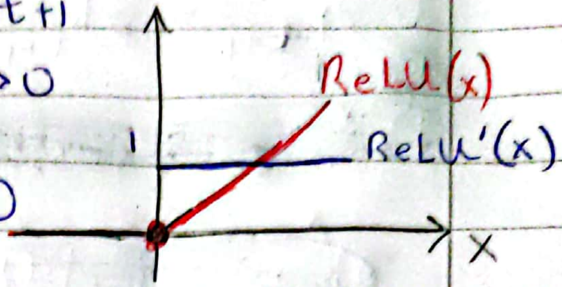
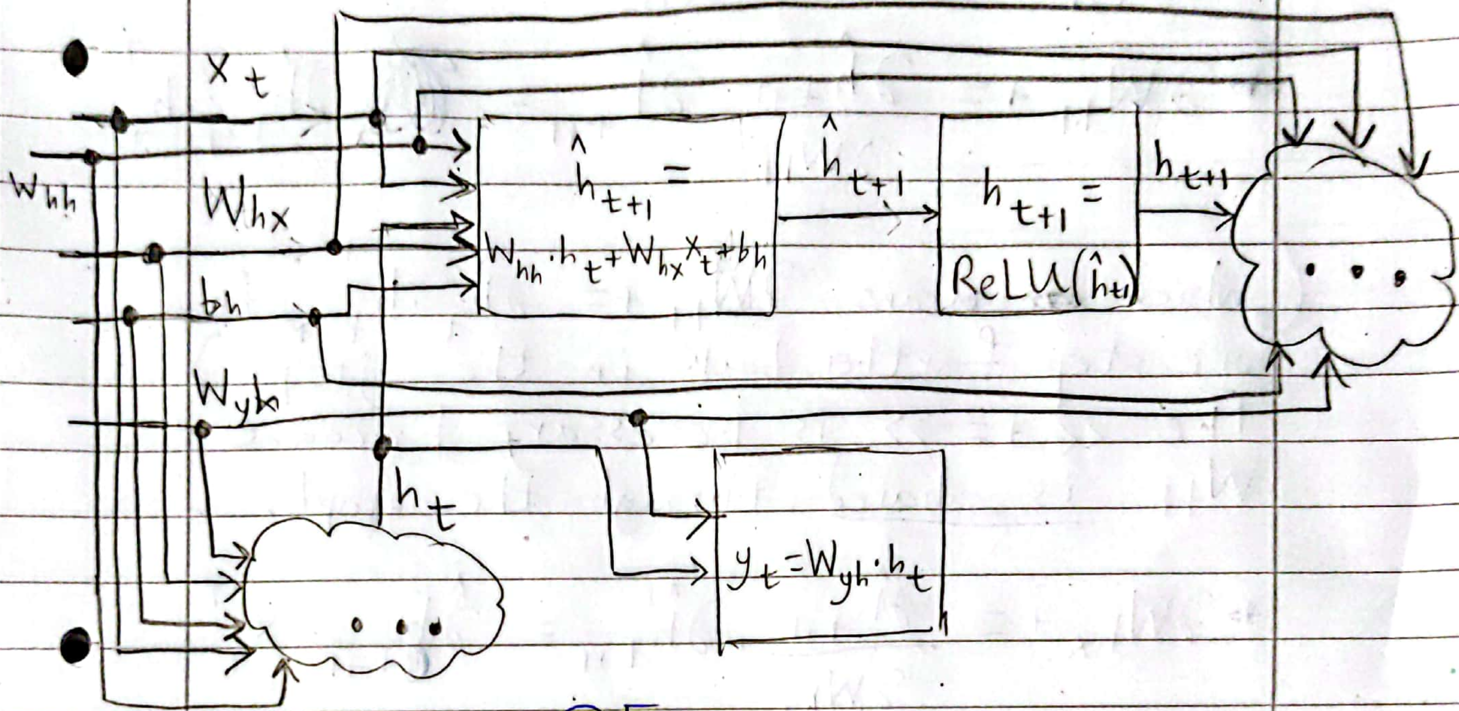
- $$h_{t+1} = ReLU(\underbrace{W_{hh} \cdot h_t + W_{hx} \cdot x_t + b_h}_{\hat{h}_{t+1}})$$

$$y_t = W_{yh} \cdot h_t$$

$$\frac{\partial ReLU(x)}{\partial x} = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } < 0 \end{cases}$$



- 

- Let $\partial v = \dfrac{\partial E}{\partial v}$ where v any variable and E the loss. Then:

→ $$\partial h_t = \frac{\partial y_t}{\partial h_t} \cdot \partial y_t + \frac{\partial \hat{h}_{t+1}}{\partial h_t} \cdot \partial \hat{h}_{t+1} =$$
$$= W_{yh}^T \cdot \partial y_t + W_{hh}^T \cdot \partial \hat{h}_{t+1}$$

- → $$\partial \hat{h}_{t+1} = ReLU'(\hat{h}_{t+1}) \cdot \partial h_{t+1}$$

Due to $ReLU'(x)$, this term is 0 whenever $\hat{h}_{t+1} < 0$ and is 1. $\partial h_{t+1} = \partial h_{t+1}$ whenever $\hat{h}_{t+1} > 0$.

~~This is equivalent to:~~

$\cancel{\partial \hat{h}_{t+1} = ReLU'(\partial h_{t+1})}$

$\rightarrow \partial W_{hh} += \frac{\partial \hat{h}_{t+1}}{\partial W_{hh}} \cdot \partial \hat{h}_{t+1} = \partial \hat{h}_{t+1} \cdot h_t^T$

(which means $\partial W_{hh} += \partial \hat{h}_t \cdot h_{t-1}^T$ from one step further back in the graph)

The « += » is necessary because W_{hh} <u>is shared</u> through the graph.

$\rightarrow \partial W_{hx} += \frac{\partial \hat{h}_{t+1}}{\partial W_{hx}} \cdot \partial \hat{h}_{t+1} = \partial \hat{h}_{t+1} \cdot x_t^T$

$\rightarrow \partial b_h += \frac{\partial \hat{h}_{t+1}}{\partial b_h} \cdot \partial \hat{h}_{t+1} = \partial \hat{h}_{t+1}$

, which means $\partial b_h += \partial \hat{h}_t$ from one step further back in the graph.