CAP6619.homework4-1

Tuesday, November 21, 2023 4:02 AM

CAP6619.homework4-1

CAP 6619 Deep Learning

2023 Fall

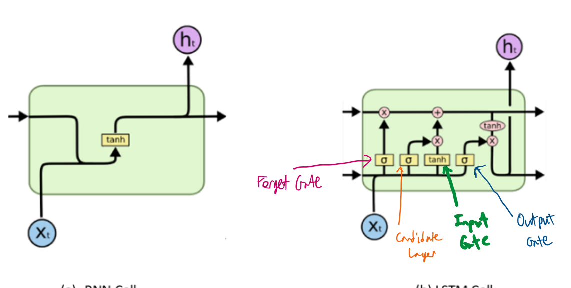
Homework 4 [8 Pts, Due: Nov 25 2023. Late Penalty: -2/day]

[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]



(a) RNN Cell (b) LSTM Cell

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]

RNN - Recurrent Neural Network

LSTM - Long-term Short-Term Memory Networks

1a) The major difference between an RNN and LSTM is in the repeating module. The repeating module of a standard RNN contains only a single layer, The repeating module in an LSTM contains four interacting layers The RNN processes inputs and has a hidden state that updated at each time step. This leads to difficulty in learning and remembering long-term dependencies. The LSTM has memory cells and gating mechanisms that provide better control of information flow and for this reason operate better with sequential data, especially when there are long range dependencies. Overall, the LSTM is more complex,

(1b) LSTMs include memory cells that can store and retrieve information over long sequences, The gating mechanisms control the flow of information through the network.

(2a) I marked the forget gate, input gate, output gate and candidate layer on the drawing above.

(2b) Forget gate - A NN w/ sigmoid

- Is responsible for determining what information from the cell state (long-term memory) should be discarded or forgotten.

candidate layer - A NN w/ Tanh

- Computes candidate values that could potentially be added to the cell state. These are stored in a vector called a candidate update.

Input gate - A NN w/ sigmoid

- Is responsible for determining what new information should be added to the cell state. Determines how much of the candidate update is added to the cell state.

Output gate - A NN w/ sigmoid

- Determines what information from the cell state should be exposed as the hidden state

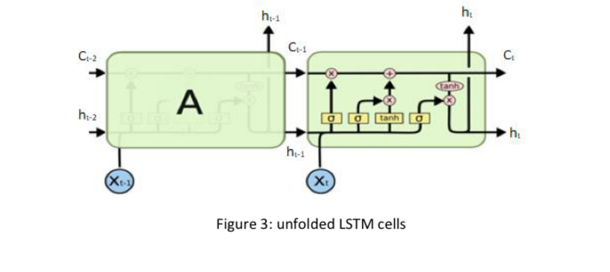


Figure 2: LSTM recurrent neural network

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'))
```

(4.1) The purpose of the Embedding() layer is used to create word embeddings for input text data. Word embeddings are vector representations of words.
The output layer of 16 means that each word will be represented by a vector of length 16.

Weights =

input $\times$ output

$1000 \times 16 = 16{,}000$

(4.2)

$4 * (32 * (16 + 32 + 1)) =$

$4 * (32 * 49) = 6{,}272$

(4.3) Dense Layer 1 = $256 * (32+1) = 8448$

Dense Layer 2 = $1 * (256+1) = 257$

Total = 8,705

Proof for all of these:

```
import tensorflow as tf
from keras import layers, models
from keras import Sequential

from keras.models import Sequential
from keras.layers import Embedding, LSTM, Flatten, Dropout, Dense



model = Sequential()
model.add(Embedding(1000, 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'))

model.summary()
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS Python

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 200, 16)	16000
lstm (LSTM)	(None, 32)	6272
flatten (Flatten)	(None, 32)	0
dropout (Dropout)	(None, 32)	0
dense (Dense)	(None, 256)	8448
dense_1 (Dense)	(None, 1)	257

```
Total params: 30977 (121.00 KB)
Trainable params: 30977 (121.00 KB)
Non-trainable params: 0 (0.00 Byte)
```