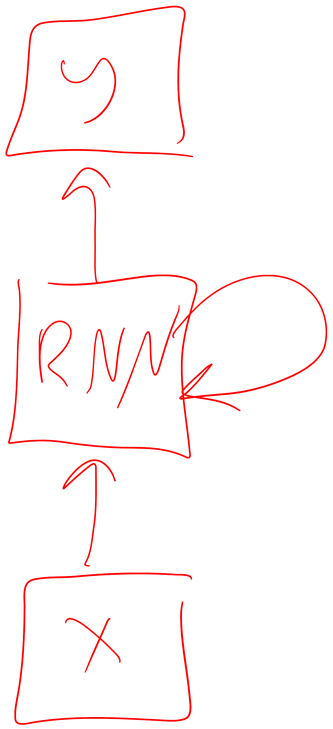
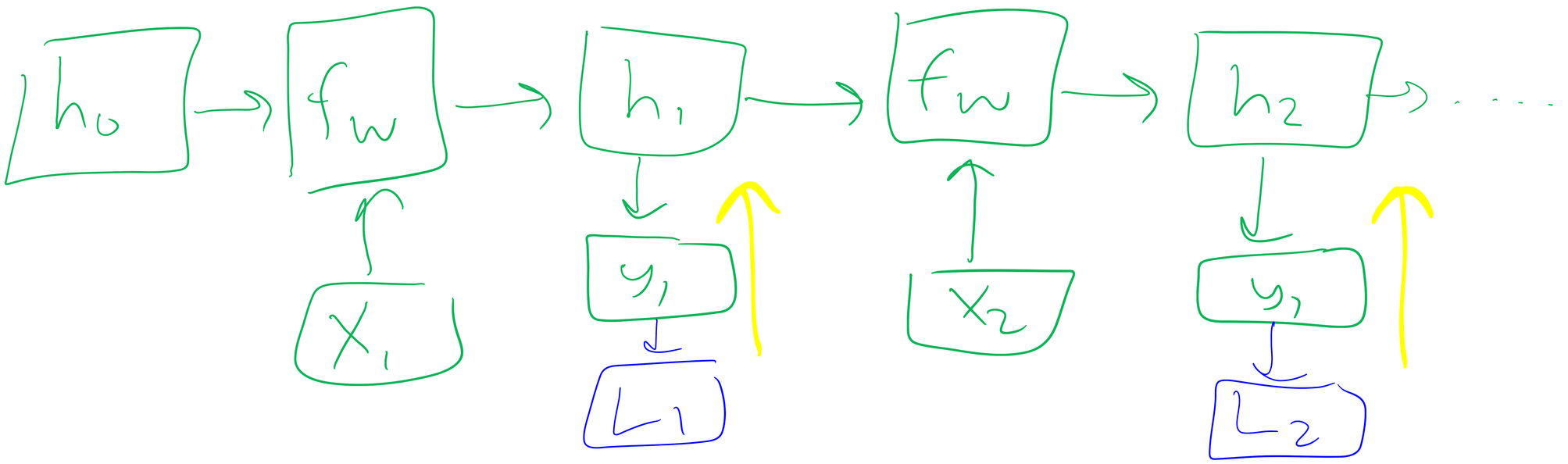



