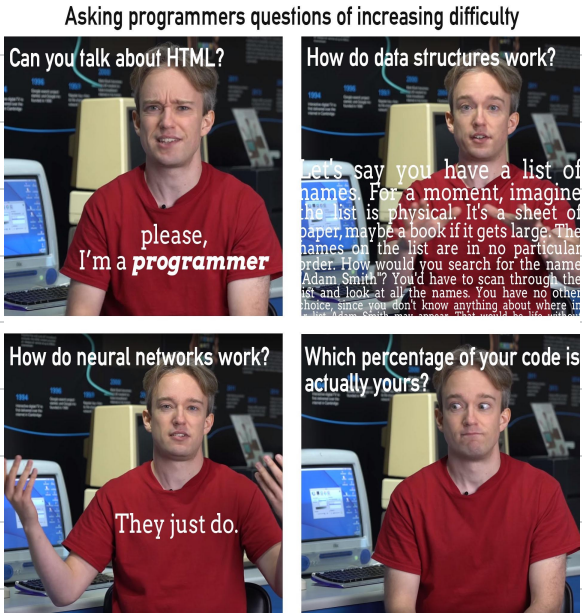
Session - 06

Recurrent Neural Networks - RNN

April 24, 2024



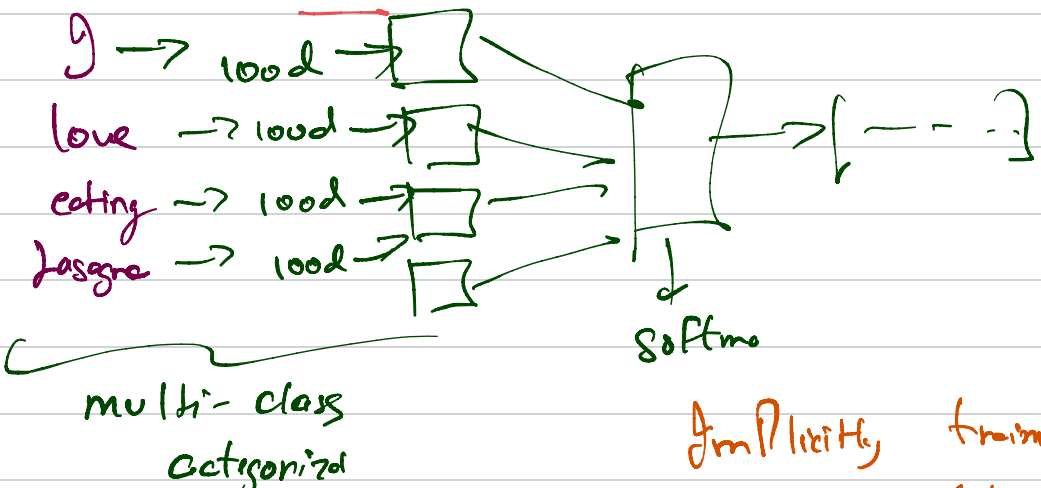
Agenda

1. RNN Architecture
2. Training a Many to One RNN model
3. Expanding training methods to other types of RNN
4. Code Implementation

