Long short-term memory networks (LSTMs)

1

Show that the simple RNN (as defined by PyTorch) *with sigmoid nonlinearity* is a subset of the LSTM (as defined by PyTorch).

Show how to set the weights of the LSTM to mimic the simple RNN. They will be functions of the RNN weight matrices.

2

Documents contain the words "bad", "good", "not", and "uh". A sentiment score is computed for each document as follows:

$$\sum \text{“good”} - \sum \text{“bad”} - 2\sum \text{“not good”} + 2\sum \text{“not bad”}$$

If the inputs to an LSTM are one-hot-encoded words (using alphabetical order), and the output at each time is the cumulative sentiment score, manually identify specific LSTM sizes and weights that solve this problem.

Turn in a .pdf file. No code required.