Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:03

RNNs :

f()

{ :

f()

}

Sequence models

(DL)

1/1

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:07

obj: sequence of inputs (RNN)

$w_3 \quad w_4 \quad w_6 \quad w_{12} \cdots$

2 / 5

Task: sequence-aware ← m-words → w_k → Clasfn

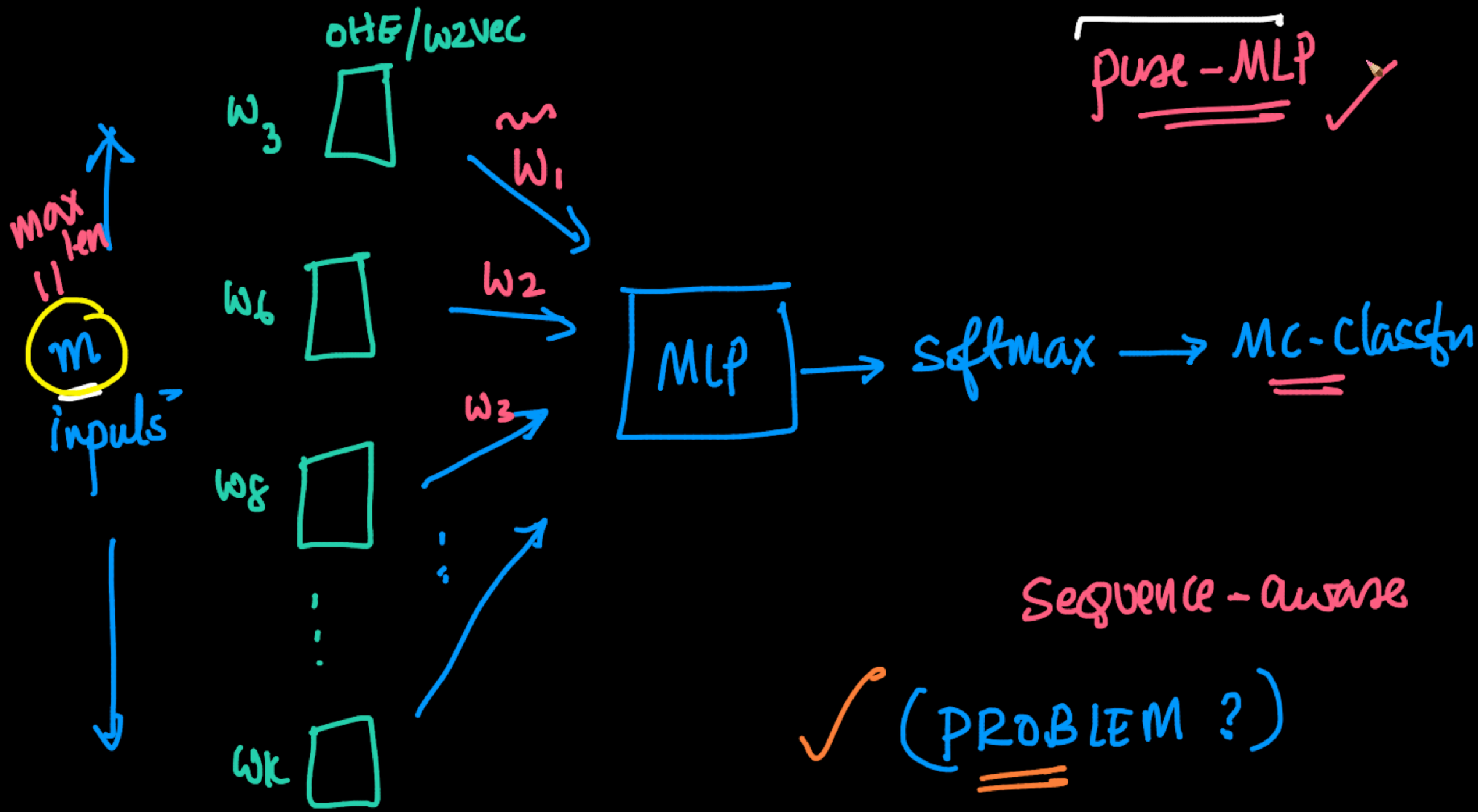
doc: w_3 w_6 w_8 w_4 ....

✓ ① TFIDF-weighted w2Vec (DL) → sequence agnostic

② HINT: MLP (?)

3 / 6

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:20



4 / 6

Problem:- input dim is large (minor)

all docs should be of same # words

→ d_1 → 10 words → C_1 (major)

d_2 → 100 words → C_2

d_3 → 5 words → C_1

d_4 → 1000 words → C_3

Soln: max-len = 1000

sparsity ✓

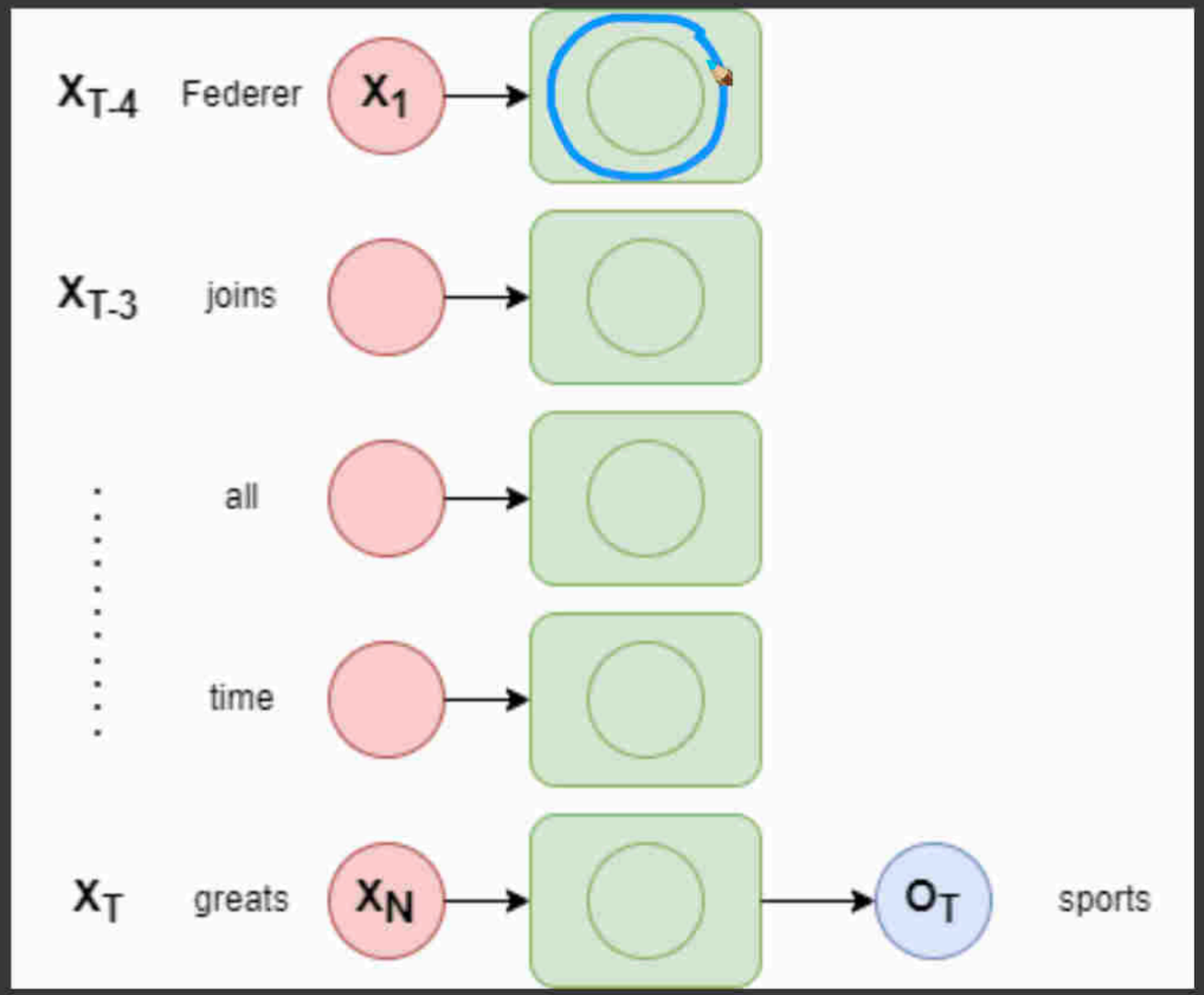
Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:22

$$\text{doc}: \quad w_1 \quad w_2 \quad \cdots \quad w_{10} \quad \overbrace{\underbrace{\langle BLK \rangle \quad \langle BLK \rangle \quad \cdots}_{\text{padding with dummy symbols}}}^{990}$$

Unnecessary overhead

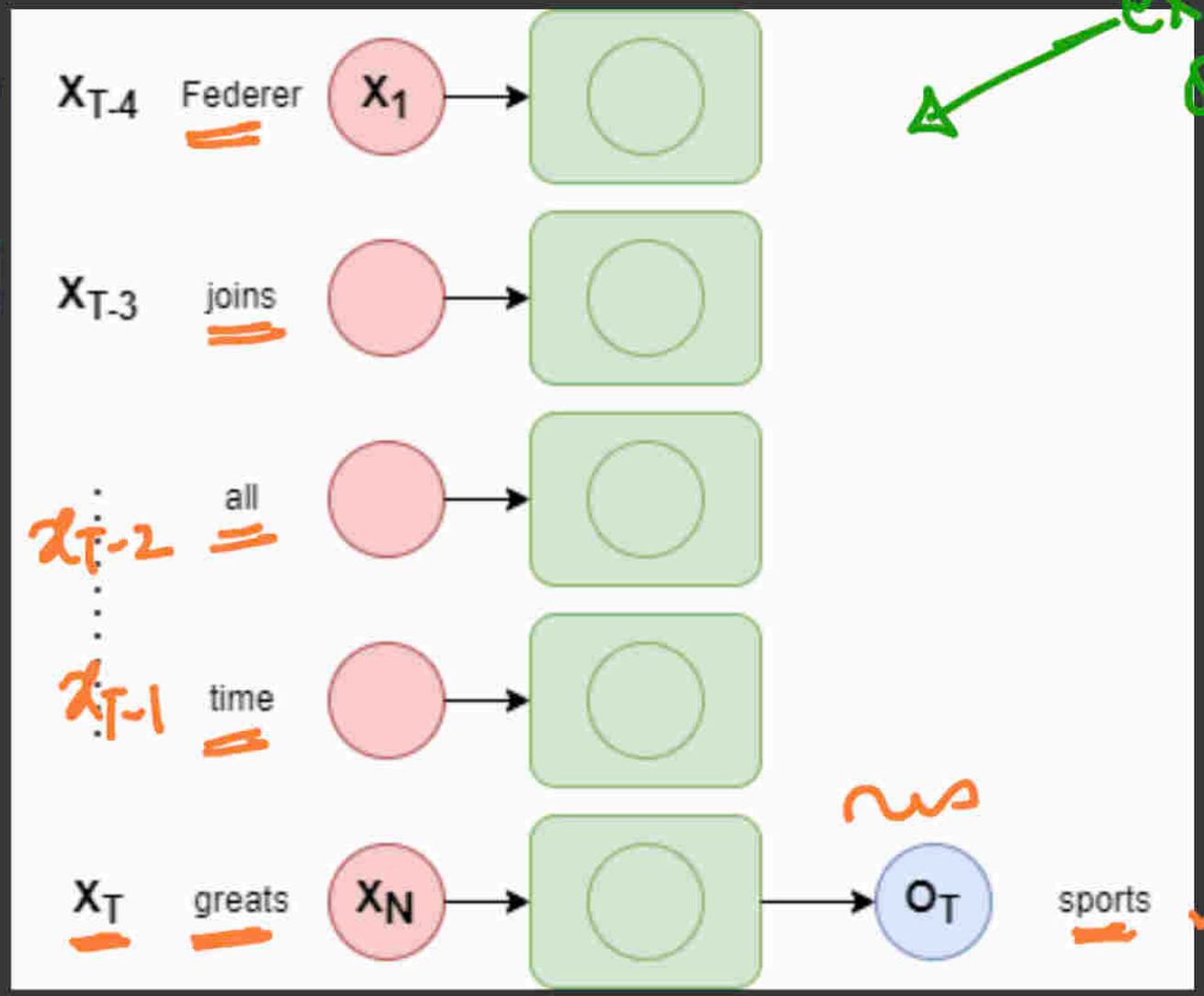
6 / 6

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:25
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