$$h_t = f_W(h_{t-1}, x_t)$$

new state — old state — input

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

$$y_t = W_{hy} h_t$$



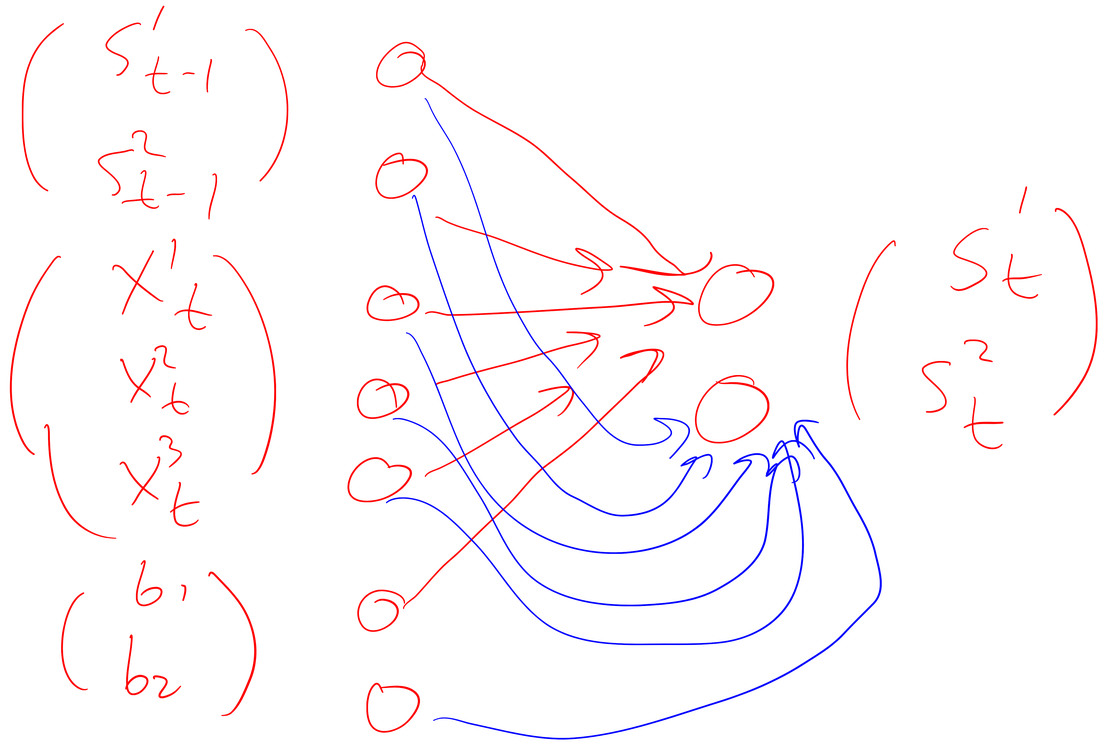
$$\boxed{S_t} = \phi\left(W S_{t-1} + U X_t + b\right)$$

$$\begin{pmatrix} s_t^1 \\ s_t^2 \end{pmatrix} = \phi\left\{ \underbrace{\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}}_{2\times 2} \begin{pmatrix} s_{t-1}^1 \\ s_{t-1}^2 \end{pmatrix} + \underbrace{\begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \end{pmatrix}}_{2\times 3} \underbrace{\begin{pmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \end{pmatrix}}_{3\times 1} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right\}$$

$$= \phi\left\{ \underbrace{\begin{pmatrix} w_{11} s_{t-1}^1 + w_{12} s_{t-1}^2 \\ w_{21} s_{t-1}^1 + w_{22} s_{t-1}^2 \end{pmatrix}}_{2\times 1} + \underbrace{\begin{pmatrix} \quad \end{pmatrix}}_{2\times 1} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right\}$$

$$= \phi\left\{ \underbrace{\left(\begin{array}{cc|ccc} w_{11} & w_{12} & u_{11} & u_{12} & u_{13} \\ w_{21} & w_{22} & u_{21} & u_{22} & u_{23} \end{array}\right)}_{2\times 5} \underbrace{\begin{pmatrix} s_{t-1}^1 \\ s_{t-1}^2 \\ \hline x_t^1 \\ x_t^2 \\ x_t^3 \end{pmatrix}}_{5\times 1} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right\}$$

