DL HW week 8

Hong Pengfei 1002949

Task 1:

1. No, h_{t-1} can be a function of C_{t-1}. $h_{t-1} = \sigma(W^o x_{t-1} + U^o h_{t-2}) \cdot \tanh(C_{t-1})$ $\therefore$ h_{t-1} can be a function of x_{t-1}, h_{t-2}, and C_{t-1}. Therefore h_{t-1} does not determine C_{t-1} but otherwise

2.
$$\frac{\partial h_t}{\partial h_{t-1}} = \frac{\partial h_t}{\partial C_t} \cdot \left(\frac{\partial C_t}{\partial f_t}\frac{\partial f_t}{\partial h_{t-1}} + \frac{\partial C_t}{\partial i_t}\frac{\partial i_t}{\partial h_{t-1}} + \frac{\partial C_t}{\partial u_t}\frac{\partial u_t}{\partial h_{t-1}} \right) + \frac{\partial h_t}{\partial O_t}\frac{\partial O_t}{\partial h_{t-1}}$$

$$\frac{\partial h_t}{\partial C_t} = O_t \circ \operatorname{sech}^2(C_t)$$

$$\frac{\partial C_t}{\partial f_t} = C_{t-1} \quad \frac{\partial C_t}{\partial i_t} = u_t \quad \frac{\partial C_t}{\partial u_t} = i_t \quad \frac{\partial h_t}{\partial O_t} = \tanh(C_t)$$

$$\therefore \frac{\partial h_t}{\partial h_{t-1}} = O_t \circ \operatorname{sech}^2(C_t) \circ C_{t-1} \circ \frac{\partial f_t}{\partial h_{t-1}} + O_t \circ \operatorname{sech}^2(C_t) \circ u_t \circ \frac{\partial i_t}{\partial h_{t-1}} + O_t \circ \operatorname{sech}^2(C_t) \circ i_t \circ \frac{\partial u_t}{\partial h_{t-1}}$$
$$+ \tanh(C_t) \circ \frac{\partial O_{t-1}}{\partial h_{t-1}}$$

In this expression:
$$\begin{cases} C_t = f_t \circ C_{t-1} + i_t \circ u_t \\ f_t = \sigma(W^f x_t + U^f h_{t-1}) \\ i_t = \sigma(W^i x_t + U^i h_{t-1}) \end{cases}$$

3. $\sigma'(z) = \sigma(z) \cdot (1 - \sigma(z))$

4. $\frac{\partial f_t}{\partial h_{t-1}} = \sigma(W^f x_t + U^f h_{t-1}) \cdot [1 - \sigma(W^f x_t + U^f h_{t-1})] \cdot U^f$

5. as sigmoid is an increasing function we need to $\operatorname{argmax}\{U^f(d) \cdot h_{t-1}\}$
when h_{t-1} is in the same direction as $U^f(d)$ the value is maximized $\therefore h_{t-1} = \frac{U^f(d)}{\|U^f(d)\|}$

6. Both No. $W^f(d) \cdot \vec{x}_t = \vec{0}$ therefore it will be not affecting both max or argmax.

Task 2.

1)
$$h_t = 1 + \operatorname{int}\left(\frac{78 + 2\times2 - 5}{3}\right) = 26$$
$$W_t = 1 + \operatorname{int}\left(\frac{84 + 2\times2 - 5}{3}\right) = 28$$
$\therefore$ output shape will be $(26, 28)$

2. $h = 1 + \text{int}\left(\frac{64-3}{2}\right) = 31$

$w = 1 + \text{int}\left(\frac{64-5}{2}\right) = 30$

$\therefore$ spatial output size is (31, 30)

3. $16 = 1 + \text{int}\left(\frac{x+2-9}{3}\right)$

$\Rightarrow 15 = \text{int}\left(\frac{x+2-9}{3}\right)$

$\Rightarrow 15 \leq \frac{x+2-9}{3} < 16$

$\Rightarrow 45 \leq x-7 < 48$

$52 \leq x < 55$

$\therefore$ x can be 52 53 54

$\therefore$ the input size can be (52, n) or (53, n) or (54, n) n can be any positive integer.

4.

① $7 \times 7 \times 64 \times 32 = 100352$

② $h = 1 + \text{int}\left(\frac{19-7}{3}\right) = 5$

$w = h = 5$

$\therefore$ total times of multiplications $= 5 \times 5 \times 100352 = 2508800$

$\therefore$ total times of addition $= 5 \times 5 \times (7^2 \times 32 - 1) \times 64 = 2507200$

③

$128 \times 512 = 65536$

$w = h = 1 + \text{int}(25+2-1) = 27$

total times of multiplication $= 27^2 \times 65536 = 47775744$

total times of addition $= (512-1) \times 128 \times 27 \times 27 = 47682432$