7 / 7

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:27
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



extension of MLP

5 words

50 words

500 words

8 / 8

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:28
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

words spoken/written

Let's look at a simple example where you will have to fill in the last word

Fill the last word

Scenarion 1: _ _ _ _ _

Scenarion 2: Sachin Tendulkar is a _

Answer: Sachin Tendulkar is a cricketer

In Scenario 1, without having access to the previous words it is literally impossible to fill the blank as **cricketer**. However in scenario 2 it's easy to fill because we know the context of what are the previous words.

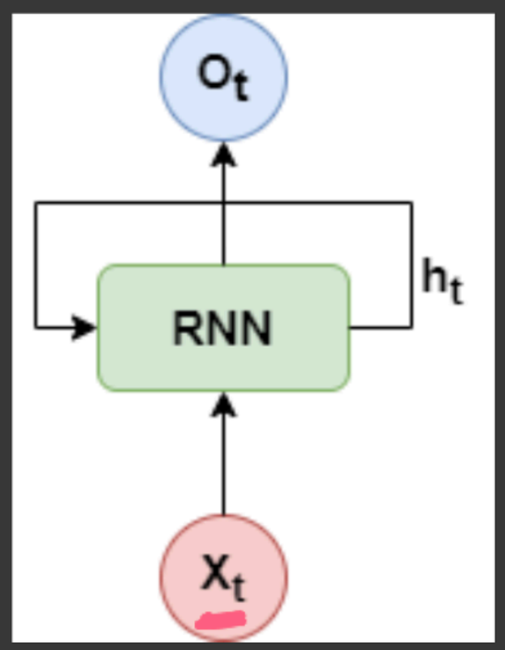
- **No connection from previous input words**
  As we pass one word at a time step **t**, the final output "**sports**" is not only a function of the last word "**greats**" passed at time **t**. But it also depends on the other words "**Federer joins all time**" as well which are passed from time steps $\mathbf{t_{n-6}}$ $\mathbf{t_{n-1}}$

9 / 9

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:30
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

along with the inputs from previous time steps

$$O_t = f(X_t, h_{t-1})$$



x_t : word → OHE / w2Vec

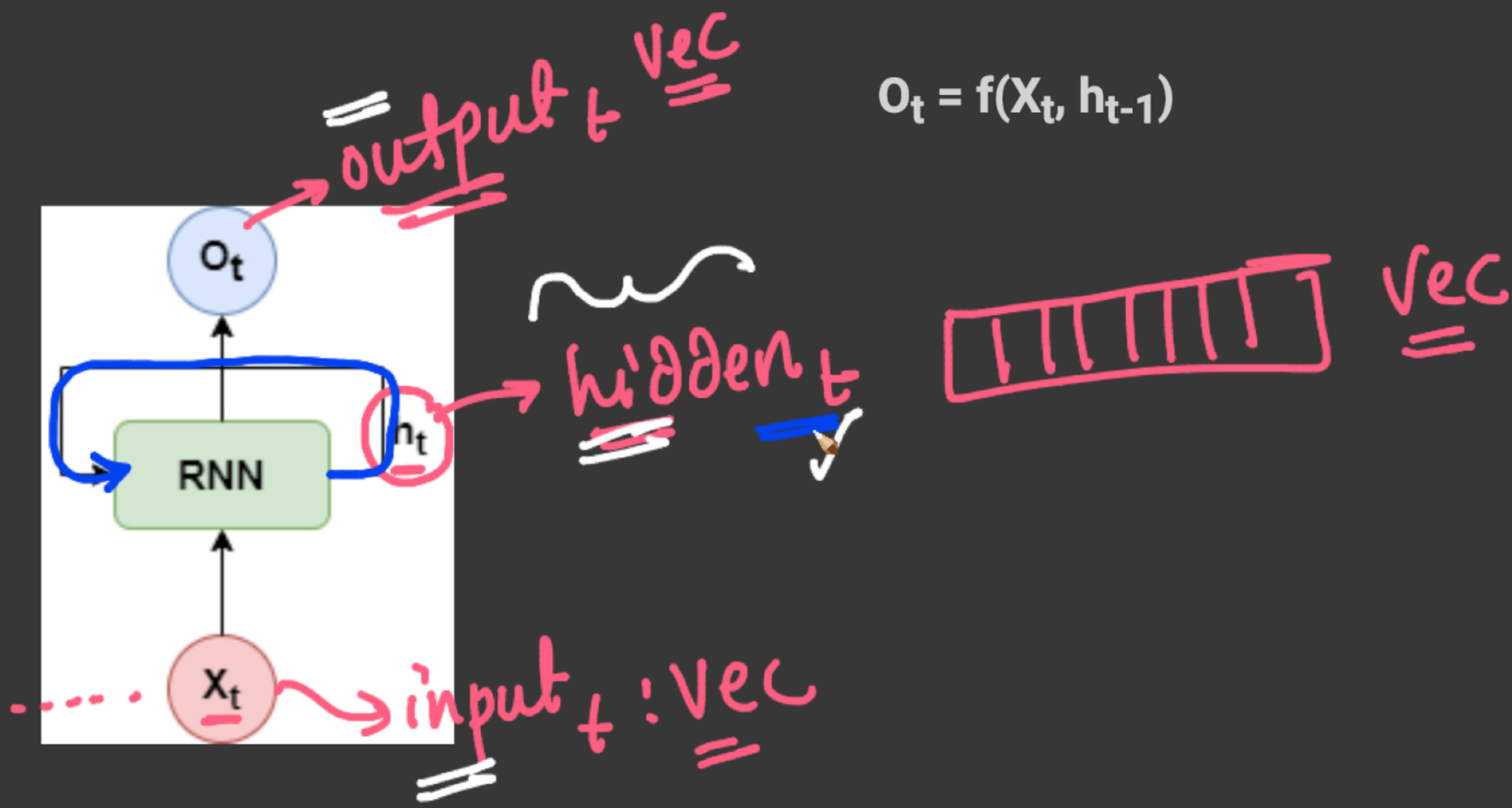
What are RNNs?

10 / 10

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:33
Vanilla-RNN_V1.ipynb - Colabo 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect

along with the inputs from previous time steps

$$O_t = f(X_t, h_{t-1})$$

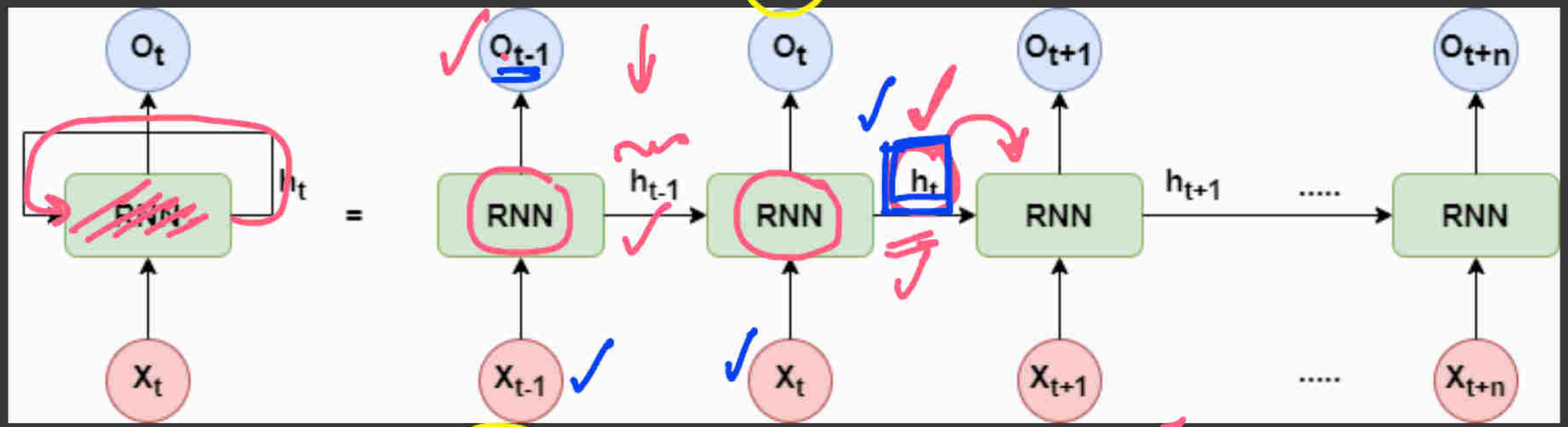


What are RNNs?

11 / 11

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:39
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect

RNN unravelled over time



What are the different types of RNN?