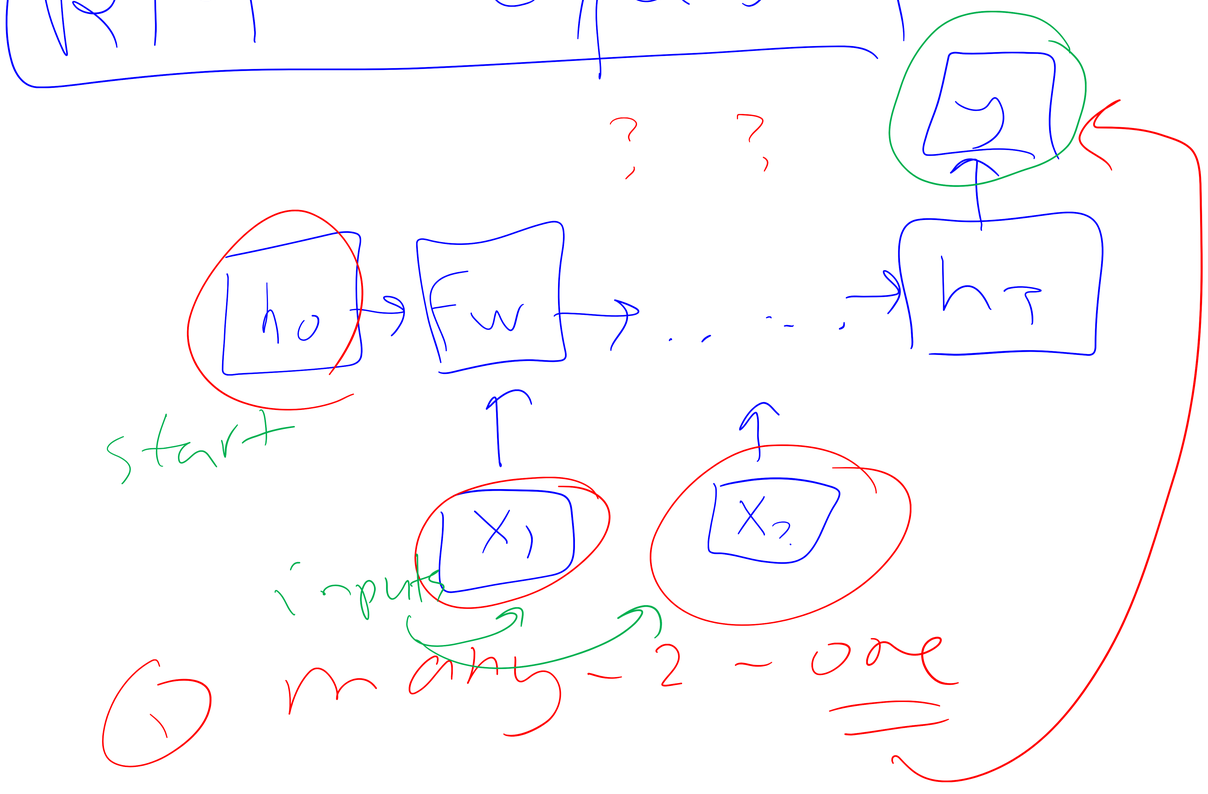
Sum



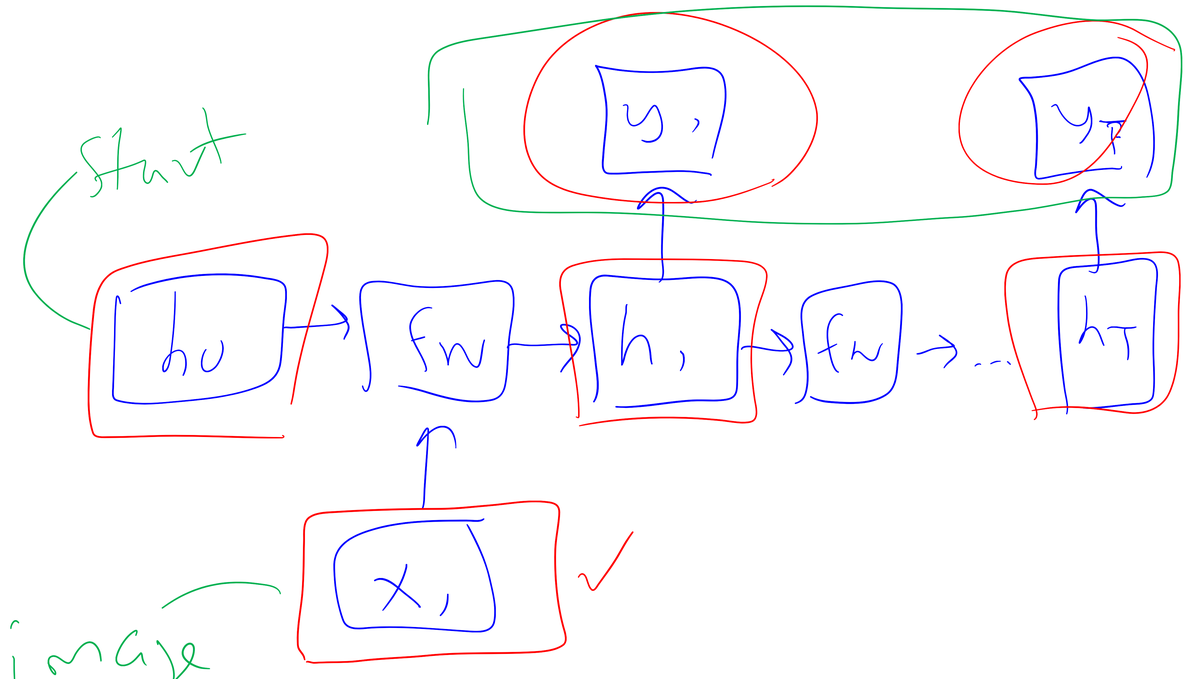

signal-2-text

Sentence

RNN types




① many-2-one

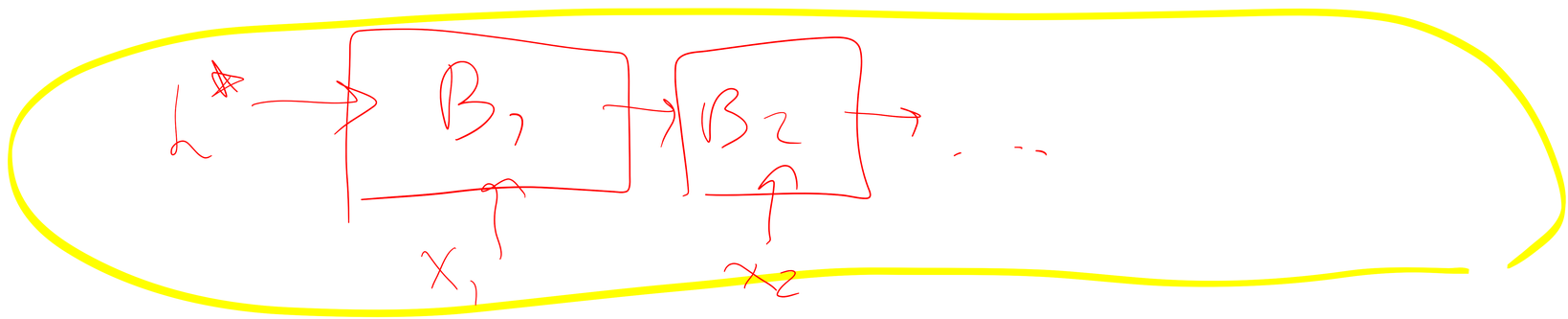


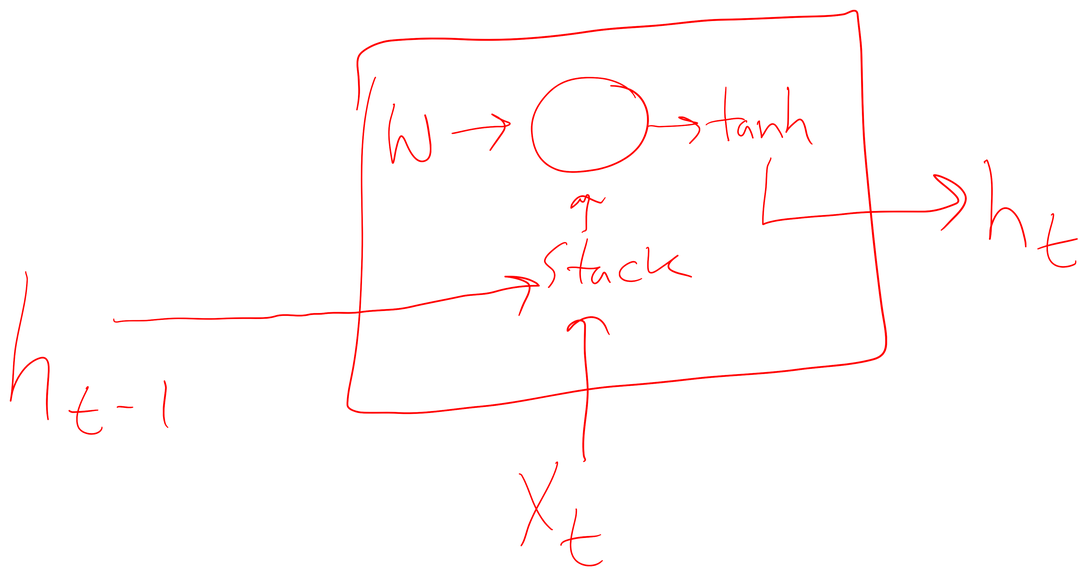

② one-2-many

(3) many-2-many (4) one-2-one

---

Block




$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

$$= \tanh\left((W_{hh}\ W_{xh})\begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$$