MLP

(1) FFNN with Word2Vec ✓ context

no Seq



Implicitly training

model → 1st → 1st neuron

(1) large no of Para V.V. Inefficient

(2) length → Padding

<PAD>

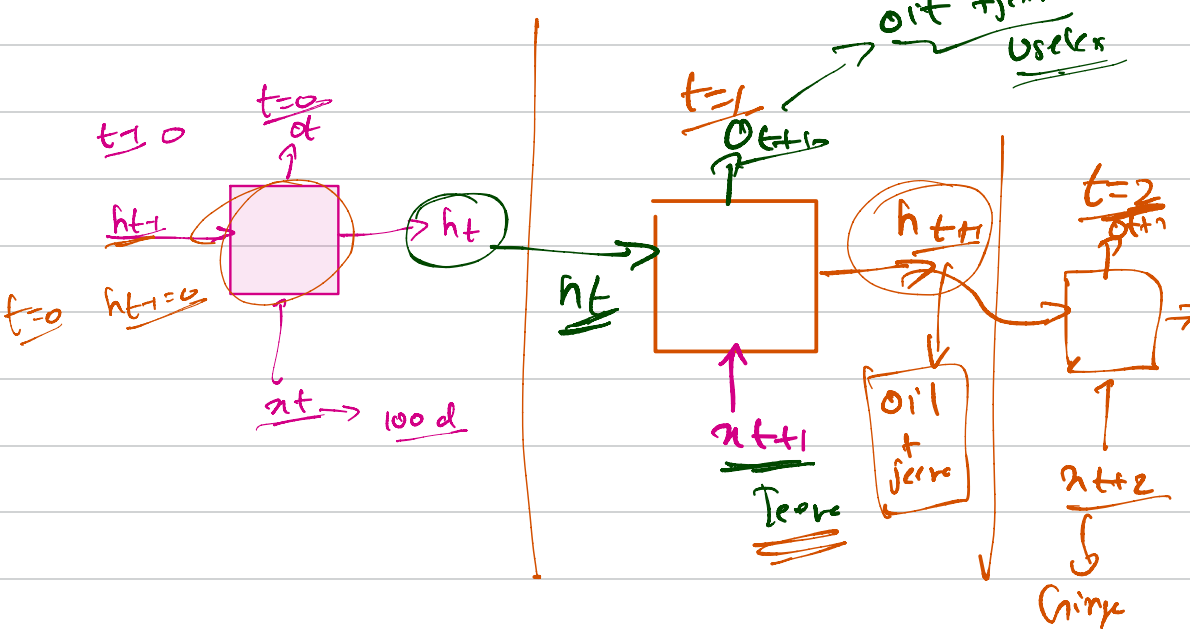
4 → 10 words

(3) All words are Present at once

Competition

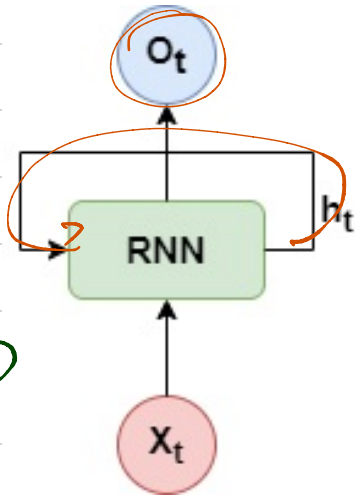
10 Players → one ingredient

RNN



9 like cats

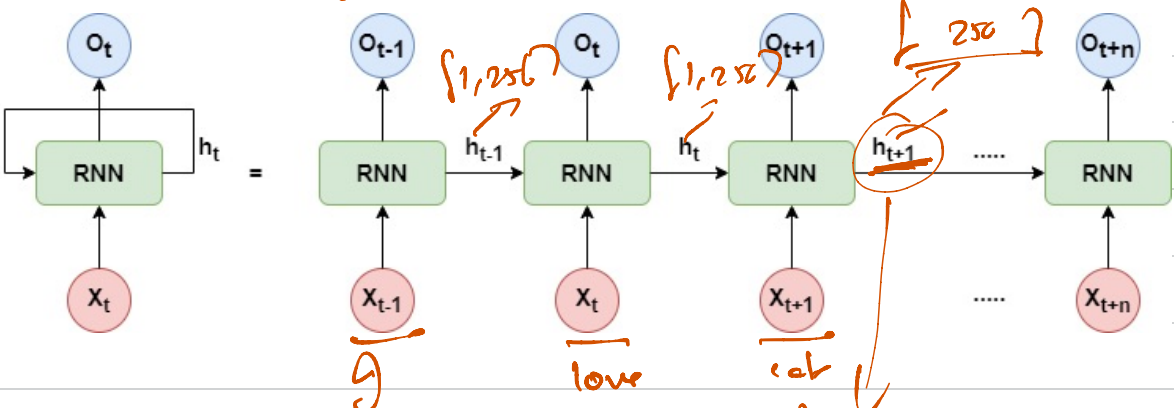
```
def minus(n)
    n = n - 1
    if n == 0:
        return 0
    else
        minus(n)
```



RNN
↓
Recurrent

f(n)
}
f(n)