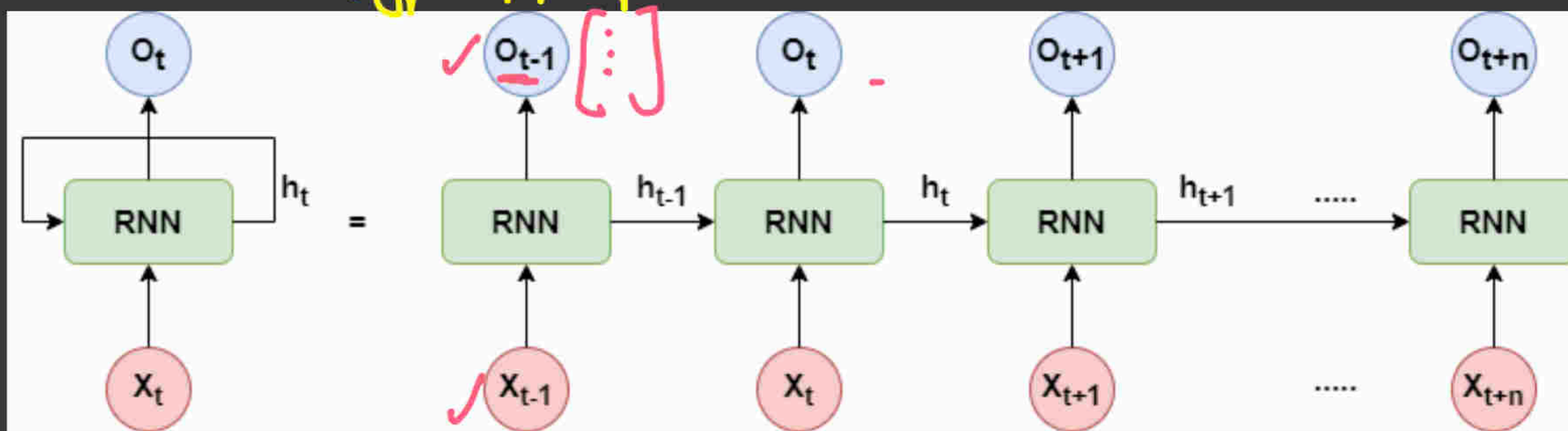
Types of input & output

12 / 12

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:40
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect

Task: Pos for each word

Adj/ N/V/Adv/



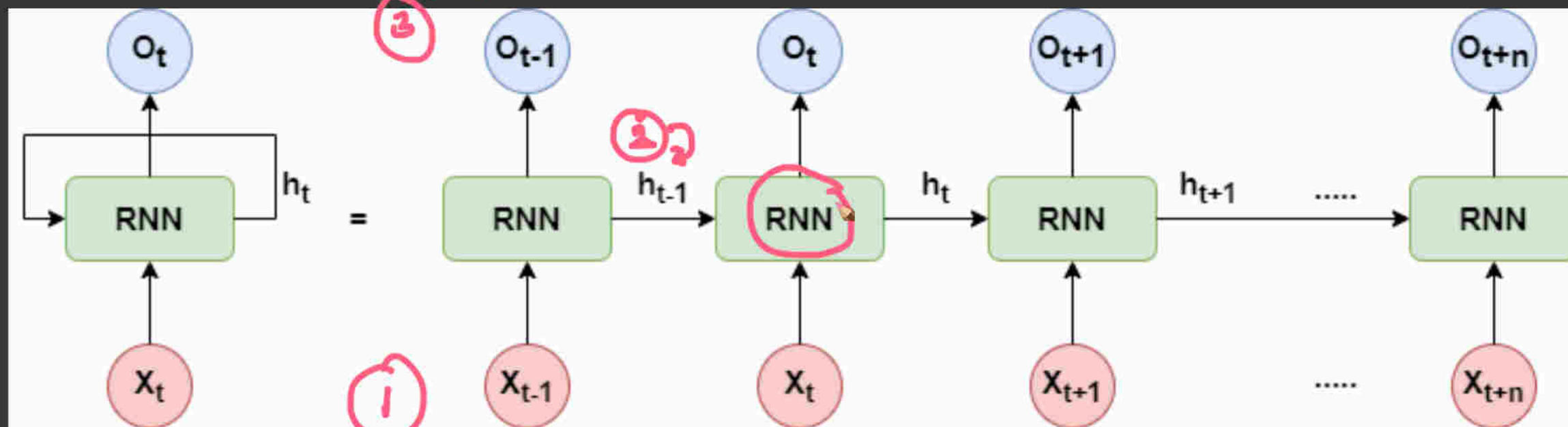
What are the different types of RNN?

Types of input & output

13 / 13

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:41
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect

RNN unravelled over time



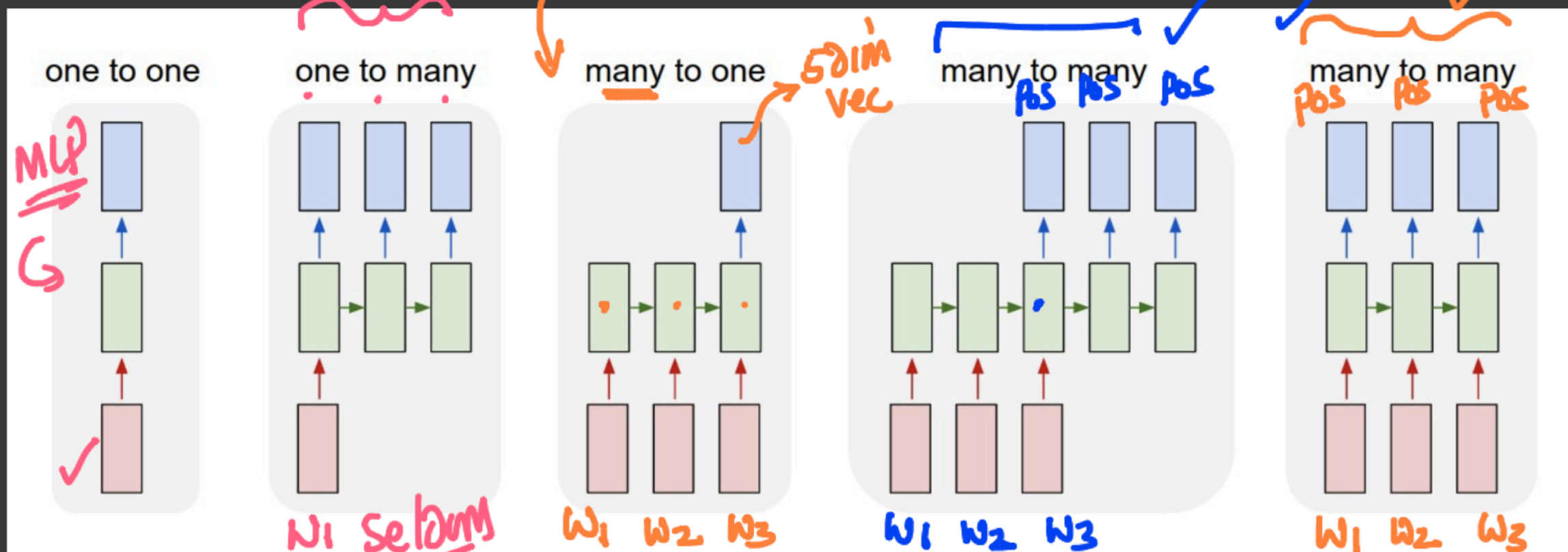
What are the different types of RNN?

Types of input & output

14 / 14

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:46
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

