Assignment 2:
Intro to Natural Language Processing

1 Programming: Text Classification with RNNs

1.1 Programming: Text Classification with RNNs

Hyperparameters:
Choice of nonlinearity = tahn
Word embedding dimension size = 16
Hidden Dimension size = 64
Dropout rate = 0.5
Choice of Optimization method = adam
Learning rate = 10^{-3}
Training Batch Size = 32
Number of Training Epochs = 20

- *Solution.*

TRAINING DATA RESULTS

	Loss	Accuracy
Epoch 1	0.6691	0.5796
Epoch 2	0.5723	0.7025
Epoch 3	0.3721	0.8251
Epoch 4	0.2221	0.9069
Epoch 5	0.1271	0.9519
Epoch 6	0.0783	0.9711
Epoch 7	0.0608	0.9793
Epoch 8	0.0652	0.9770
Epoch 9	0.0712	0.9745
Epoch 10	0.0388	0.9866
Epoch 11	0.0360	0.9877
Epoch 12	0.0301	0.9896
Epoch 13	0.0376	0.9868
Epoch 14	0.0345	0.9878
Epoch 15	0.0328	0.9884
Epoch 16	0.0251	0.9916
Epoch 17	0.0288	0.9908
Epoch 18	0.0380	0.9867
Epoch 19	0.0408	0.9861
Epoch 20	0.0271	0.9911

TEST DATA RESULTS

	Loss	Accuracy
Epoch 1	0.6987	0.5084
Epoch 2	0.5178	0.7405
Epoch 3	0.2342	0.9047
Epoch 4	0.1398	0.9449
Epoch 5	0.1040	0.9608
Epoch 6	0.0785	0.9708
Epoch 7	0.0674	0.9753
Epoch 8	0.0570	0.9796
Epoch 9	0.0513	0.9812
Epoch 10	0.0484	0.9828
Epoch 11	0.0456	0.9837
Epoch 12	0.0460	0.9836
Epoch 13	0.0450	0.9844
Epoch 14	0.0362	0.9869
Epoch 15	0.0412	0.9854
Epoch 16	0.0396	0.9863
Epoch 17	0.0315	0.9896
Epoch 18	0.0311	0.9895
Epoch 19	0.0276	0.9914
Epoch 20	0.0214	0.9928

1.2 Programming: Text Classification with LSTMs

Hyperparameters:

- Choice of nonlinearity = tahn
- Word embedding dimension size = 16
- Hidden Dimension size = 64
- Dropout rate = 0.5
- Choice of Optimization method = adam
- Learning rate = 10^{-3}
- Training Batch Size = 32
- Number of Training Epochs = 20

Solution.

TRAINING DATA RESULTS

	Loss	Accuracy
Epoch 1	0.5194	0.7403
Epoch 2	0.3449	0.8556
Epoch 3	0.1965	0.9275
Epoch 4	0.1498	0.9477
Epoch 5	0.1189	0.9596
Epoch 6	0.1014	0.9659
Epoch 7	0.0777	0.9737
Epoch 8	0.0648	0.9788
Epoch 9	0.0616	0.9799
Epoch 10	0.0504	0.9835
Epoch 11	0.0467	0.9849
Epoch 12	0.0385	0.9877
Epoch 13	0.0335	0.9905
Epoch 14	0.0310	0.9901
Epoch 15	0.0248	0.9915
Epoch 16	0.0187	0.9938
Epoch 17	0.0176	0.9949
Epoch 18	0.0168	0.9945
Epoch 19	0.0276	0.9948
Epoch 20	0.0156	0.9951

TEST DATA RESULTS

	Loss	Accuracy
Epoch 1	0.5132	0.7415
Epoch 2	0.3172	0.8677
Epoch 3	0.1820	0.9335
Epoch 4	0.1386	0.9503
Epoch 5	0.1144	0.9595
Epoch 6	0.0790	0.9742
Epoch 7	0.0654	0.9775
Epoch 8	0.0545	0.9818
Epoch 9	0.0548	0.9825
Epoch 10	0.0410	0.9873
Epoch 11	0.0315	0.9901
Epoch 12	0.0322	0.9901
Epoch 13	0.0253	0.9925
Epoch 14	0.0208	0.9928
Epoch 15	0.0217	0.9935
Epoch 16	0.0177	0.9940
Epoch 17	0.0159	0.9950
Epoch 18	0.0163	0.9949
Epoch 19	0.0140	0.9957
Epoch 20	0.0101	0.9970

LSTM has much better accuracy than SimpleRNN. This can be seen most clearly by comparing Epoch 1 of both. In SimpleRNN, Epoch 1 reports an accuracy of 0.5084, while when using LSTM, Epoch 1 reports an accuracy of 0.7403, a difference of 0.2319

2 Theory: Deriving the Viterbi Algorithm

2.1

Solution.

While j = 1,

$$v_j(y_j) = max_{y_{j-1}} s(x, j, y_{j-1}, y_j)$$

Assume that for all j,

$$v_j(y_j) = max_{y_{j-1}} [s(x, j, y_{j-1}, y_j) + v_{j-1}(y_{j-1})]$$

We want prove that for

$$j+1, v_{j+1}(y_{j+1}) = max_{y_j}[s(x,j,y_j,y_{j+1}) + v_j(y_j)]$$

$$v_{j+1}(y_{j+1}) = max_{y_1,..,y_j}\sum_{i=1}^{j+1} s(x,j,y_{j-1},y_j) = max_{y_j}[s(x,j,y_j,y_{j+1}) + max_{y_{j-1}}[s(x,j,y_{j-1},y_j) + max_{y_{j-2}}......)]]$$

2.2

$$O(K^2N)$$

where K is the number of state, N is the length of the sequence

3 Programming: Implementing the Viterbi algorithm

3.1 Coding the Viterbi Algorithm

Solution.

(Parts 1 2)

Currently, our solution passes the test decoder test.
However, it fails on the test and dev sets.

Dev Set:

processed 51578 tokens with 5917 phrases; found: 1 phrases; correct: 0.

accuracy: 24.34

accuracy: 4.06

LOC: precision: 0.00

MISC: precision: 0.00

ORG: precision: 0.00

PER: precision: 0.00

Test Set:

processed 46666 tokens with 5616 phrases; found: 1 phrases; correct: 0.

accuracy: 23.66

accuracy: 4.11

LOC: precision: 0.00

MISC: precision: 0.00

ORG: precision: 0.00

PER: precision: 0.00