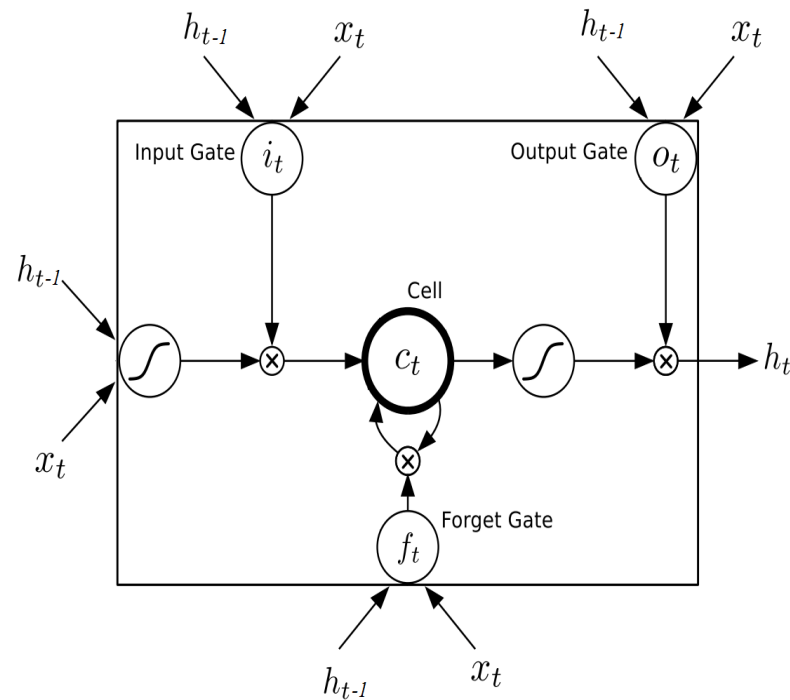![edX](edX logo)

Course > Unit 3 Neural networks (2.5 weeks) > Homework 4 > 2. LSTM

2. LSTM

Extension Note: Homework 4 due date has been extended by 1 day to **July 27 23:59UTC** .

The diagram below shows a single LSTM unit that consists of Input, Output, and Forget gates.

The behavior of such a unit as a recurrent neural network is specified by a set of update equations. These equations define how the gates, "memory cell" c_t and the "visible state" h_t are updated in response to input x_t and previous states c_{t-1}, h_{t-1}. For the LSTM unit,

$$f_t \quad = \operatorname{sigmoid}\left(W^{f,h} h_{t-1} + W^{f,x} x_t + b_f\right)$$

$$i_t \quad = \operatorname{sigmoid}\left(W^{i,h} h_{t-1} + W^{i,x} x_t + b_i\right)$$

$$o_t \quad = \operatorname{sigmoid}\left(W^{o,h} h_{t-1} + W^{o,x} x_t + b_o\right)$$

$$c_t \quad = f_t \odot c_{t-1} + i_t \odot \tanh\left(W^{c,h} h_{t-1} + W^{c,x} x_t + b_c\right)$$

$$h_t = o_t \odot \tanh(c_t)$$

where symbol $\odot$ stands for element-wise multiplication. The adjustable parameters in this unit are matrices $W^{f,h}$, $W^{f,x}$, $W^{i,h}$, $W^{i,x}$, $W^{o,h}$, $W^{o,x}$, $W^{c,h}$, $W^{c,x}$, as well as the offset parameter vectors b_f, b_i, b_o, and b_c. By changing these parameters, we change how the unit evolves as a function of inputs x_t.

To keep things simple, in this problem we assume that x_t, c_t, and h_t are all scalars. Concretely, suppose that the parameters are given by

$$
\begin{array}{llllll}
W^{f,h} &= 0 & W^{f,x} &= 0 & b_f &= -100 & W^{c,h} &= -100 \\
W^{i,h} &= 0 & W^{i,x} &= 100 & b_i &= 100 & W^{c,x} &= 50 \\
W^{o,h} &= 0 & W^{o,x} &= 100 & b_o &= 0, & b_c &= 0
\end{array}
$$

We run this unit with initial conditions $h_{-1} = 0$ and $c_{-1} = 0$, and in response to the following input sequence: [0, 0, 1, 1, 1, 0] (For example, $x_0 = 0$, $x_1 = 0$, $x_2 = 1$, and so on).

LSTM states

1.0/1 point (graded)
Calculate the values h_t at each time-step and enter them below as an array $[h_0, h_1, h_2, h_3, h_4, h_5]$. For ease of calculation, you can assume that you round h_t to the closest integer in every time-step. E.g., assume $\text{sigmoid}(50) \approx 1$ and $\tanh(-50) \approx -1$.