Task: sentence -> Class

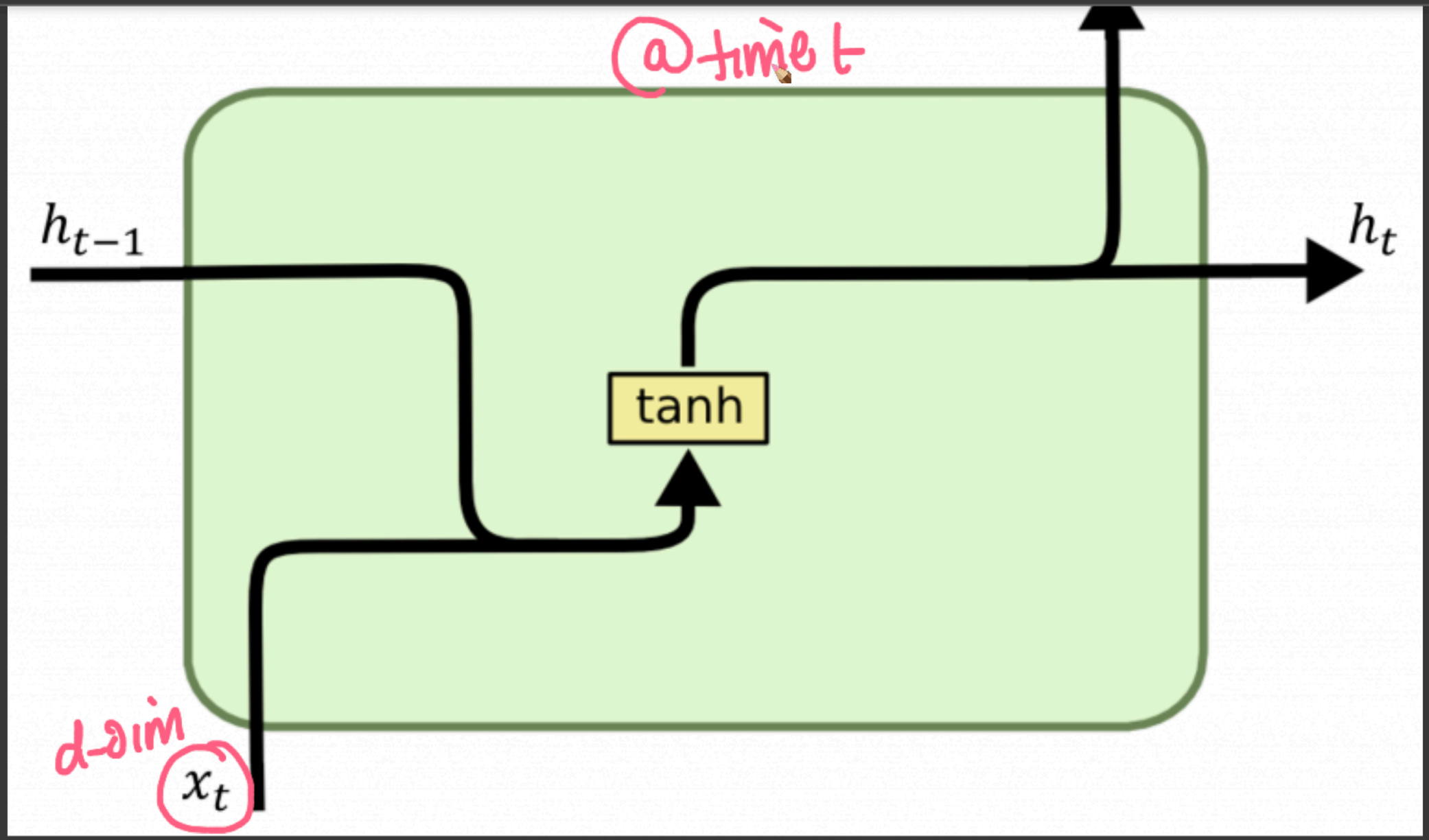


One to One: $T_x=1$, $T_y=1$

Example: Normal Feedforward Neural Network

15 / 15

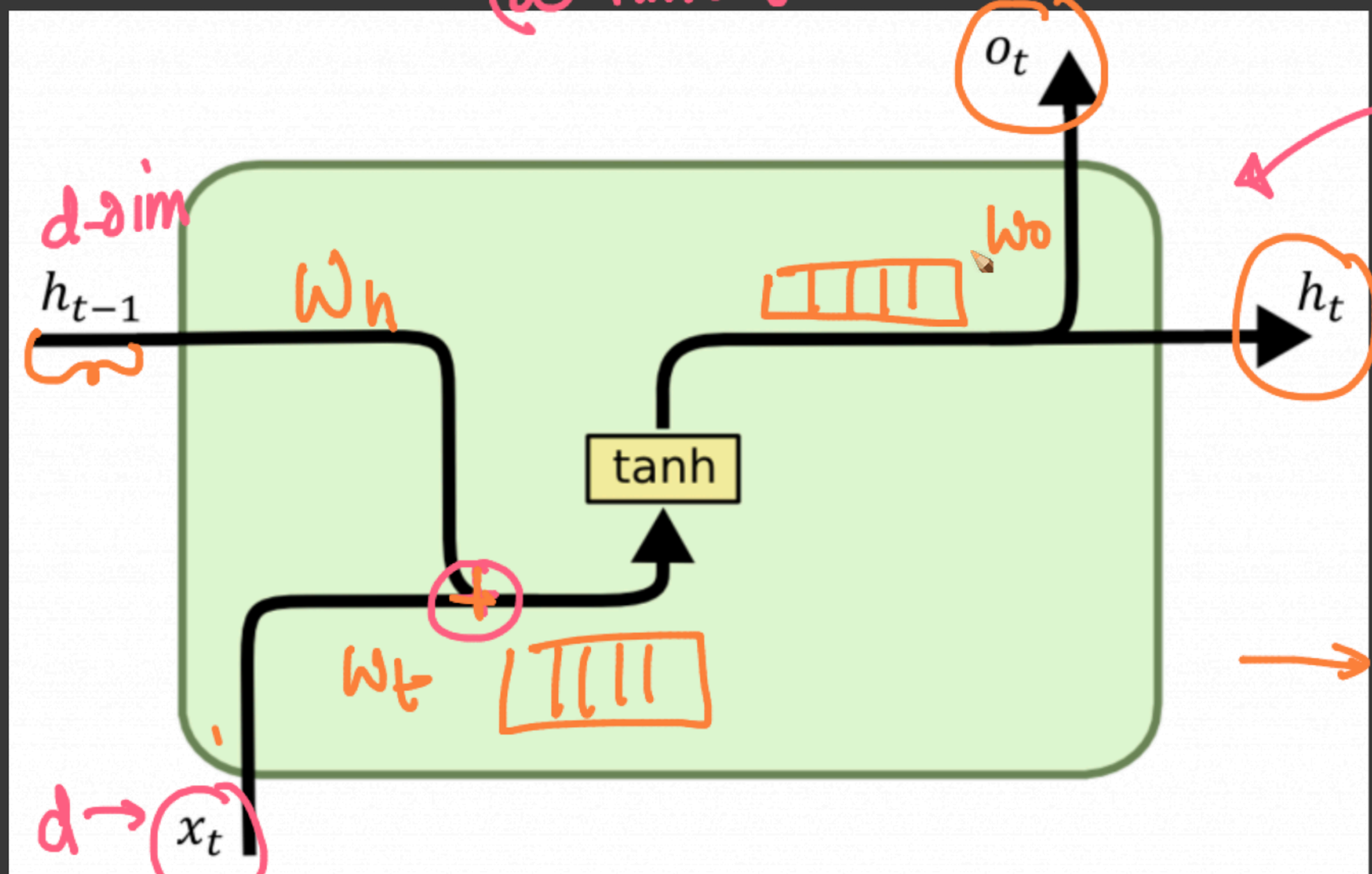
Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:48
Vanilla-RNN_V1.ipynb - Colabo 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



16 / 16

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:52
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

@time t



d-dim

W_h

W_o

W_t

d → x_t

1990's

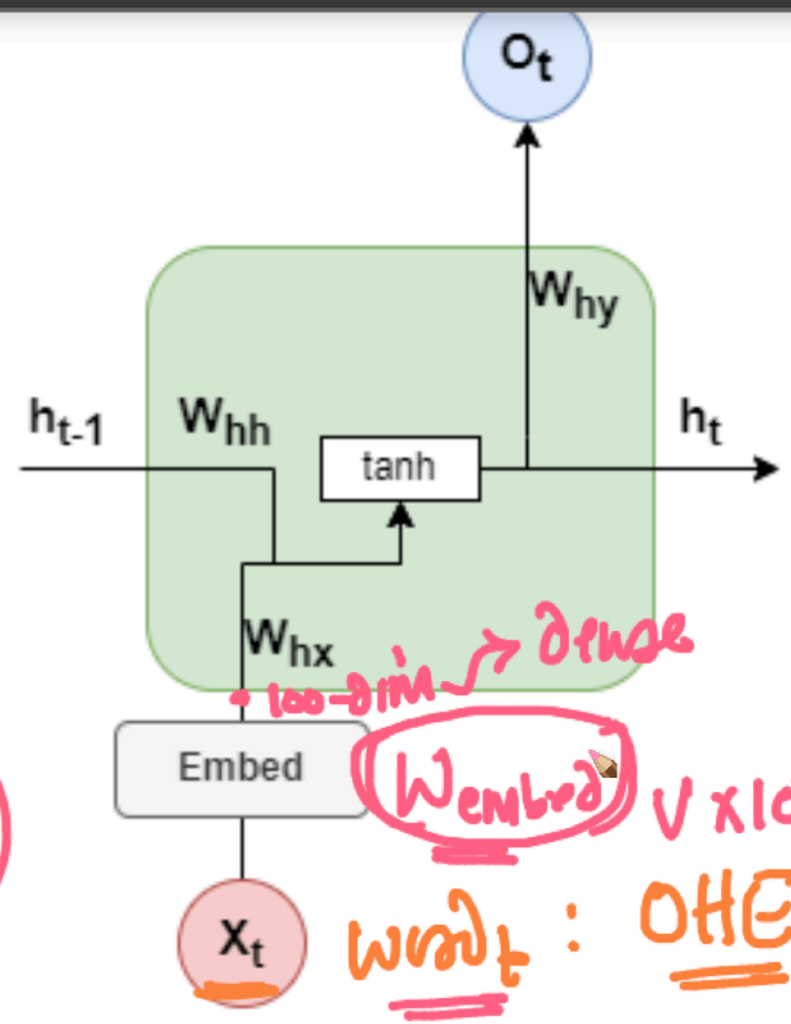
Simple RNN (2015 – ....)

more complex

LSTM
GRU

17 / 17

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:57
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



① W2Vec : 512dim

② embedding (FC - matrix mul)

100-dim → dense

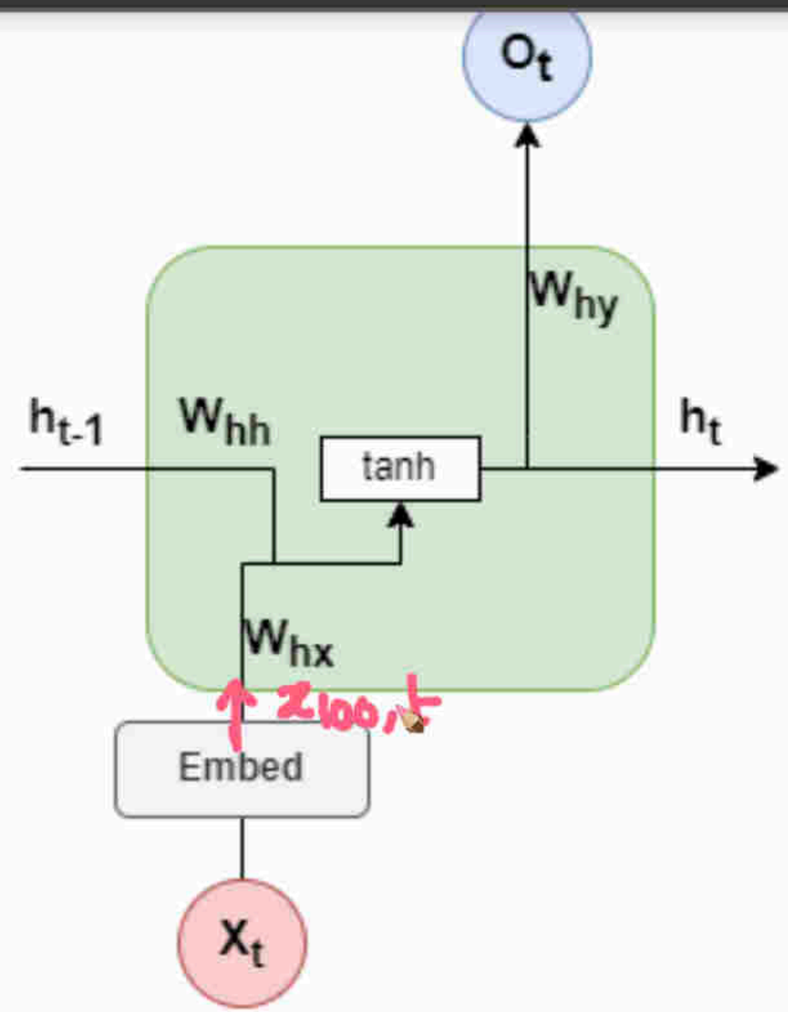
FC-layer

$V \times 100$

word$_t$: OHE $\in \mathbb{R}^{V}$, $V = 100{,}000$ (sparse)

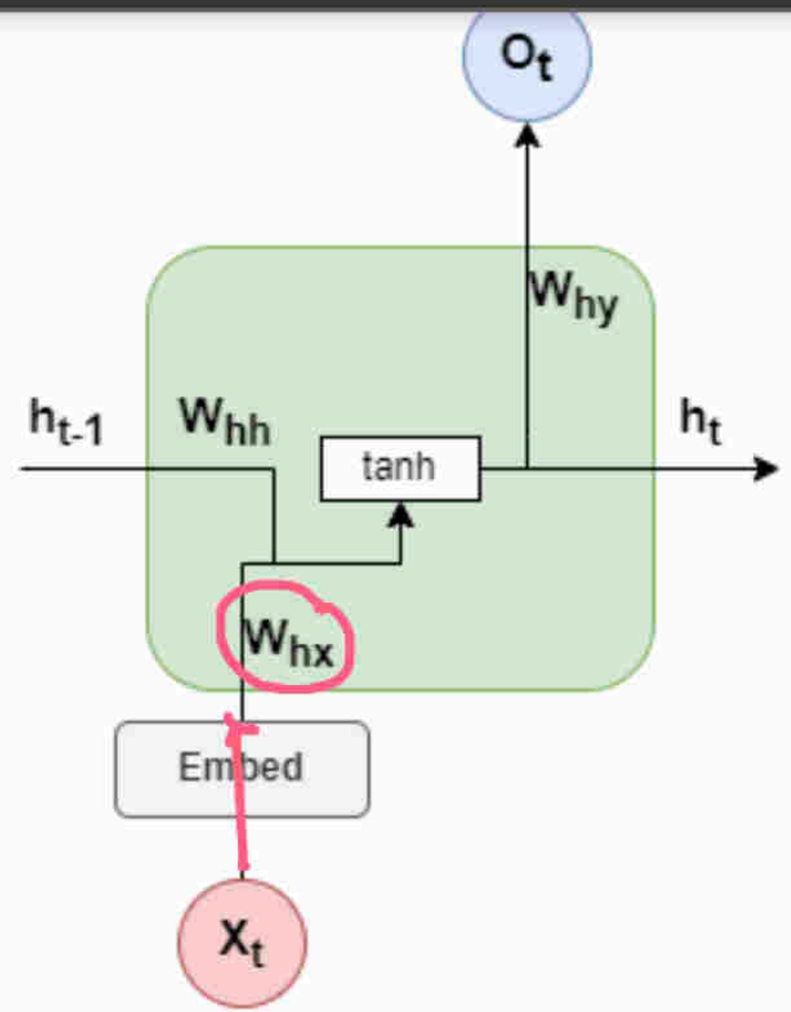
18 / 18

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:58
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect



