Yehoshua Stern 314963927

Eyal Cohen 207947086

**Part 1 (tagger 1):**

**Parameters:** (POS)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Learning rate | Epochs | Batch size | Batches per epoch | Middle dimension |
| 0.1 divided by 2 every 7 epochs | 15 | 256 | 4096 | (DIM \* 5 + out\_dim) / 2  = 148 |

**Parameters:** (NER)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Learning rate | Epochs | Batch size | Batches per epoch | Middle dimension |
| 0.1 divided by 2 every 7 epochs | 35 | 256 | 4096 | (DIM \* 5 + out\_dim) / 2  = 138 |

The training process was with mini batches, every batch was taken randomly from the training data, we did not make a full pass for every epoch, we just took every epoch 4096 random batches of 256 samples.

**Considerations:**

Q: What do you do with a word that appears in the train set and not in the dev set

A: We just called it UNK and we assigned the vector of the UNK special word.

Q: What vectors will you use for the words surrounding the first and last words in the sequence?

A: we had special words START and END that had their own vector in the embedding layer.

**Graphs: (x is epochs, y is value)**