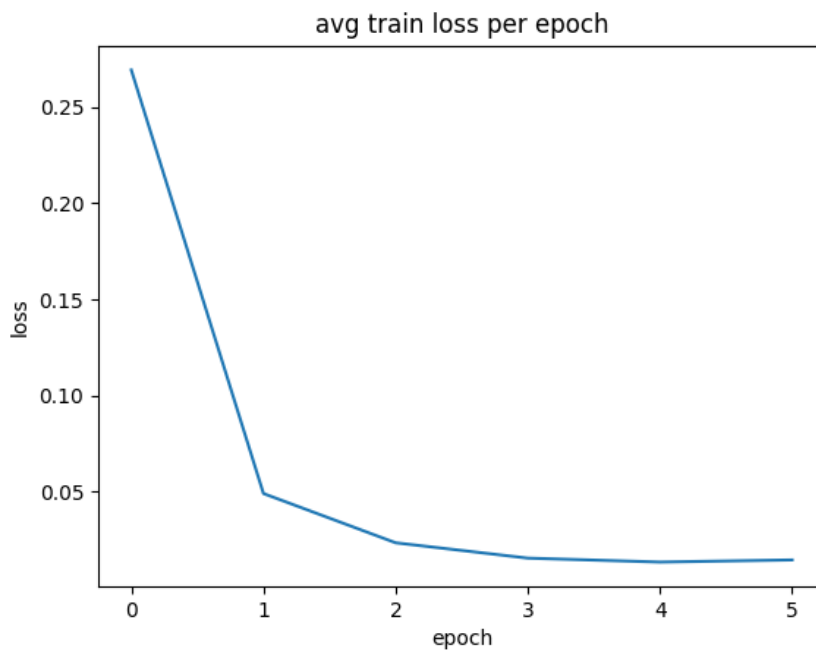
POS model:
after 6 epochs:
average loss: ~ 0.315
accuracy: $\sim 92.5\%$



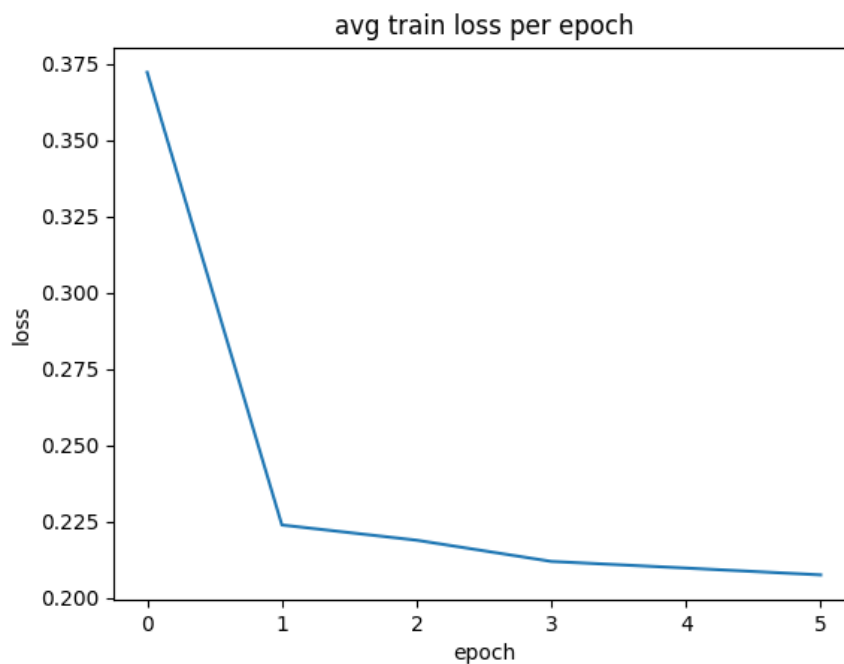
Oren Cohen
305164295
DL ass2 part1

**It's seems to me that the loss in the dev section is wrong (because its increase in time), so, I checked the loss in the training process – and it was FINE!
After all we are training to minimize the train loss (and it is defiantly worked)
I still don't know why the dev loss is increasing.**

(I add the avg train loss of both models):
NER:



POS:



Oren Cohen
305164295
DL ass2 part1

conclusion:

the POS model was much more accurate!

Considerations:

what do you do with a word that appears in the dev set and not in the train set?

Ans:

I added to the train set spacial word: "UNK", and for any word that not fount in the word set (in dev and test) I'm using this word.

which word embeddings will you assign to it?

"UNK"

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

For every sentence I used padding of START START W1 W2 ...Wn END END
its geranty that every word in a sentence will have window of 5 words,
and also solve the problem of the first and last word..