19 / 19

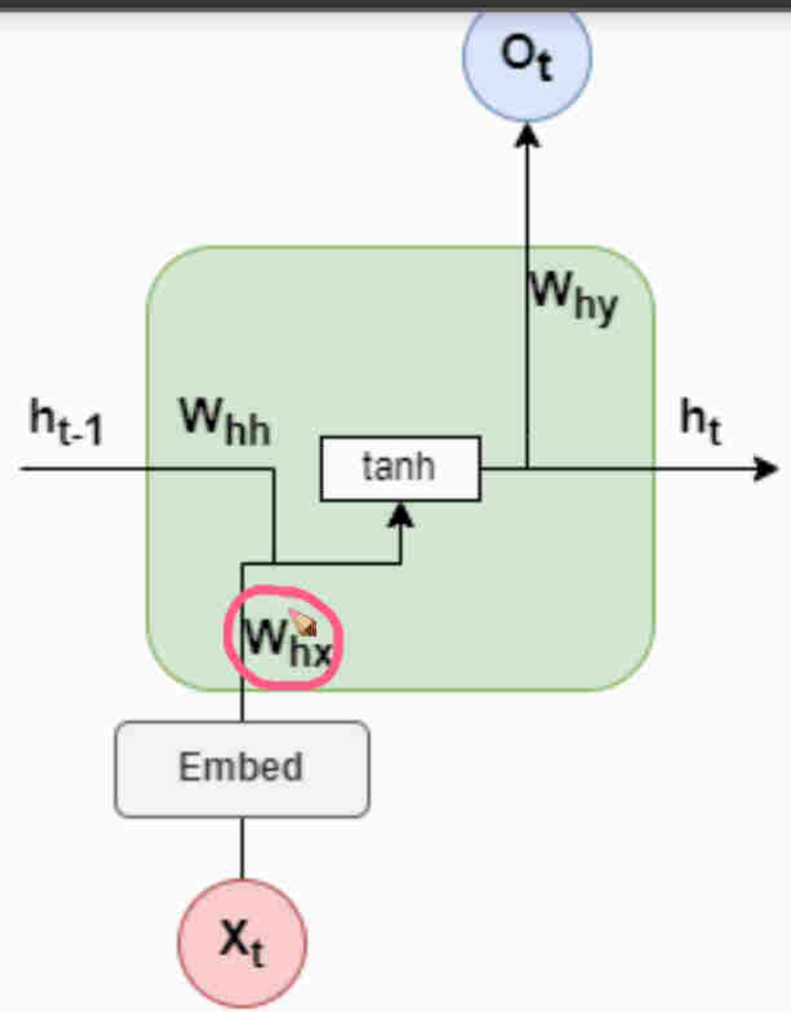
Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:58
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect



h: hidden
x: input
y: output

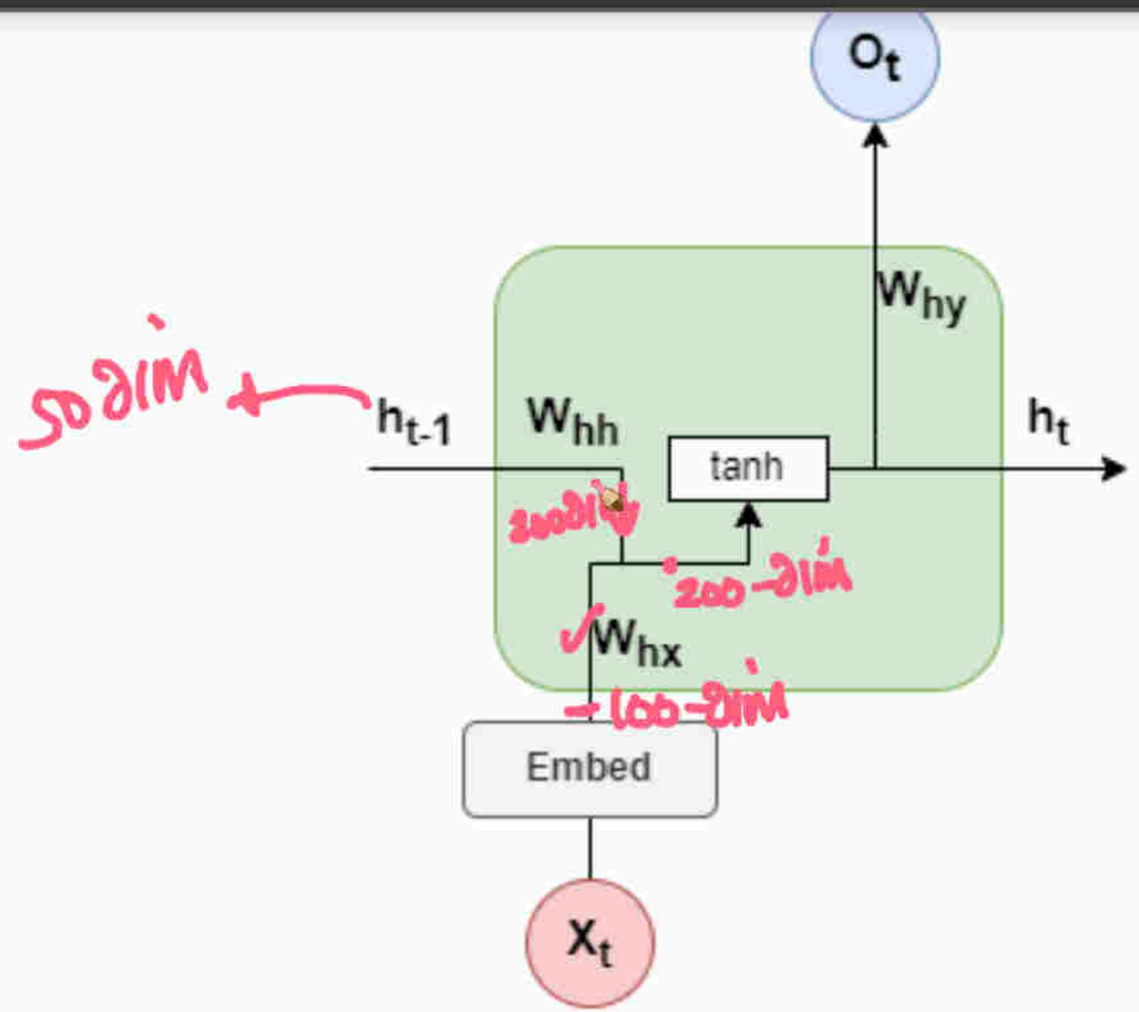
20 / 20

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:58
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



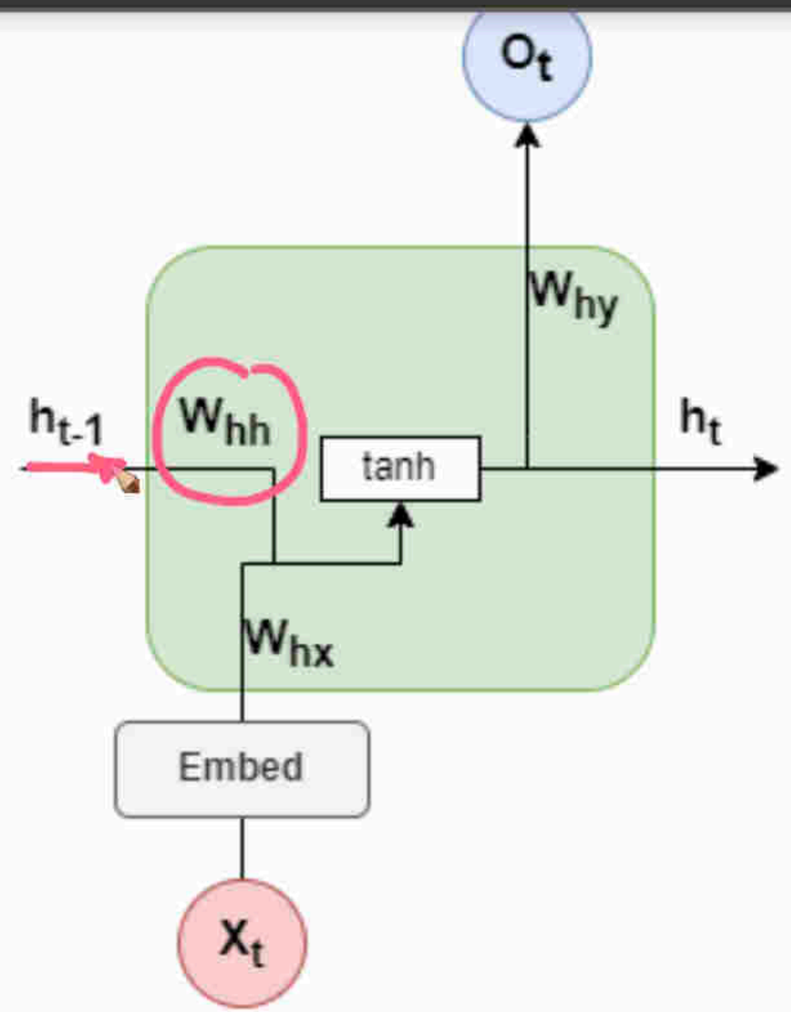
21 / 21

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:59
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



