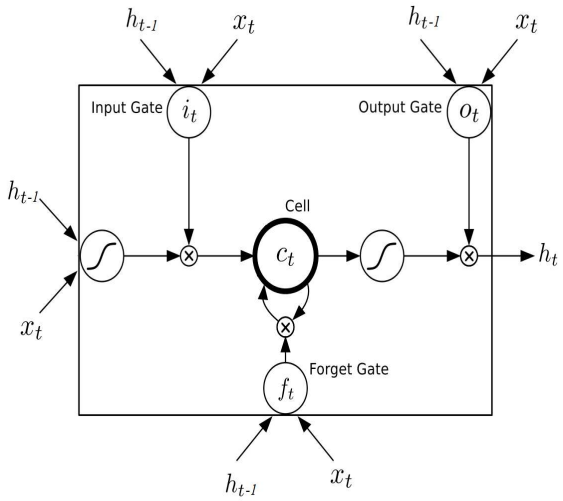
## 2. LSTM

Bookmark this page

Homework due Oct 28, 2020 05:29 IST *Completed*

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,

$$
\begin{aligned}
f_t &= \text{sigmoid}\left(W^{f,h}h_{t-1} + W^{f,x}x_t + b_f\right) \\
i_t &= \text{sigmoid}\left(W^{i,h}h_{t-1} + W^{i,x}x_t + b_i\right) \\
o_t &= \text{sigmoid}\left(W^{o,h}h_{t-1} + W^{o,x}x_t + b_o\right) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tanh\left(W^{c,h}h_{t-1} + W^{c,x}x_t + b_c\right) \\
h_t &= o_t \odot \tanh\left(c_t\right)
\end{aligned}
$$

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 $[h_0, h_1, h_2, h_3, h_4, h_5]$.

(Please round $h_t$ to the closest integer in every time-step. If $h_t = \pm 0.5$, then round it to $0$.
For ease of calculation, assume that $\text{sigmoid}\left(x\right) \approx 1$ and $\tanh\left(x\right) \approx 1$ for $x \geq 1$, and $\text{sigmoid}\left(x\right) \approx 0$ and $\tanh\left(x\right) \approx -1$ for $x \leq -1$.)

[ 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.

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

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:

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

Continue in this manner.

Submit You have used 2 of 5 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 $[h_0, h_1, h_2, h_3, h_4]$.

(Please round $h_t$ to the closest integer in every time-step. If $h_t = \pm 0.5$, then round it to $0$.
For ease of calculation, assume that $\text{sigmoid}\left(x\right) \approx 1$ and $\tanh\left(x\right) \approx 1$ for $x \geq 1$, and $\text{sigmoid}\left(x\right) \approx 0$ and $\tanh\left(x\right) \approx -1$ for $x \leq -1$.)

[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 5 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 1 of 1 attempt

Answers are displayed within the problem

## Discussion

Hide Discussion

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

Add a Post

Show all posts by recent activity

| | | |
|---|---|---|
| ? | LSTM States. All values rights except one!<br>On LSTM States 1 and 2 I'm getting all values right in my array except for one! On LSTM 1 I'm getting [ 0, 0, 1, -1, 1, -1,] and on LSTM 2 I'... | 2 |
| ☑ | What are we supposed to learn from "LSTM Info"?<br>Except for the fact that LSTMs can be very expressive? | 3 |
| 💬 | Be careful with LSTM state update | 9 |
| 💬 | Risky questions<br>I got right in the above questions, but it was a **extremely** handwork exercise. Even using a software aid it was extremely painful and... | 10 |
| 💬 | Hint, for last question 'LSTM info'<br>Run the update with the following sequence, [1, 0, 1, 1, 1, 1, 0] Why? Total number of 0/1 and number of consecutive 0/1 whether it is od... | 4 |
| ? | (staff) The meaning of LSTM info<br>I am the only one who cannot understand the meaning.. Consecutive means f,i,o,c,h at the same order? And there are three kinds of h. ... | 4 |
| 💬 | Could not format HTML for problem.<br>It says: "Could not format HTML for problem. Contact course staff in the discussion forum for assistance." What am I to do? I just tried to... | 2 |
| ☑ | Interpretation of choices in "LSTM Info"<br>"Whether the number of consecutive ones is odd": Does occurrence of a single appearance of 1 meet this criteria or do we need a mini... | 2 |
| 💬 | Sigmoid function<br>Hi, What is sigmoid function exactly?? I thought that it is 1/(e^-x +1), but I think it is not the function. please any one can help me?? thanks | 4 |
| 💬 | LSTM States with Python (general steps)<br>Community TA | 4 |
| 💬 | Another approach<br>Highly recommend pulling values and calculation through a Google sheet / Excel. Implement tanh and sigmoid with a couple of IF(). I did... <br>Community TA | 3 |

Previous

Next



## edX

About
Affiliates
edX for Business
Open edX
Careers
News

## Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

## Connect

Blog
Contact Us
Help Center
Media Kit
Donate




/