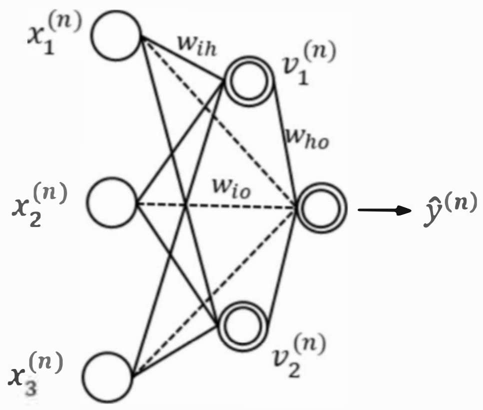
2.



a) $\hat{y}^{(n)} = \sigma(\sum_h w_{ho} v_h^{(n)} + \sum_i w_{io} x_i^{(n)})$，其中 $v_h^{(n)} = \sigma(\sum_i w_{ih} x_i^{(n)})$。

b) 记 $\hat{y}^{(n)} = \sigma(z^{(n)})$，$z^{(n)} = \sum_h w_{ho} v_h^{(n)} + \sum_i w_{io} x_i^{(n)}$，有 $\frac{\partial Loss}{\partial w_{io}} = \sum_n \frac{\partial Loss}{\partial \hat{y}^{(n)}} \cdot \frac{\partial \hat{y}^{(n)}}{\partial z^{(n)}} \cdot \frac{\partial z^{(n)}}{\partial w_{io}}$

其中，$\frac{\partial Loss}{\partial \hat{y}^{(n)}} = \frac{1}{N} 2(\hat{y}^{(n)} - y^{(n)})$。由

$$\sigma(x) = \frac{1}{1+e^{-x}} \Rightarrow \sigma'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \sigma(x)\cdot(1-\sigma(x))$$

得 $\frac{\partial \hat{y}^{(n)}}{\partial z^{(n)}} = \hat{y}^{(n)}(1-\hat{y}^{(n)})$。又有 $\frac{\partial z^{(n)}}{\partial w_{io}} = x_i^{(n)}$，得

$$\frac{\partial Loss}{\partial w_{io}} = \frac{1}{N}\sum_n 2(\hat{y}^{(n)} - y^{(n)})\, \hat{y}^{(n)}(1-\hat{y}^{(n)})\, x_i^{(n)}.$$

c) 先求 $\frac{\partial Loss}{\partial x_i^{(n)}}$。$\frac{\partial Loss}{\partial x_i^{(n)}} = \frac{\partial Loss}{\partial \hat{y}^{(n)}} \cdot \frac{\partial \hat{y}^{(n)}}{\partial z^{(n)}} \cdot \frac{\partial z^{(n)}}{\partial x_i^{(n)}}$，前两项 b) 已求出，下求最后一项

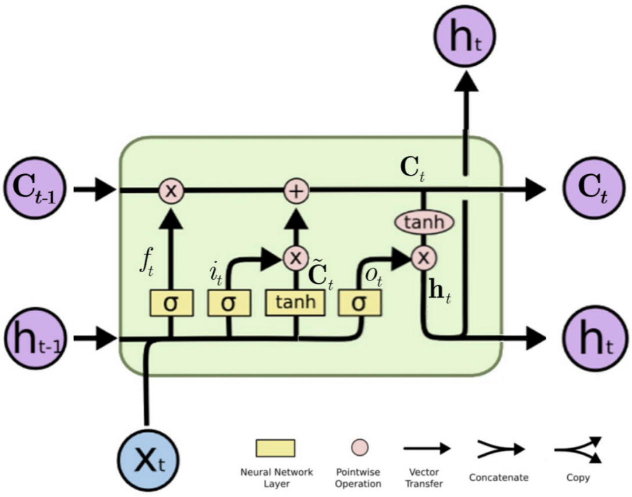
$$\frac{\partial z^{(n)}}{\partial x_i^{(n)}} = \frac{\partial}{\partial x_i^{(n)}}(\sum_h w_{ho} v_h^{(n)}) + w_{io}$$