[0,0,1,-1,1,0]

✔ **Answer:** [0, 0, 1, -1, 1, 0]

Solution:

Approximating the functions to the nearest integer and assuming that x_t is only 0 or 1 simplifies the equation to the following.

$$
\begin{aligned}
f_t &= \text{sigmoid}\,(-100) = 0 \\
i_t &= \text{sigmoid}\,(100x_t + 100) = 1 \\
o_t &= \text{sigmoid}\,(100x_t) \\
c_t &= 0 \odot c_{t-1} + 1 \odot \tanh\,(-100h_{t-1} + 50x_t) = \tanh\,(-100h_{t-1} + 50x_t) \\
h_t &= o_t \odot \tanh\,(c_t)
\end{aligned}
$$

Notice that for c_t, the h_{t-1} term overpowers the x_t one, unless h_{t-1} is 0.

Based on our simplifications above, we can find the values for each h_t.

Input 1:

$$
\begin{aligned}
f_0 &= 0 & i_0 &= 1 & o_0 &= 0.5 & c_0 &= \tanh\,(-100\,(0) + 50\,(0)) = 0 & h_0 &= 0 \odot \tanh\,(0) = 0 \\
f_1 &= 0 & i_1 &= 1 & o_1 &= 0.5 & c_1 &= \tanh\,(-100\,(0) + 50\,(0)) = 0 & h_1 &= 0 \\
f_2 &= 0 & i_2 &= 1 & o_2 &= 1 & c_2 &= \tanh\,(0 + 50) = 1 & h_2 &= 1\tanh\,(1) = .76 \text{ rounded to } 1
\end{aligned}
$$

Continue in this manner.

Submit   You have used 2 of 3 attempts

ⓘ Answers are displayed within the problem

LSTM states 2

1.0/1 point (graded)

Now, we run the same model again with the same parameters and same initial conditions as in the previous question. The only difference is that our input sequence in now: [1, 1, 0, 1, 1].

Calculate the values h_t at each time-step and enter them below as an array $[h_0, h_1, h_2, h_3, h_4]$. For ease of calculation, you can assume that you round h_t to the closest integer in every time-step. E.g., assume $\mathrm{sigmoid}\,(50) \approx 1$ and $\tanh\,(-50) \approx -1$.

[1,-1,0,1,-1]   ✔ **Answer:** [1, -1, 0, 1, -1]

Solution:

The computation is similar to the previous question.

| Submit | You have used 1 of 3 attempts

ⓘ Answers are displayed within the problem

LSTM info

1/1 point (graded)

What information is carried in the state h_t?

○ Whether the total number of zeros is odd.

○ Whether the number of consecutive zeros is odd.

○ Whether the total number of ones is odd.

◉ Whether the number of consecutive ones is odd. ✔

Solution:

We can observe that the network counts the number of consecutive 1's. If it is currently seeing a 0 it outputs 0, otherwise it outputs a 1 if it has seen an odd number of 1's so far, and a -1 if it is even.

Submit	You have used 0 of 1 attempt

ℹ   Answers are displayed within the problem

Discussion   **Hide Discussion**

Topic: Unit 3 Neural networks (2.5 weeks):Homework 4 / 2. LSTM

Add a Post

Show all posts ▼
by recent activity ▼

💬 [staff] Can we have clarification on what should be rounded?

📌 Pinned   👤 Community TA

6

☑ [STAFF] Can you please check my LSTM state and State 2 Answer.

I have made two unsuccessful attempts at calculating the above values. i have skimmed over my calculation but still unaware of the mistake im d...

2

☑ [Staff] LSTM info

I got the first two correct. Can you check my answer? I feel the state can be interpreted like that. Also can we increase number of attempts to 2 si...

6

💬 For sigmoid(0)=0.5, do you round it to 1 or keep it?

In my first iteration t=0, I get a sigmoid(0)=0.5, should I round it or leave it? Staff: could you check if my answers are off because of that? Thanks!

2

☑ [Staff] LSTM states 2, could you check my answer!

I have used Excel to answer the first question and I got it right finally. The second question is exactly same as the first one except for the input, I j...

3

💬 H_0 and H_1?

The first question says, initialize H_1=0. But then the solution asks for H_0. I suppose the latter is equal to the former?

2

💬 [STAFF] could you plz help check my answer (LSTM state)

I compute each round with the parameters in the question, I tried both none np.round() and np.round(), but the results were incorrect. Please he...

9

💬 [STAFF]Could you please at least add some marks if there is a inner link between 2 questions?

I've spent a lot of time calculating LSTM states on hand, and I got the right answer. Then I found that I have to do that whole bunch of calculation ...

1

❓ I don't know what's wrong about mymy LSTM state Answer, please check it.Thank you!

There is one problem about the LSTM state , - If we need to round sigmond(0)=1 , or keep it in 0.5 I have used up all my chance about to this que...

1

💬 LSTM - R code

Below is the R code for the LSTM calculation: ft = round(sigmoid(w_fh * h_t_1 + w_fx * x_t + b_f)) it = round(sigmoid(w_ih * h_t_1 + w_ix * x_t + b_i)...

3

must round ht to the nearest integer

i think that you should write that we must round ht to the nearest integer. I entered the decimal values and my answer was wrong

1

No success after 2 attempts. Seems very wrong to me.

This is simply wrong !!! I have 27 years of programing and several patents in machine learning. This subject of LSTM is new to me and most recur...

12 new_

@Staff mistake in the last question

3