CSE 143 Writeup Assignment 3
Aaron Steinberg, Andrew Moshkovich, James Gu

1 Programming: Text Classification with Neural Networks

1.1 Programming: Text Classification with RNNs

SimpleRNN Epochs 1-10:

Epoch:	1/20	2/20	3/20	4/20	5/20	6/20	7/20	8/20	9/20	10/20
Loss:	0.6990	0.6484	0.4099	0.2618	0.1952	0.1547	0.1286	0.1043	0.0844	0.0802
Accuracy:	0.5044	0.6119	0.8048	0.8899	0.9229	0.9387	0.9508	0.9627	0.9685	0.9709

SimpleRNN Epochs 11-20:

Epoch:	11/20	12/20	13/20	14/20	15/20	16/20	17/20	18/20	19/20	20/20
Loss:	0.692	0.0681	0.0680	0.0675	0.0599	0.0535	0.0544	0.0555	0.0643	0.0791
Accuracy:	0.9748	0.9754	0.9746	0.9761	0.9784	0.9809	0.9808	0.9806	0.9768	0.9717

1.2 Programming: Text Classification with LSTMs

LSTM Epochs 1-10:

Epoch:	1/20	2/20	3/20	4/20	5/20	6/20	7/20	8/20	9/20	10/20
Loss:	0.5155	0.3369	0.1932	0.1586	0.1418	0.1156	0.0914	0.0692	0.0608	0.0508
Accuracy:	0.7441	0.8600	0.9299	0.9445	0.9502	0.9590	0.9691	0.9762	0.9794	0.9832

LSTM Epochs 11-20:

Epoch:	11/20	12/20	13/20	14/20	15/20	16/20	17/20	18/20	19/20	20/20
Loss:	0.0472	0.0422	0.0356	0.0275	0.0300	0.0229	0.0190	0.0198	0.0152	0.0169
Accuracy:	0.9845	0.9857	09887	0.9915	0.9909	0.9930	0.9933	0.9940	0.9952	0.9948

Same number of epochs as RNN, but accuracies were betterafter longer sequences. At half the epochs (Epoch 10) LSTM was already better than the best accuracy for RNN.

2 Theory: Deriving the Viterbi Algorithm

1.

Given: $S(x,y) = \sum_{i=1}^{n+1} s(x,\ i,\ y_{i-1}, y_i)$

Let, for every possible value of y_j, $v_j(y_j) = max_{y_1,\ldots,y_{j-1}} \sum_{i=1}^{j} s(x,\ i,\ y_{i-1}, y_i)$

Proof:

$$v_j(y_j) = max_{y_1,\ldots,y_{j-1}} \sum_{i=1}^{j} s(x,\ i,\ y_{i-1}, y_i)$$

$$v_j(y_j) = max[\ max_{y_1,\ldots,y_j} \sum_{i=1}^{j-1} s(x,\ i,\ y_{i-1}, y_i) + max \sum_{i=j}^{j} s(x,\ i,\ y_{i-1}, y_i)]$$

$$v_j(y_j) = max[\ max_{y_1,\ldots,y_j} \sum_{i=1}^{j-1} s(x,\ i,\ y_{i-1}, y_i) + max\ s(x,\ j,\ y_{j-1}, y_j)]$$

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

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

2.

We predict that the will be time complexity O(n) and the space complexity will be O(n) as well.

3 Programming: Implementing the Viterbi Algorithm

3.1 Coding the Viterbi Algorithm

We weren't able to get a full solution working for #3, as we tried to follow the pseudo code given in the textbook. We were able to initialize the matrices and figure out the for loops, as well as M and K. However, we especially had trouble with the max and argmax statements, so we weren't able to write in the solutions for those parts. We also had trouble figuring out the dimensions of y, so we couldn't implement that. Apart from those two things, we were able to follow the pseudo code given to us.