$$= \tanh\left(W\begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$$

Issues with back prop

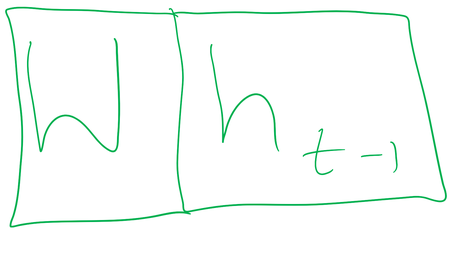
mul by 0's or mul > 1, ... vanishing and exploding gradients

# BPTT back prop through time

* simple, shared weights and unrolling

$$h_t = \tanh(Ux_t + Wh_{t-1})$$

$$\hat{y}_t = \text{softmax}(Vh_t)$$




$\hat{y}_t$ ↑

$h_t$ ↓

softmax ⇒

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

$\Sigma = 1$ and $[0,1]$

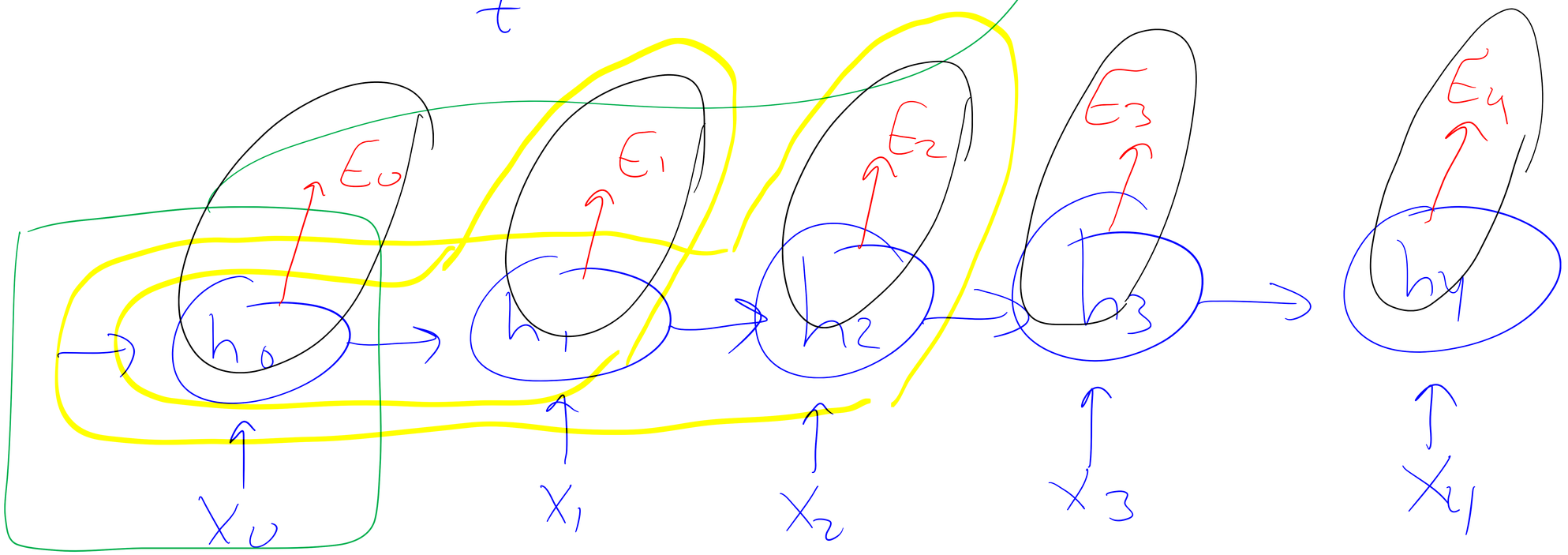
Error fx

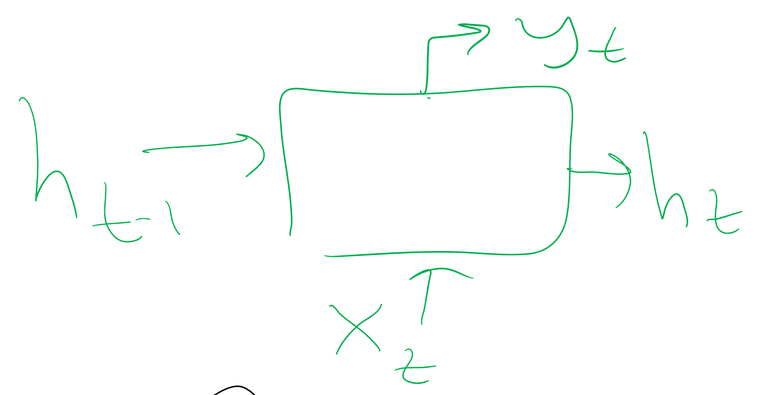
$$E_t(y_t, \hat{y}_t) = -y_t \ln \hat{y}_t$$

cross entropy

$$E(y, \hat{y}) = \sum_t E_t(y_t, \hat{y}_t)$$

$$= -\sum_t y_t \ln(\hat{y}_t)$$




Back prop

Question $\frac{\partial E}{\partial W} = \sum_t \frac{\partial E_t}{\partial W}$ longer....

Try $\frac{\partial E}{\partial V} = \sum_t \frac{\partial E_t}{\partial V}$ ✓

