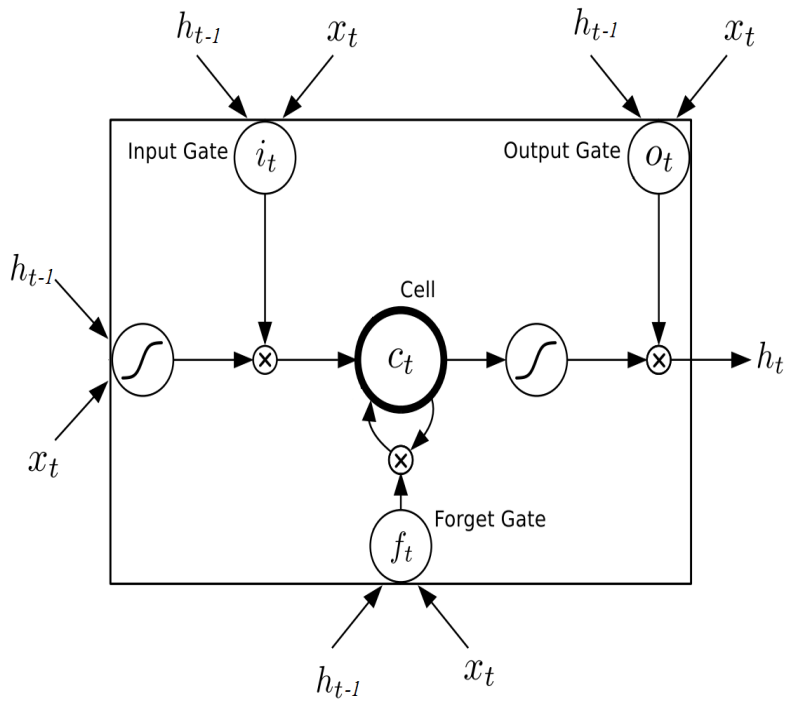
2. LSTM

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,

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

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

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

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 $\text{sigmoid}\left(50\right) \approx 1$ and $\tanh\left(-50\right) \approx -1$.

[0.0, 0.0, 1.0, -1.0, 1.0, 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:

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

Continue in this manner.

Submit You have used 3 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 $\text{sigmoid}\left(50\right) \approx 1$ and $\tanh\left(-50\right) \approx -1$.

[1.0, -1.0, 0.0, 1.0, -1.0] ✔ **Answer:** [1, -1, 0, 1, -1]

Solution:

The computation is similar to the previous question.

Submit You have used 3 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

Answers are displayed within the problem

Discussion

Show Discussion

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