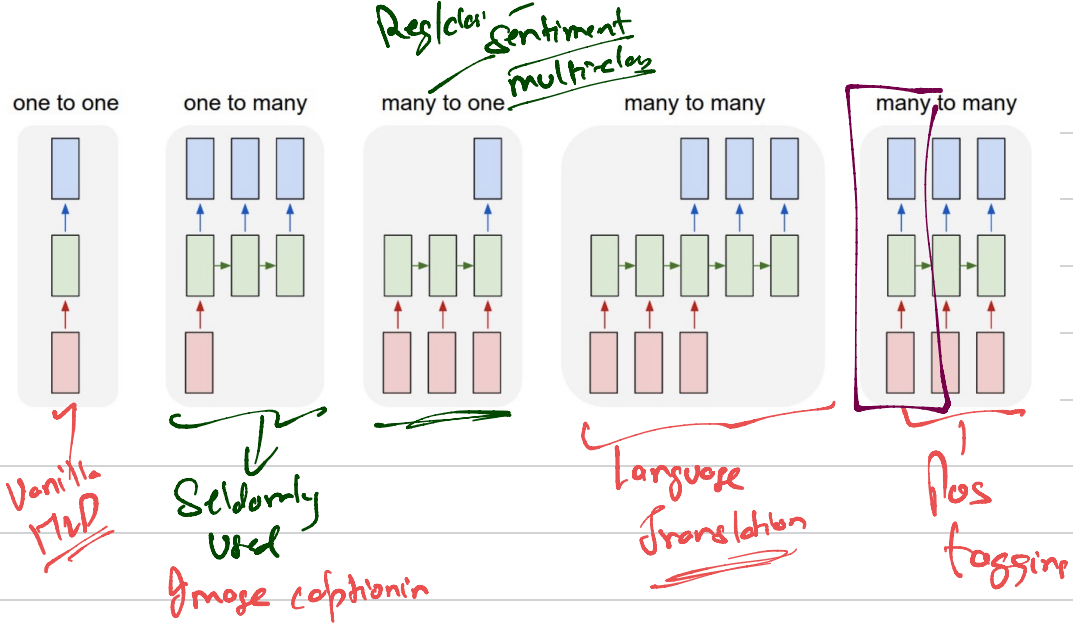
unrolling



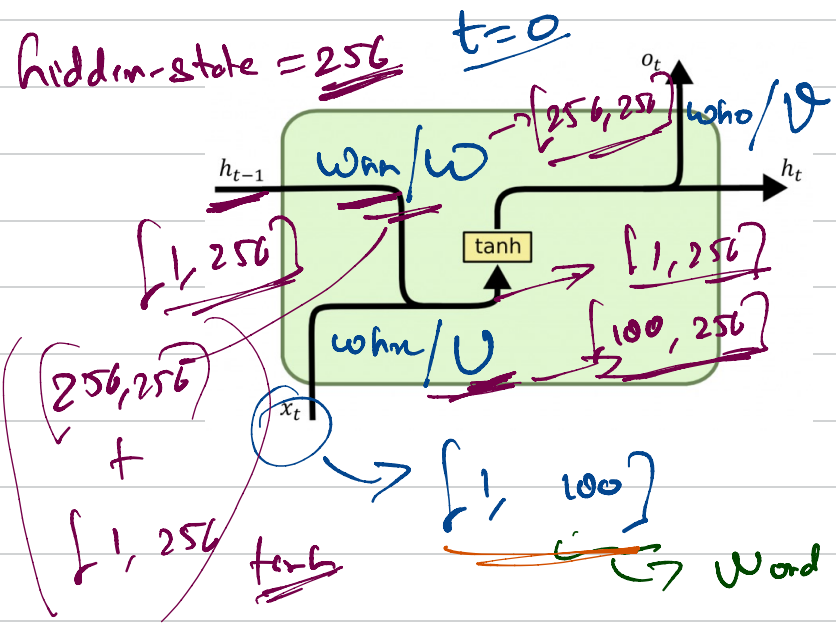
Includes inform. from previous time steps

