ex2 - part 1

Eitan kerzhner   205697139
itamar Shachen tov   207497769

Model's Params:

POS:
o Dropout: p=0.5 (default), applied after Tanh activation.
o Optimizer: Adam.
o Learning rate: 1e-3 (Adam's default).
o Hidden layer size: 128 units.
o Word embedding size: 50 dimensions.
o Context window: 5 words.
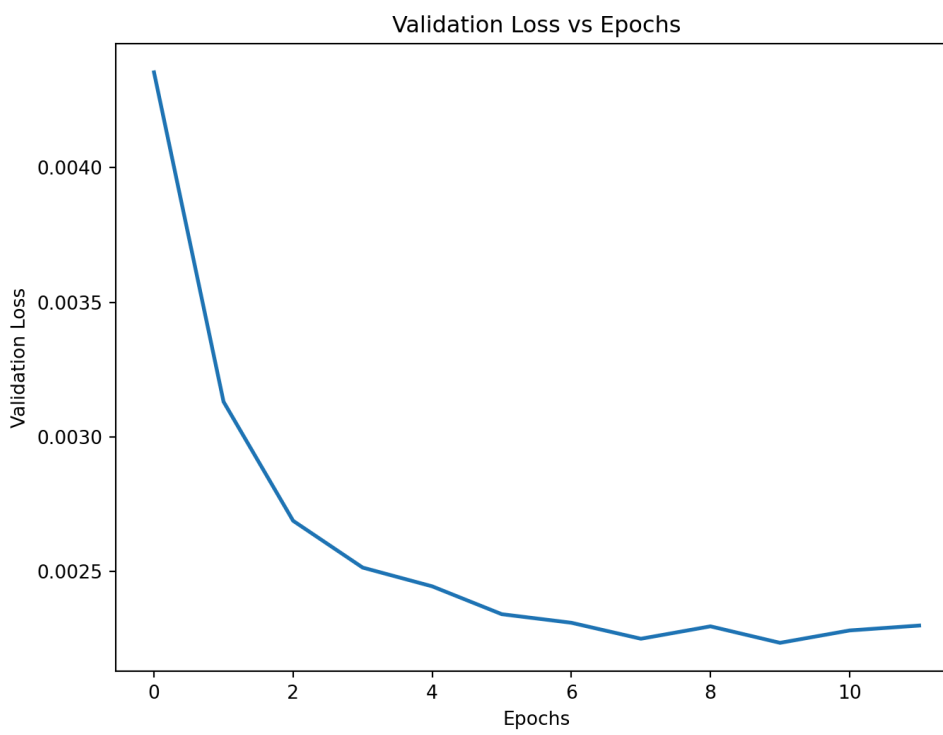o Batch size (train & dev): 64.
o Number of epochs: 12.

NER:
o Dropout: p=0.5 (default), applied after Tanh activation.
o Optimizer: Adam.
o Learning rate: 1e-3 (Adam's default).
o Hidden layer size: 128 units.
o Word embedding size: 50 dimensions.
o Context window: 5 words.
o Batch size (train & dev): 64.
o Number of epochs: 12.

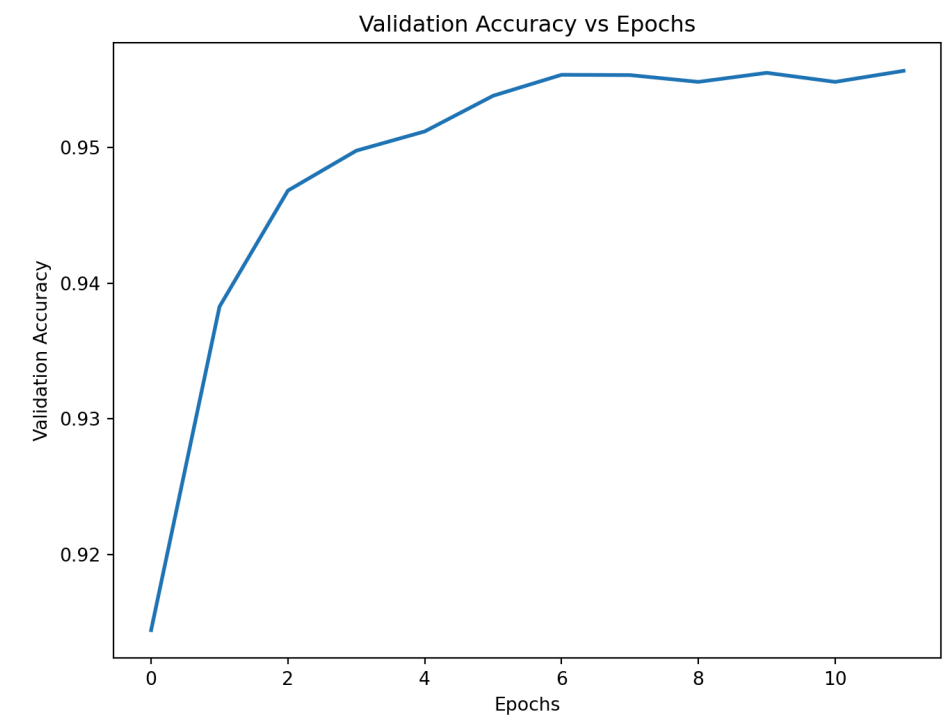Considerations:

o Unseen/rare words (frequency <=1 in train) are mapped to an '`<unknown>`' token. This token has its own embedding, trained by treating these rare training words as '`<unknown>`'.

o To handle context for initial words, two '`<s>`' (start-of-sentence) tokens are prepended to each sentence. These tokens have learnable embeddings and provide left context.

o Similarly, two '`<e>`' (end-of-sentence) tokens are appended for the right context of final words, also with their own learnable embeddings.

POS:

Validation Accuracy vs Epochs

Validation Loss vs Epochs

NER:

Validation Accuracy vs Epochs

Validation Loss vs Epochs