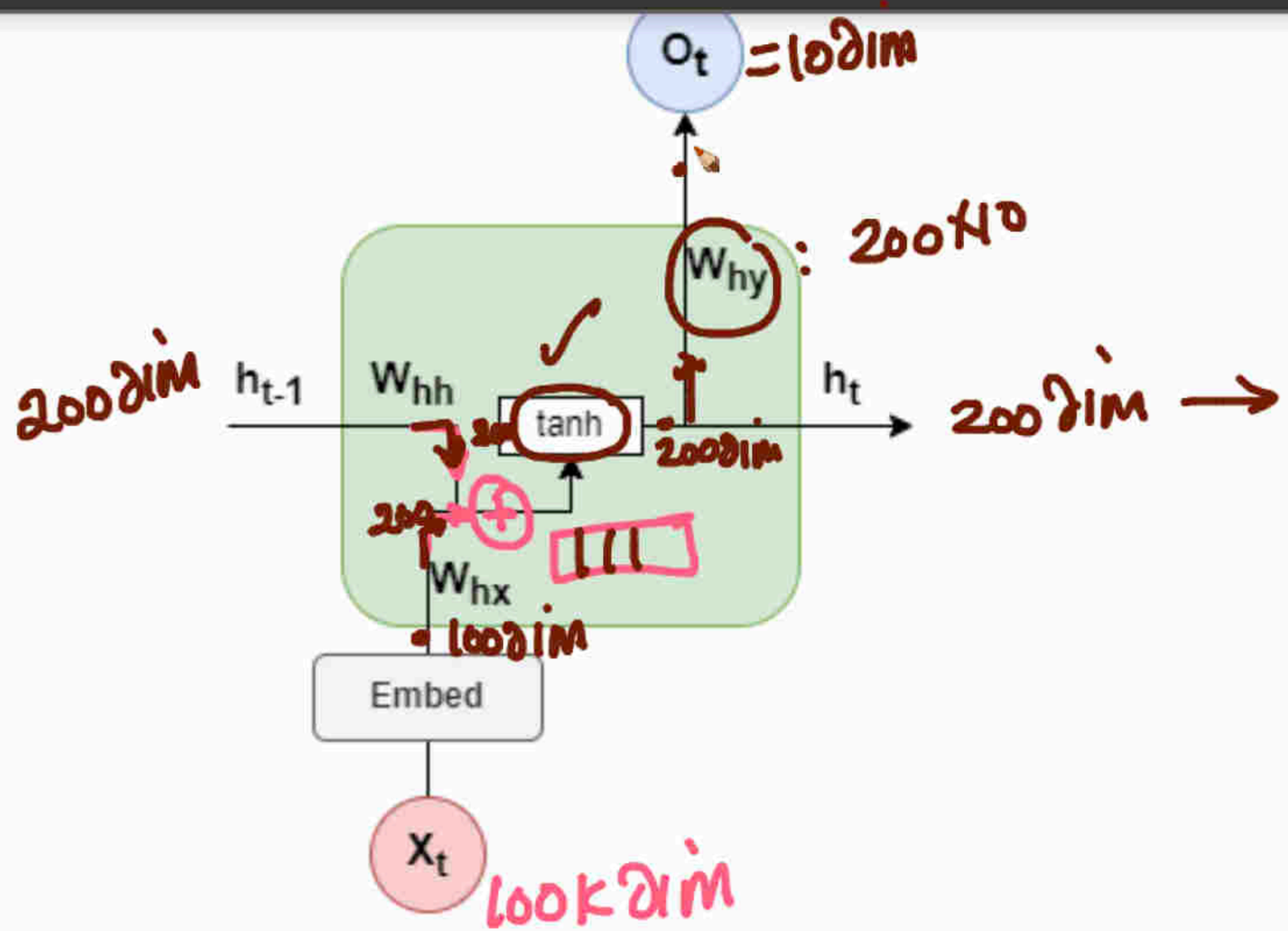
22 / 22

Chrome File Edit View History Bookmarks Profiles Tab Window Help 9:59
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



23 / 23

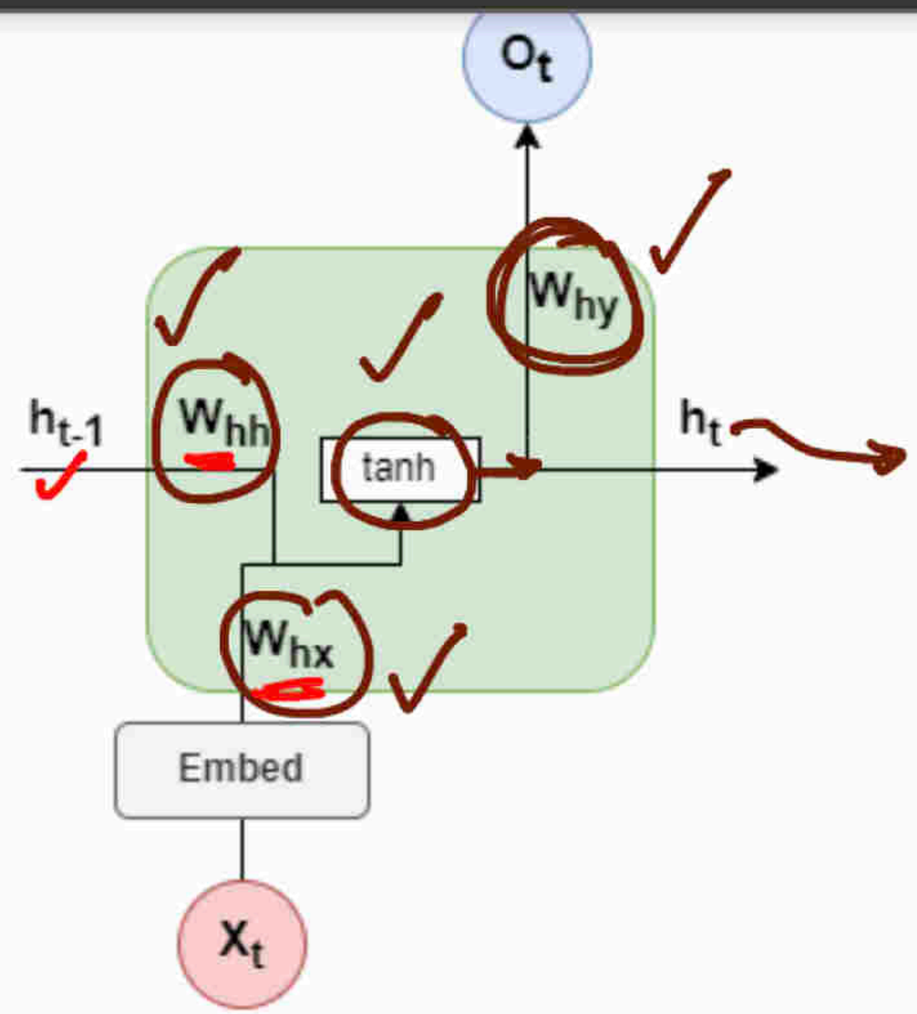
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:01
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect



24 / 24

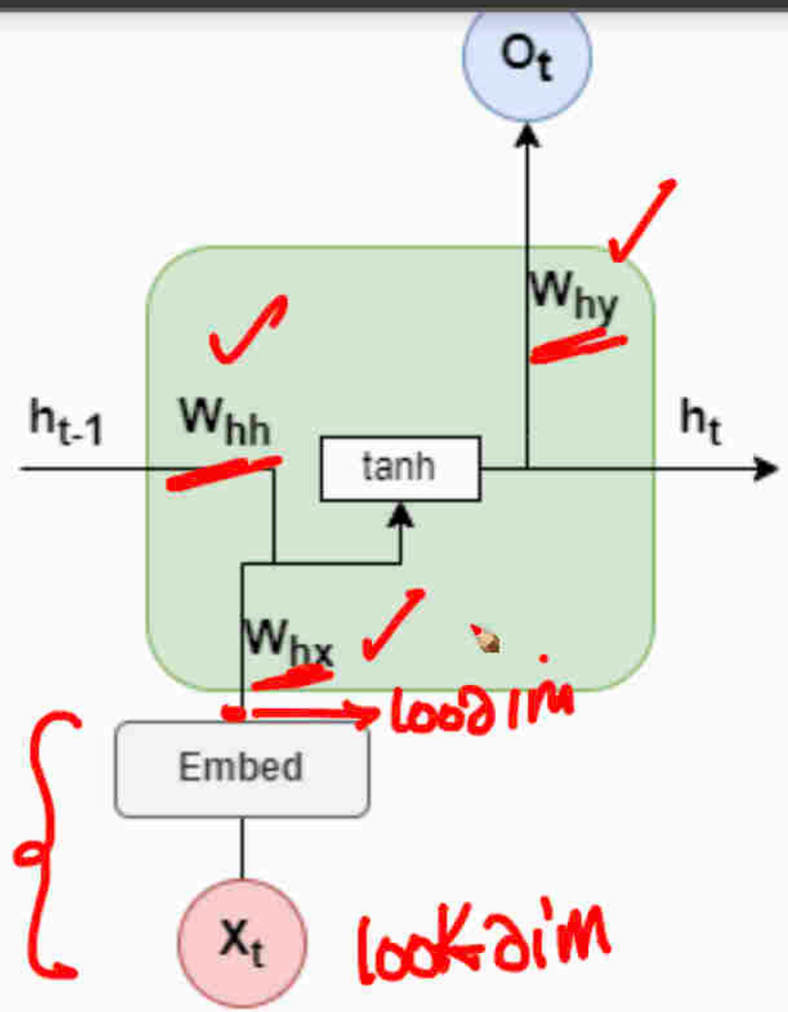
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:02
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect

Simple RNN



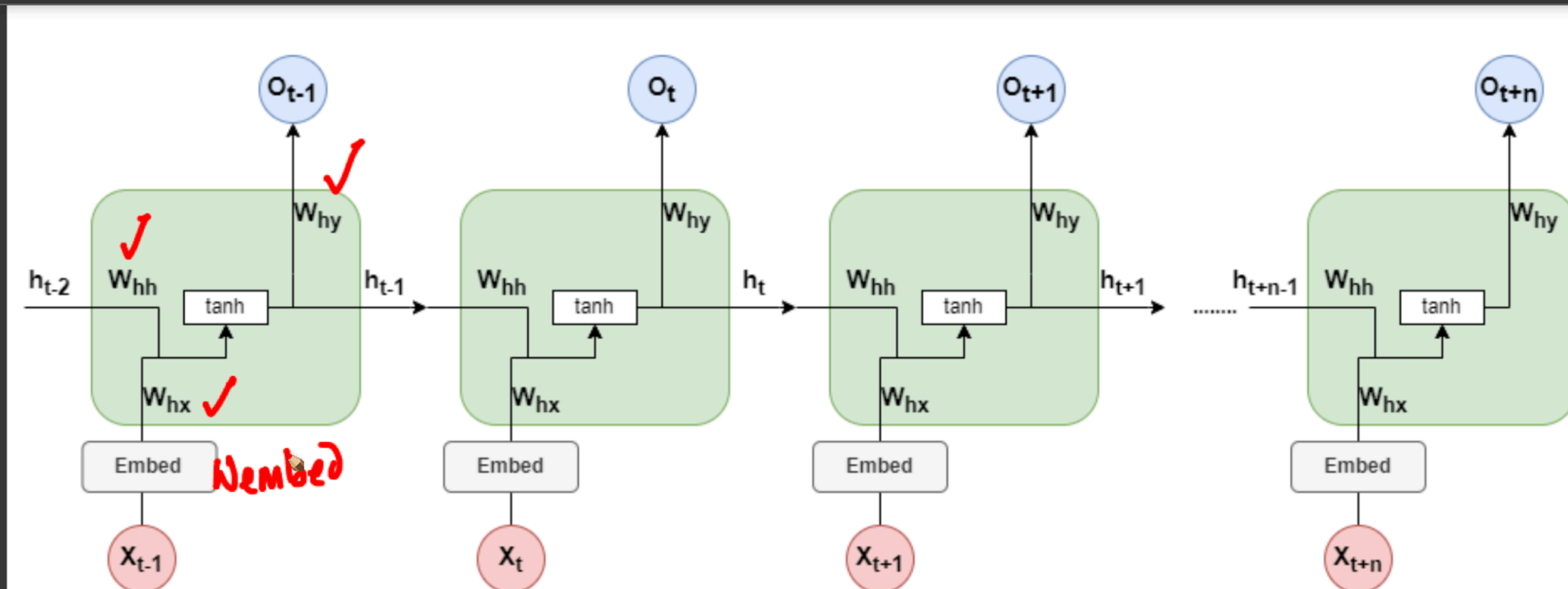
25 / 25

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:03
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect