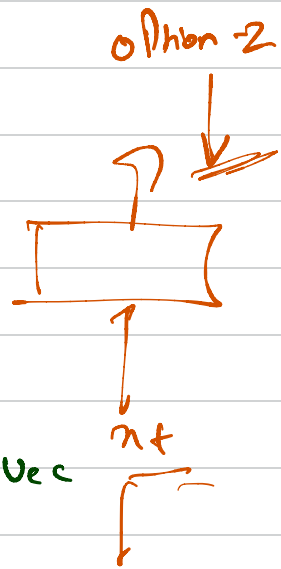
→ Any Sequential data

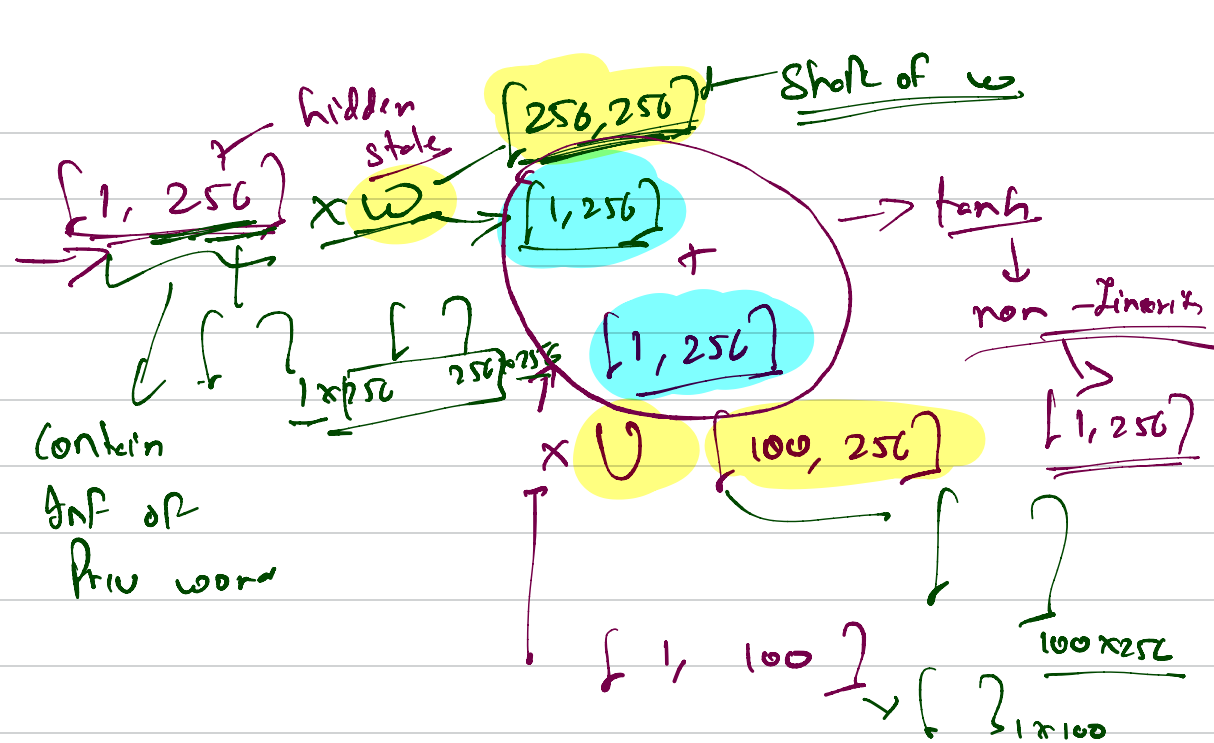
1. Time-series. Eg. Stock mark
2. Signal Eg. audio / speech
3. Text / Video



→ How RNN works internally







S - class

outflut from last timest

[1, 256] ⊙ v → [1, 5]

[256]

