DL4Seq - Assignment 3

Otmazgin, Shon 305394975

Rassin, Royi 311334734

May 15th, 2021

1 Report 1

Q: How large were the training and test sets,

A: Train: 1000 examples. 500 positive, 500 negative. Test: 100 examples. 50 positive, 50 negative.

Q: did your network succeed in distinguishing the two languages (it should)?

A: Yes. An LSTM network can learn a sequence of "something" followed by "something" else. In our case, b-sequence follows c-sequence.

Q: how long did it take(both wall-clock time (i.e., number of seconds), and number of iterations)?

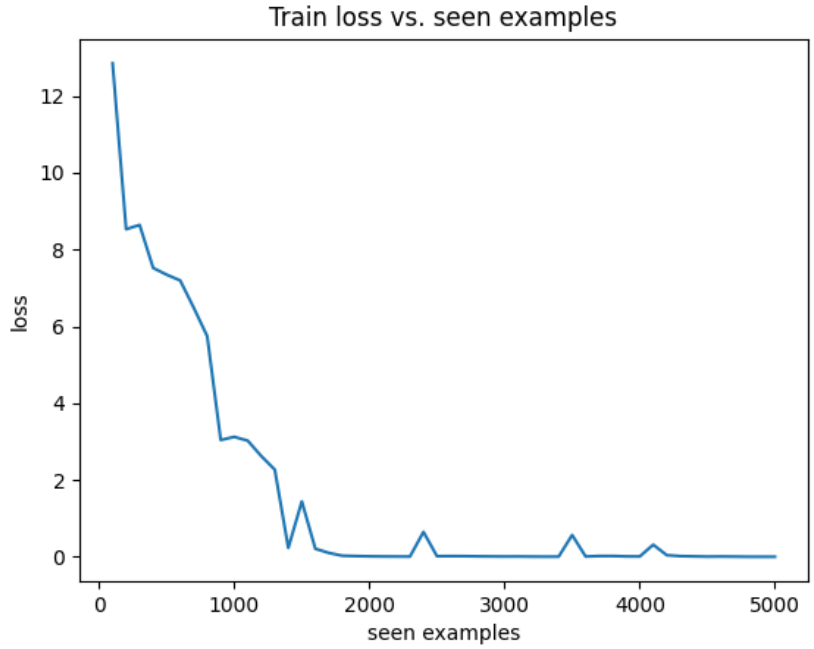
A: We implemented batching in training. In particular, we used a batch size of 10 and padded each example to the max word length for a total of five epochs. Therefore, we had 100 iterations per epoch, and a total of 5 epochs ($\frac{1000}{10} * 5 = 500$). The model learned to distinguish between the two languages after 150 iterations. The wall clock after 150 iterations was $\sim$ 39 seconds on GPU.

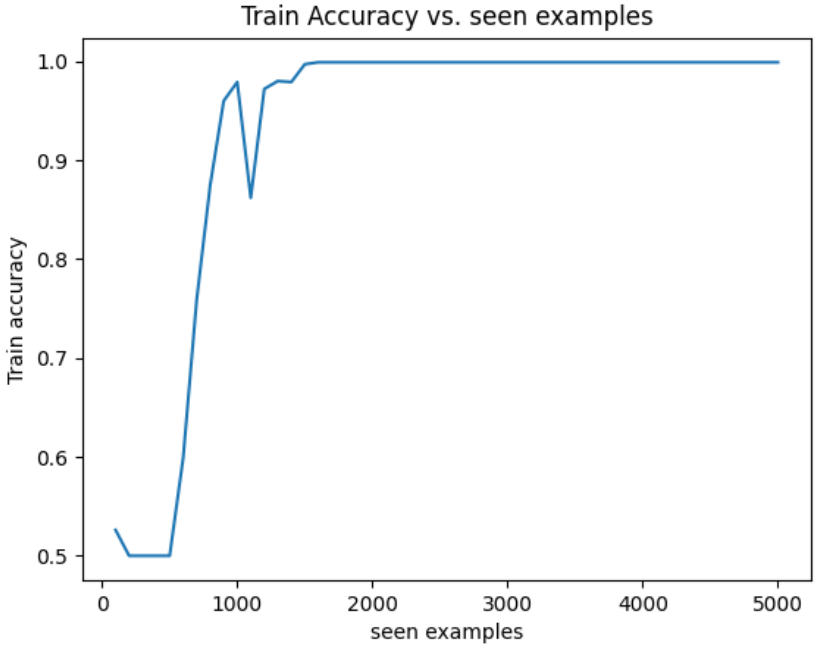
Q: did it succeed only on the train and not on the test?

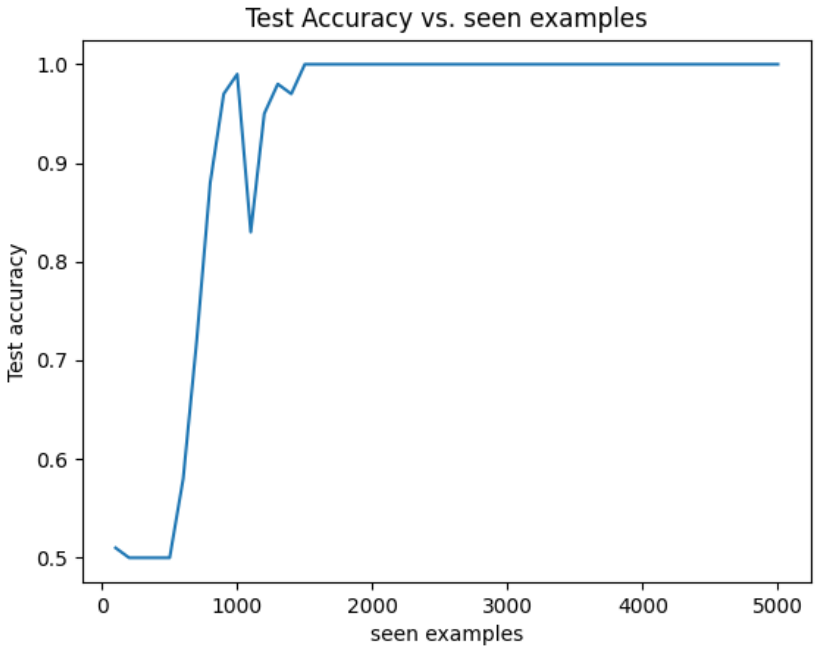
A: It succeeded on both. After every iteration (as mentioned above, each iteration is 100 seen examples) we measured the model accuracy both on train and test sets. **See graphs below for full details.**

Q: what you did in order to make it work, etc.

A: We implemented 1 layer LSTM(hidden dim = 1024) network followed by MLP with 1 hidden layer(dim=1024) network with dropout(p=0.3) and batch norm. For char embedding we used dim of 300. In addition to Adam optimizer with 0.001 learning rate. Lastly, as mentioned earlier, we trained for five epochs (although after 1.5 epochs the model solved the problem) and the batch-size was 10.


Train loss vs. seen examples
loss
seen examples
0
2
4
6
8
10
12
0
1000
2000
3000
4000
5000


Train Accuracy vs. seen examples
Train accuracy
seen examples
0.5
0.6
0.7
0.8
0.9
1.0
0
1000
2000
3000
4000
5000


Test Accuracy vs. seen examples
Test accuracy
seen examples
1.0
0.9
0.8
0.7
0.6
0.5
0
1000
2000
3000
4000
5000