

CAP 6619 Deep Learning

2023 Fall

Homework 4 [8 Pts, Due: Nov 25 2023. Late Penalty: -2/day]

[If two homework submissions are found to be similar to each other, both submissions will receive 0 grade]

[Homework solutions must be submitted through Canvas. No email submission is accepted. If you have multiple files, please include all files as one zip file, and submit zip file online (only zip, pdf, or word files are allowed). You can always update your submissions. Only the latest version will be graded.]

Question 1 [2 pts]: Figure 1 shows the structure of an RNN cell vs. an LSTM cell.

- Summarize major difference between RNN cell vs LSTM cell in terms of their neural architectures [1 pt]
- Why LSTM can achieve long-short term memory, whereas RNN cannot [1 pt]

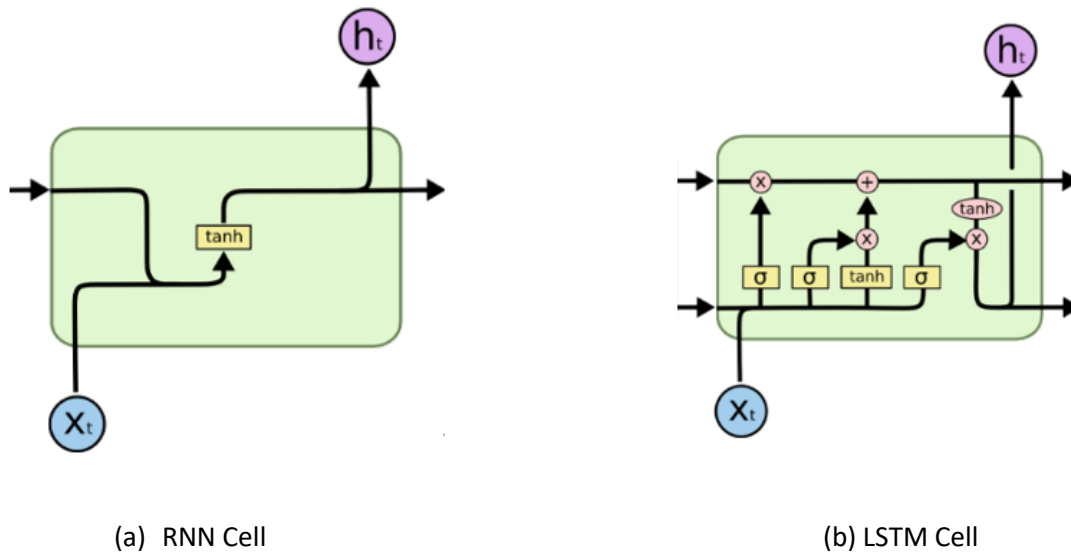


Figure 1: RNN vs LSTM Cell

Question 2 [2 pts]: Figure 1 shows the structure of an LSTM cell, and its connection to adjacent cells to form a recurrent neural network.

- Please mark forget gate, input gate, output gate, and candidate layer, respectively [1 pt]
- Explain the main role/functionality of the forget gate, input gate, output gate, and candidate layer, respectively [1 pt]

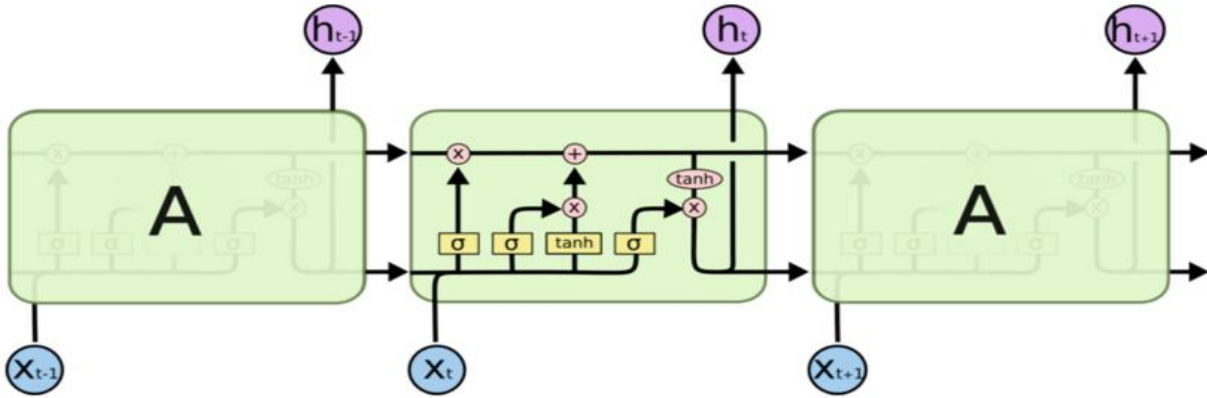


Figure 2: LSTM recurrent neural network

Question 3 [2 pts]: Figure 3 shows unfolded LSTM network with two consecutive cells. Using h_t and c_t to denote output and cell memory of the cell at time point t . Use f_t , i_t , o_t to denote forget gate, input gate, output gate of the cell at time point t . Use \tilde{c}_t to denote candidate layer output at time point t .

1. Use mathematical equations to show the relationship between cell memory at time point t , with respect to cell memory and output at time point $t-2$. [0.5 pt]
2. Use δC_t to denote change of network error with respect to cell state at time point t , i.e., $\delta C_t = \frac{\partial E}{\partial C_t}$.
 - a. Derive relationship between δC_{t-2} and δC_t [0.5 pt]
 - b. Explain why LSTM cell can alleviate weight vanishing or exploding in deep neural network learning. [1 pt]

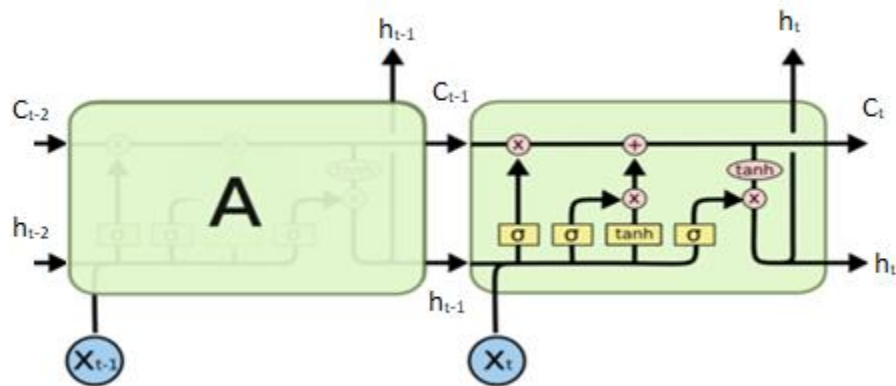


Figure 3: unfolded LSTM cells

Question 4 [2 pts]: The following Keras codes show a deep learning network for text classification (only the network structure part).

1. What is the purpose of the Embedding()? The Embedding() layer output size is 16, what does this mean?, what is the number of weight parameters for the Embedding() layer (show your solutions) [1 pt]
2. What is the number of weight parameters for the for the LSTM () layer (Show your solutions) [0.5 pt].
3. What is the total number of weight parameters for the last two dense layers (show your solutions) [0.5 pt].

```
model = Sequential()  
model.add(Embedding(1,000, 16, input_length=200))  
model.add(LSTM(32, dropout=0.1, recurrent_dropout=0.1))  
model.add(Flatten())  
model.add(Dropout(0.1))  
model.add(Dense(256, activation='sigmoid'))  
model.add(Dense(1, activation='sigmoid'))
```