$$\frac{\partial E_3}{\partial \boxed{V}} = \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial V}$$

just depends on anx "current" values $\hat{y}_3, y_3, h_3$

$$= \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial z_3} \frac{\partial z_3}{\partial V}$$

$z_3 = V h_3$

since you know

post thur :-)

$$\frac{\partial z_3}{\partial V} = h_3$$

$$= (\hat{y}_3 - y_3) \otimes h_3$$

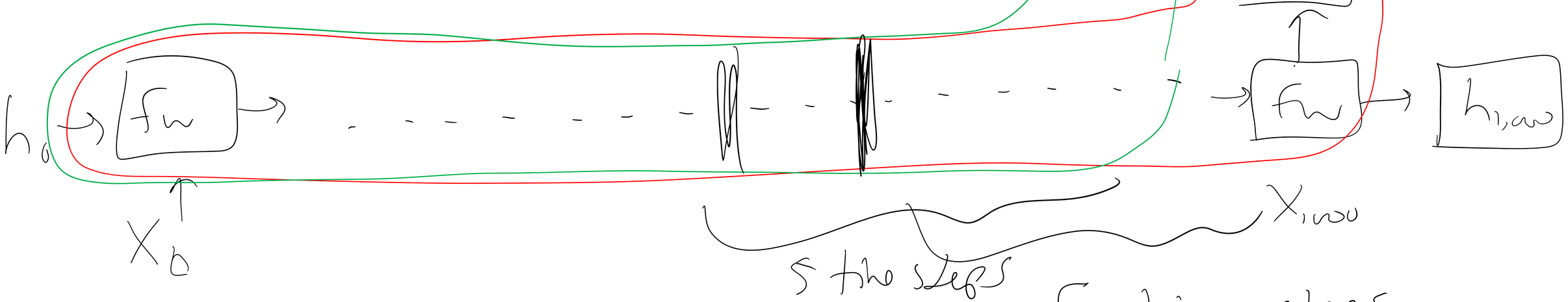
↑ outer product

$$\frac{\partial E_3}{\partial W} = \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial h_3} \frac{\partial h_3}{\partial W}$$

note: $h_3 = \tanh(U x_3 + W h_2)$

↑↑ depends on $h_2$, which depends on $h_1$

$$\frac{\partial E_3}{\partial W} = \sum_{k=0}^{3} \left(\frac{\partial E_3}{\partial \hat{y}_3}\frac{\partial \hat{y}_3}{\partial h_3}\right)\left(\frac{\partial h_3}{\partial h_k}\frac{\partial h_k}{\partial W}\right)$$




5 time steps

TBTT → truncated BTT

( fixed hunizon

|

approx....