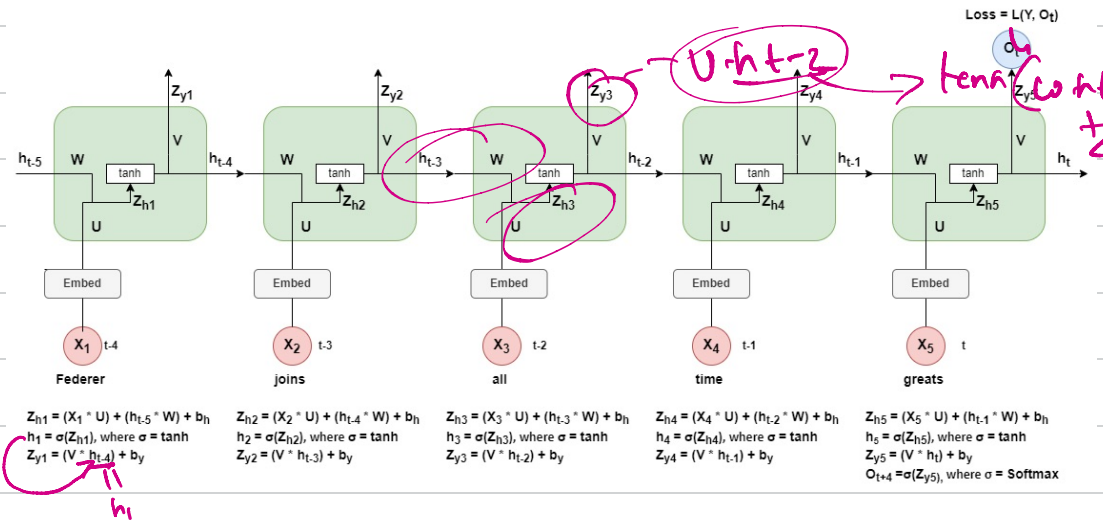
softmax

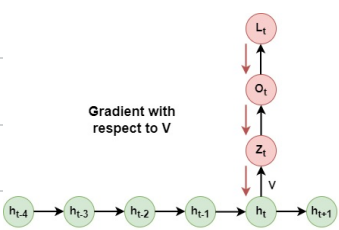
ht ∝ Amount of information you want to share across timestep

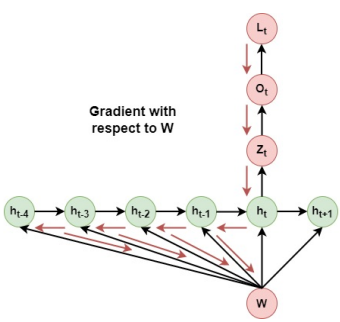
loss only w.r.t. last time

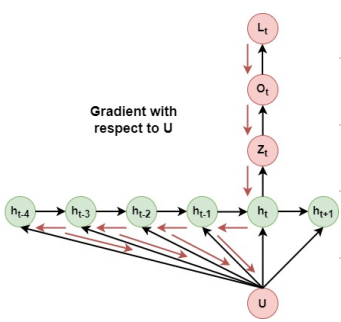
ht & how much model memorize

LETS TRAIN RNN









$$\frac{\partial h_t}{\partial W} = \frac{\partial h_t}{\partial W} + \frac{\partial h_t}{\partial h_{t-1}} * \frac{\partial h_{t-1}}{\partial W}$$

$$\frac{\partial h_{t-1}}{\partial W} = \frac{\partial h_{t-1}}{\partial W} + \frac{\partial h_{t-1}}{\partial h_{t-2}} * \frac{\partial h_{t-2}}{\partial W}$$

$$\frac{\partial h_{t-2}}{\partial W} = \frac{\partial h_{t-2}}{\partial W} + \frac{\partial h_{t-2}}{\partial h_{t-3}} * \frac{\partial h_{t-3}}{\partial W}$$

$$\frac{\partial h_{t-3}}{\partial W} = \frac{\partial h_{t-3}}{\partial W} + \frac{\partial h_{t-3}}{\partial h_{t-4}} * \frac{\partial h_{t-4}}{\partial W}$$

$$\frac{\partial h_{t-4}}{\partial W} = \frac{\partial h_{t-4}}{\partial W} + \frac{\partial h_{t-4}}{\partial h_{t-5}} * \frac{\partial h_{t-5}}{\partial W}$$