26 / 26

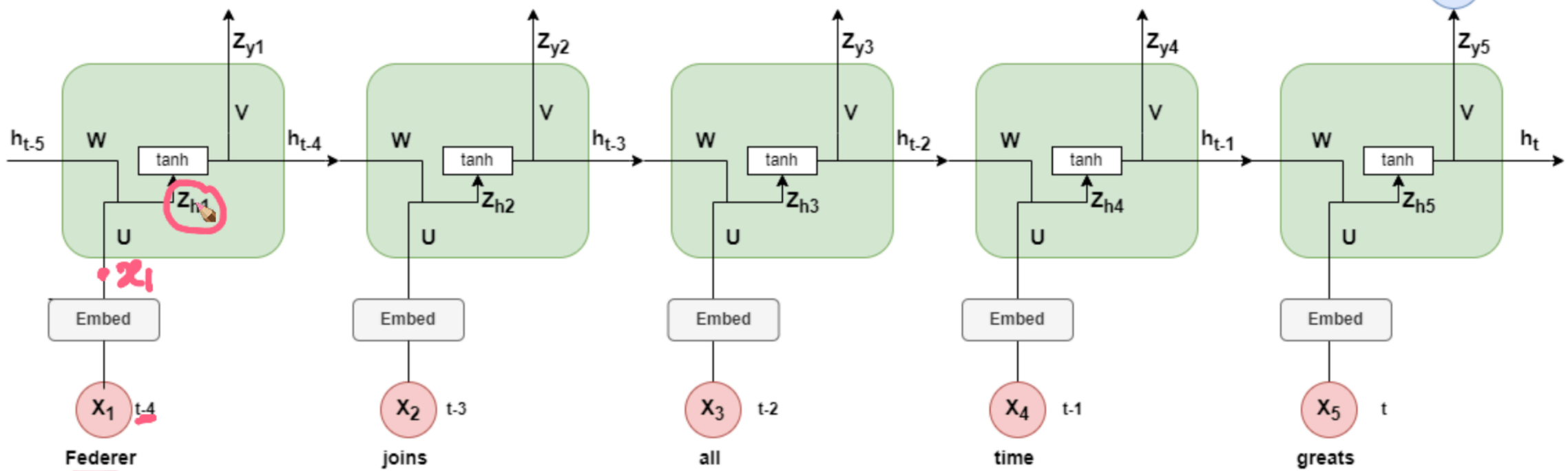
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:04
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect



How Forward propagation is different in RNN?

27 / 27

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:06
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



$Z_{h1} = (X_1 * U) + (h_{t-5} * W) + b_h$
$h_1 = \sigma(Z_{h1})$, where σ = tanh
$Z_{y1} = (V * h_{t-4}) + b_y$

$Z_{h2} = (X_2 * U) + (h_{t-4} * W) + b_h$
$h_2 = \sigma(Z_{h2})$, where σ = tanh
$Z_{y2} = (V * h_{t-3}) + b_y$

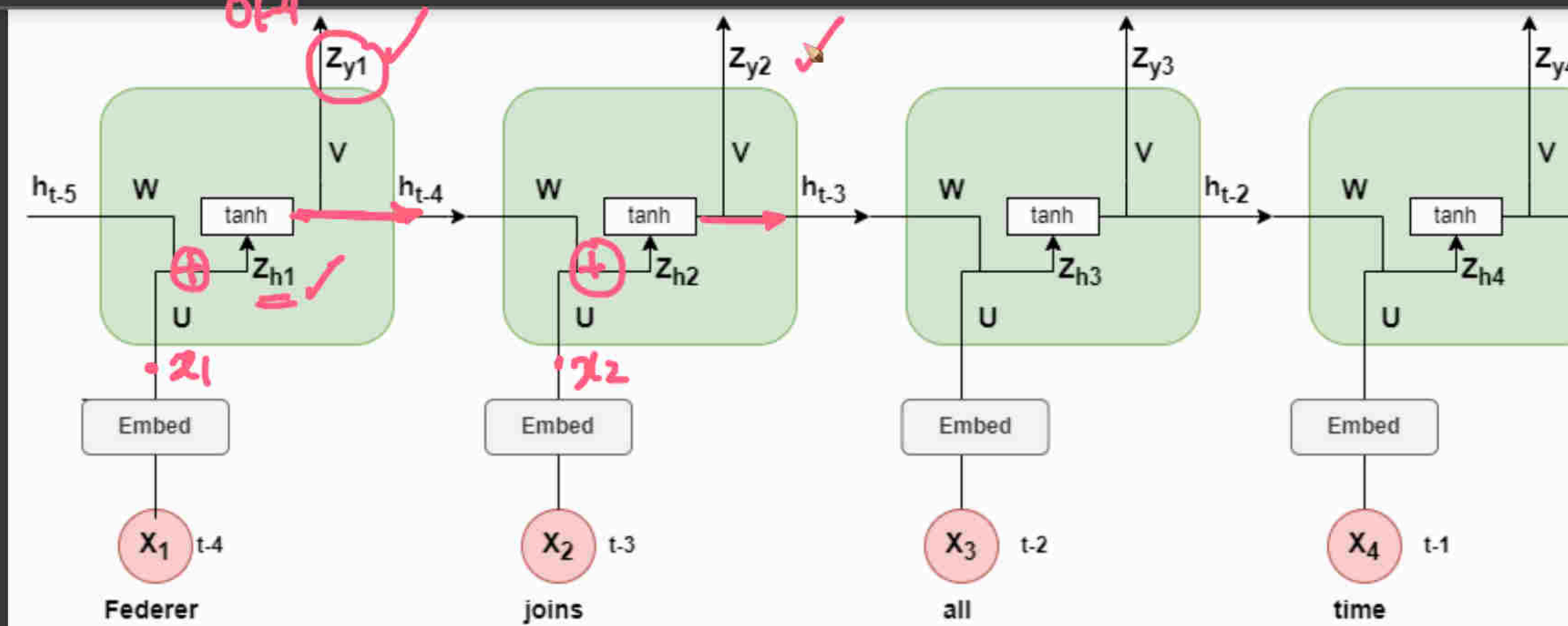
$Z_{h3} = (X_3 * U) + (h_{t-3} * W) + b_h$
$h_3 = \sigma(Z_{h3})$, where σ = tanh
$Z_{y3} = (V * h_{t-2}) + b_y$

$Z_{h4} = (X_4 * U) + (h_{t-2} * W) + b_h$
$h_4 = \sigma(Z_{h4})$, where σ = tanh
$Z_{y4} = (V * h_{t-1}) + b_y$

$Z_{h5} = (X_5 * U) + (h_{t-1} * W) + b_h$
$h_5 = \sigma(Z_{h5})$, where σ = tanh
$Z_{y5} = (V * h_t) + b_y$
$O_{t+4} = \sigma(Z_{y5})$, where σ = Softmax

28 / 28

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:08
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect



$Z_{h1} = (X_1 * U) + (h_{t-5} * W) + b_h$
$h_1 = \sigma(Z_{h1})$, where $\sigma = \tanh$
$Z_{y1} = (V * h_{t-4}) + b_y$

$Z_{h2} = (X_2 * U) + (h_{t-4} * W) + b_h$
$h_2 = \sigma(Z_{h2})$, where $\sigma = \tanh$
$Z_{y2} = (V * h_{t-3}) + b_y$

$Z_{h3} = (X_3 * U) + (h_{t-3} * W) + b_h$
$h_3 = \sigma(Z_{h3})$, where $\sigma = \tanh$
$Z_{y3} = (V * h_{t-2}) + b_y$

$Z_{h4} = (X_4 * U) + (h_{t-2} * W) +$
$h_4 = \sigma(Z_{h4})$, where $\sigma = \tanh$
$Z_{y4} = (V * h_{t-1}) + b_y$

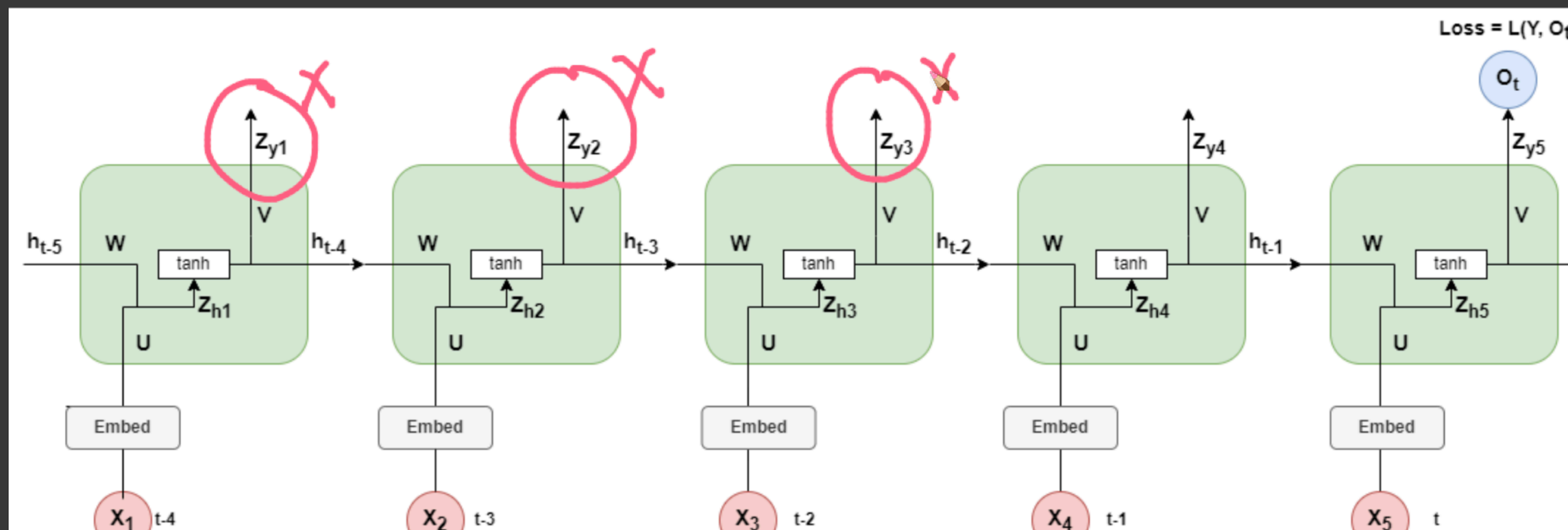
29 / 29

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:08
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