其中 $\frac{\partial v_h^{(n)}}{\partial x_i^{(n)}} = v_h^{(n)}(1-v_h^{(n)}) w_{ih} \Rightarrow \frac{\partial z^{(n)}}{\partial x_i^{(n)}} = (\sum_h v_h^{(n)}(1-v_h^{(n)}) w_{ih} w_{ho}) + w_{io}$

得 $\frac{\partial Loss}{\partial x_i^{(n)}} = 2(\hat{y}^{(n)} - y^{(n)})\, \hat{y}^{(n)}(1-\hat{y}^{(n)}) \left(\sum_h v_h^{(n)}(1-v_h^{(n)}) w_{ih} w_{ho} + w_{io}\right)$

故参数更新公式为 $x_i^{(n)} \leftarrow x_i^{(n)} - \eta \frac{\partial Loss}{\partial x_i^{(n)}}$，其中 $\frac{\partial Loss}{\partial x_i^{(n)}}$ 由上式给出。

4、



(1) $f_t = \sigma(W_f[h_{t-1}, x_t]) = \sigma([0.5\ \ 0.5]\begin{bmatrix}0\\1\end{bmatrix}) = \sigma(0.5) = 0.6225$

$i_t = \sigma(W_i[h_{t-1}, x_t]) = \sigma([0.4\ \ 0.4]\begin{bmatrix}0\\1\end{bmatrix}) = \sigma(0.4) = 0.5987$

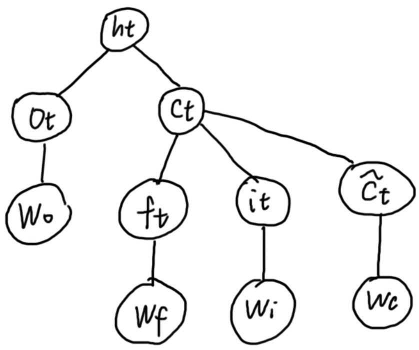
$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t]) = \tanh([0.4\ \ 0.4]\begin{bmatrix}0\\1\end{bmatrix}) = \tanh(0.4) = 0.3799$

$O_t = \sigma(W_o[h_{t-1}, x_t]) = \sigma([0.5\ \ 0.5]\begin{bmatrix}0\\1\end{bmatrix}) = \sigma(0.5) = 0.6225$

$C_t = f_t C_{t-1} + i_t \tilde{C}_t = 0.6225 \times 0 + 0.5987 \times 0.3799 = 0.2275$

$h_t = O_t \cdot \tanh(C_t) = 0.6225 \times \tanh(0.2275) = 0.1392$

(2) 我们认为h_{t-1}，C_{t-1}与x_t无需梯度（此处只关注单个时间步），先画出简单的计算图



记$W_o = [W_{o1}, W_{o2}]$. $\frac{\partial L}{\partial W_{oi}} = \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial W_{oi}}$，其中$\frac{\partial L}{\partial h_t} = h_t - d = 0.1392 - 0.6 = -0.4608$

$\frac{\partial h_t}{\partial W_{oi}} = \frac{\partial h_t}{\partial O_t} \cdot \frac{\partial O_t}{\partial W_{oi}}$，其中$\frac{\partial h_t}{\partial O_t} = \tanh(C_t) = \tanh(0.2275) = 0.2236$

$\frac{\partial O_t}{\partial W_{o1}} = O_t(1-O_t)h_{t-1} = 0$，$\frac{\partial O_t}{\partial W_{o2}} = O_t(1-O_t)x_t = 0.6225 \times (1-0.6225) \times 1 = 0.2350$

故$\frac{\partial L}{\partial W_{o1}} = 0$，$\frac{\partial L}{\partial W_{o2}} = \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial O_t} \cdot \frac{\partial O_t}{\partial W_{o2}} = -0.4608 \times 0.2236 \times 0.2350 = -0.0242$

反向传播更新：$W_o \leftarrow W_o - \alpha(\frac{\partial L}{\partial W_{o1}}\ \ \frac{\partial L}{\partial W_{o2}}) = [0.5\ \ 0.5] - 0.1 \times [0\ \ -0.0242]$

$= [0.5000\ \ 0.5024]$

注：此处数值计算是用计算机进行的，保留四位小数抄写在此，实际的计算精度要更高一些。