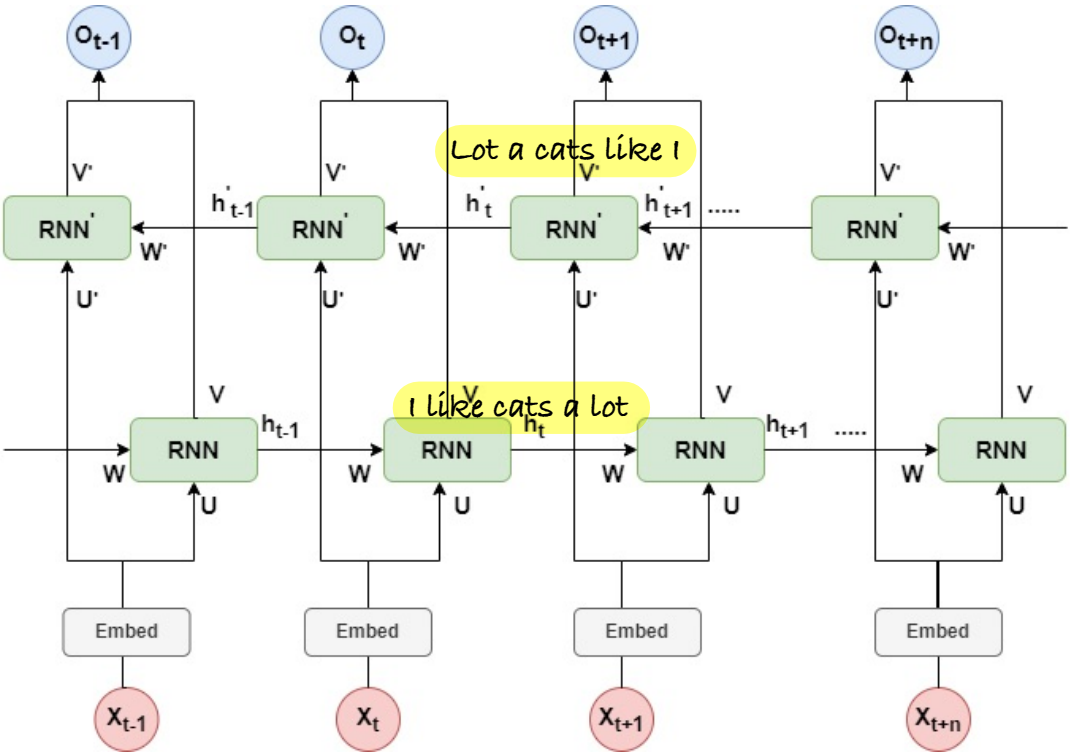
$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial h_t} \times \frac{\partial h_t}{\partial W} + \frac{\partial L}{\partial h_t} \times \frac{\partial h_t}{\partial h_{t-1}} \times \frac{\partial h_{t-1}}{\partial W}$$

$\frac{\partial L}{\partial h_{t-5}} \times \frac{\partial h_{t-5}}{\partial w}$

$P \rightarrow \frac{\partial G_1}{} \; -- \; \epsilon_R$

Back-Propagation through time

Before I went to bed, I turned of the _____, It was very loud and I Needed to fall asleep.



Bidirectional RNN processes sequences in both forward and backward directions.

$z = [0.1, 0.2, 0.3]$

softmax = $[0, 1]$

$$\left[\frac{e^{0.1}}{e^{0.1}+e^{0.2}+e^{0.3}} \;/\; \frac{e^{0.2}}{e^{0.1}+e^{0.2}+e^{0.3}} \;/\; \frac{e^{0.3}}{e^{0.1}+e^{0.2}+e^{0.3}}\right]$$

$h_t \rightarrow (w \times h_{t-1} + U \times x_t)$ $\quad \frac{1}{1+e^{-z}}$

$h_{t-1} \quad (w \times h_{t-2}) + (\ldots$

$h_{t-2} \quad (w \times h_{t-3}) \ldots$

$\downarrow$

0, 1

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial h_t} \times \frac{\partial h_t}{\partial w} + \frac{\partial L}{h_{t-1}} \times \frac{\partial h_{t-1}}{\partial w} + \frac{\partial L}{\partial h_{t-2}} \times \frac{\partial h_{t-2}}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial w}$$

Last time step

$$\frac{\partial L}{\partial h_t} \times \frac{\partial h_t}{\partial h_{t-1}} \times \frac{\partial h_{t-1}}{\partial w}$$

U

$$\frac{\partial h_t}{\partial W} = \frac{\partial h_t}{\partial W} + \frac{\partial h_t}{\partial h_{t-1}} * \frac{\partial h_{t-1}}{\partial W}$$

$$\frac{\partial h_{t-1}}{\partial W} = \frac{\partial h_{t-1}}{\partial W} + \frac{\partial h_{t-1}}{\partial h_{t-2}} * \frac{\partial h_{t-2}}{\partial W}$$

$$\frac{\partial h_{t-2}}{\partial W} = \frac{\partial h_{t-2}}{\partial W} + \frac{\partial h_{t-2}}{\partial h_{t-3}} * \frac{\partial h_{t-3}}{\partial W}$$

$$\frac{\partial h_{t-3}}{\partial W} = \frac{\partial h_{t-3}}{\partial W} + \frac{\partial h_{t-3}}{\partial h_{t-4}} * \frac{\partial h_{t-4}}{\partial W}$$

$$\frac{\partial h_{t-4}}{\partial W} = \frac{\partial h_{t-4}}{\partial W} + \frac{\partial h_{t-4}}{\partial h_{t-5}} * \frac{\partial h_{t-5}}{\partial W}$$

going all together