σ = Activation function

Y = Actual

L = Loss function



30 / 30

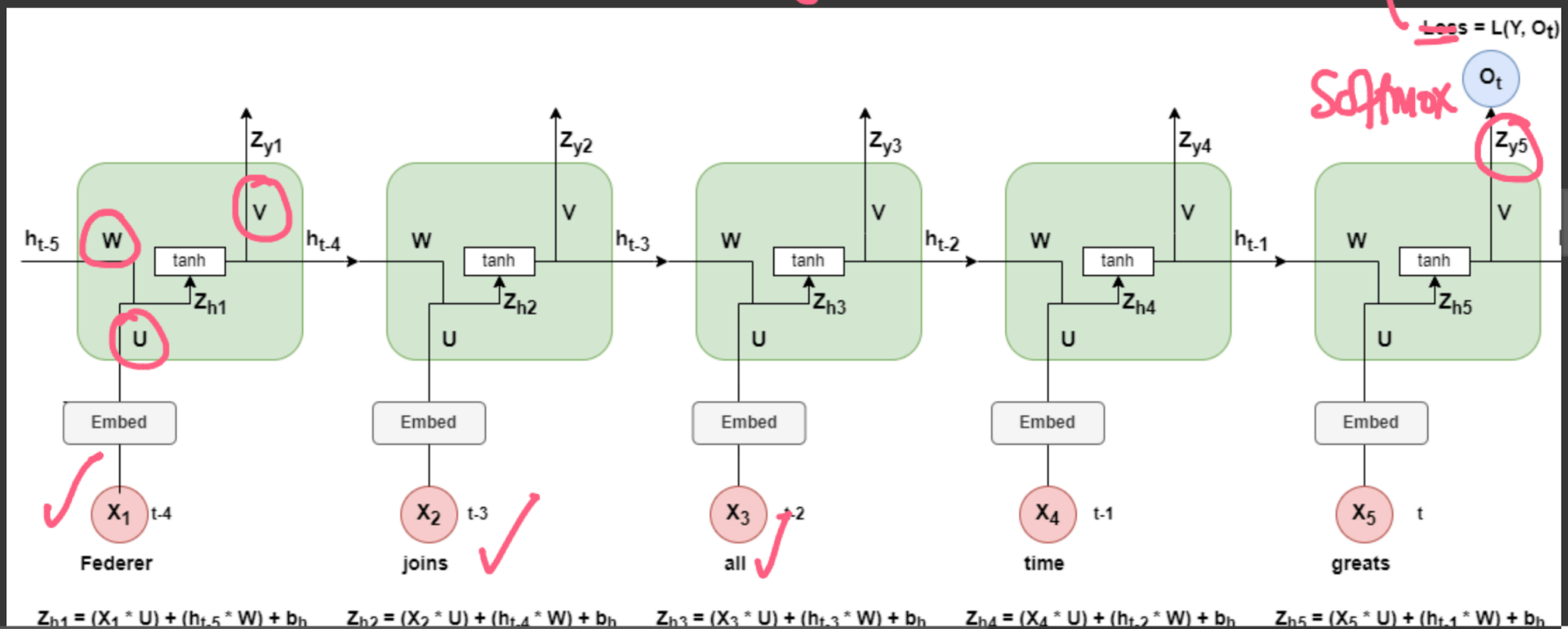
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:09
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

Y = Actual

L = Loss function

many-one

BCE

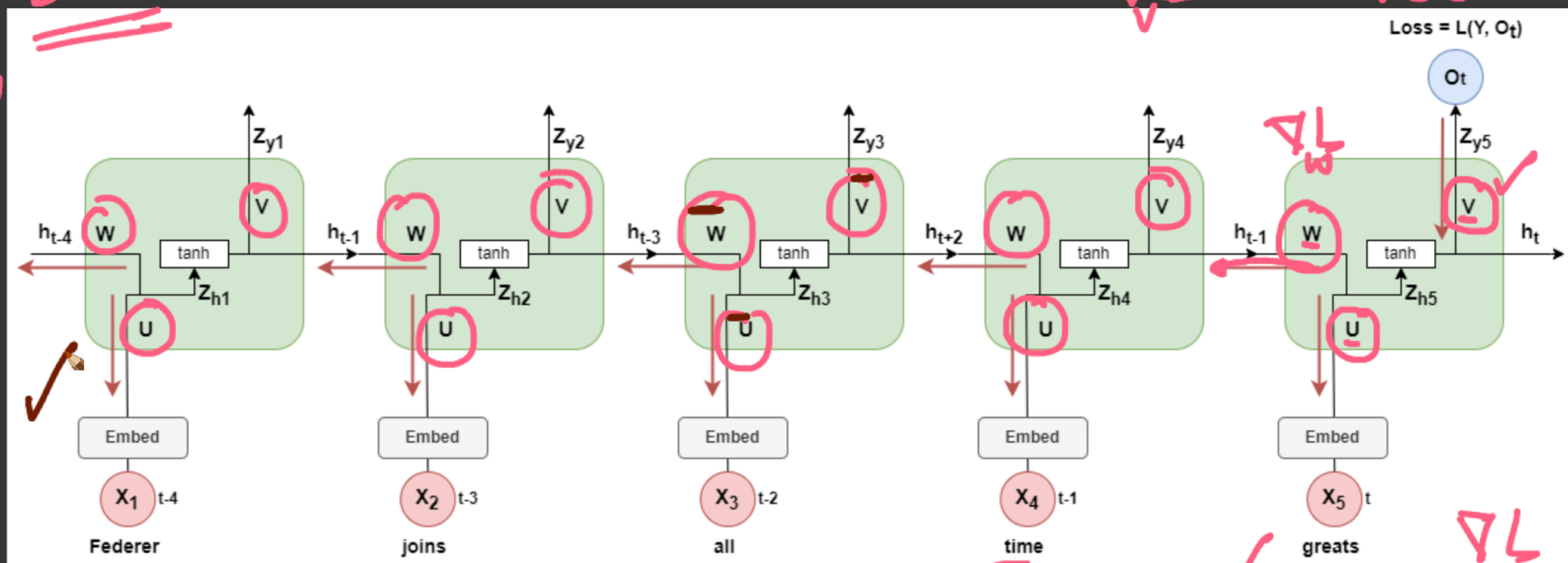


31 / 31

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:15
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

BPTT

W, V, U

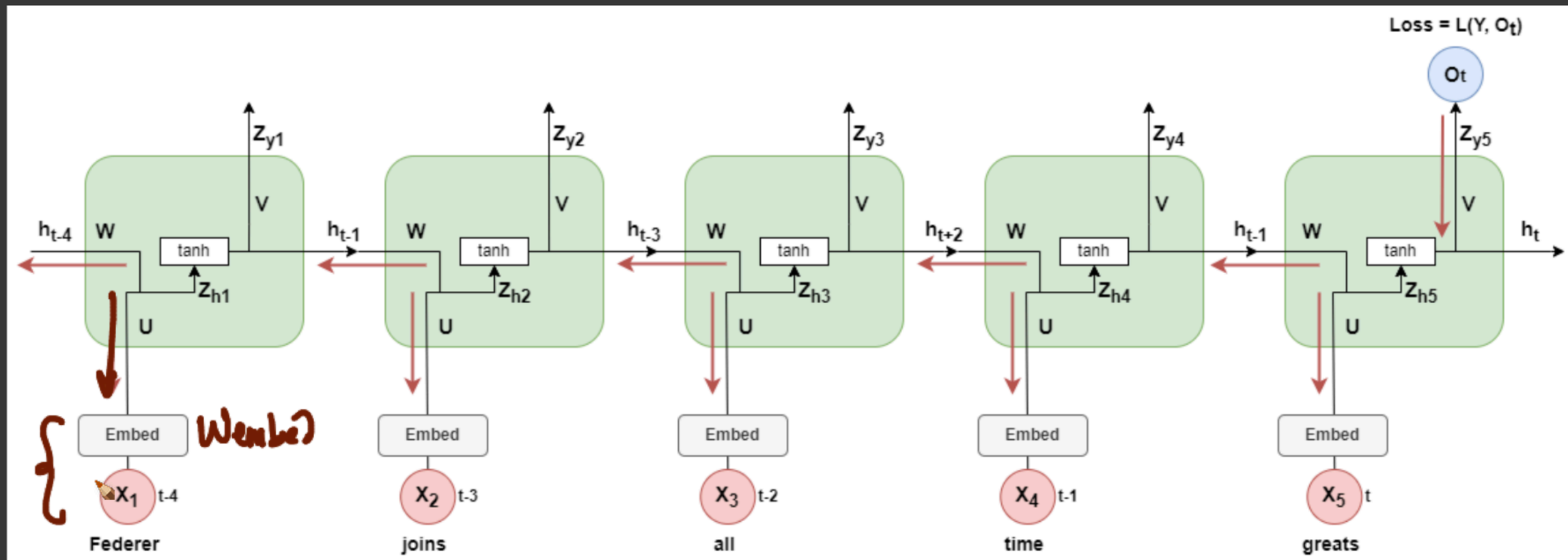


→ time

Usual steps involved in updating parameter weights:

32 / 32

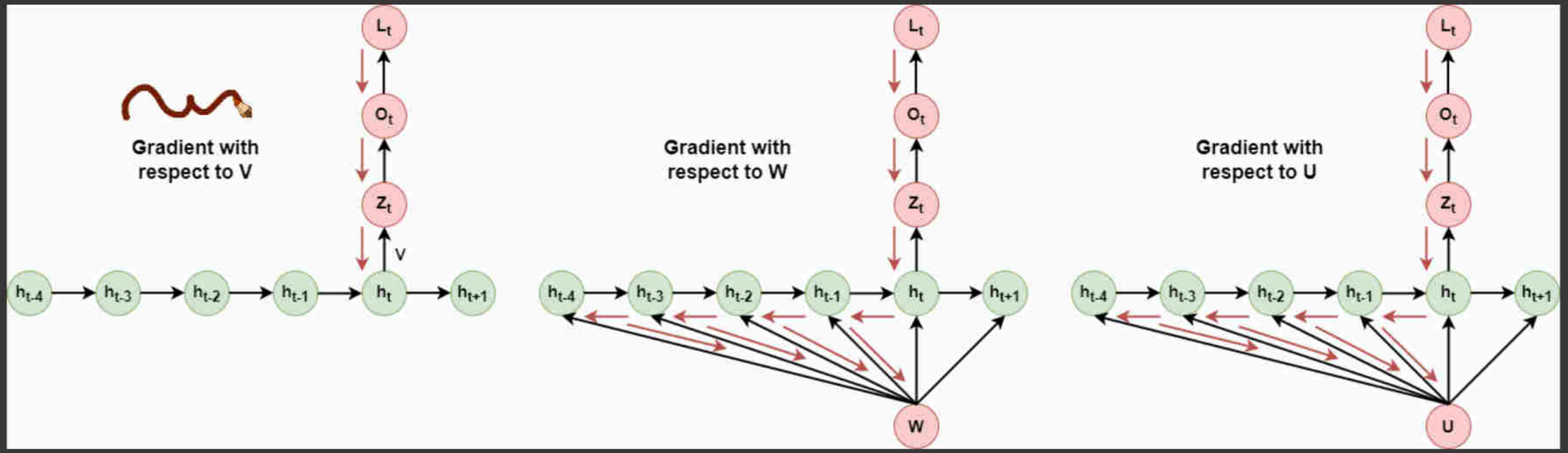
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:15
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



Usual steps involved in updating parameter weights:

33 / 33

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:16
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect



Loss function

What is a Loss function and why we need?

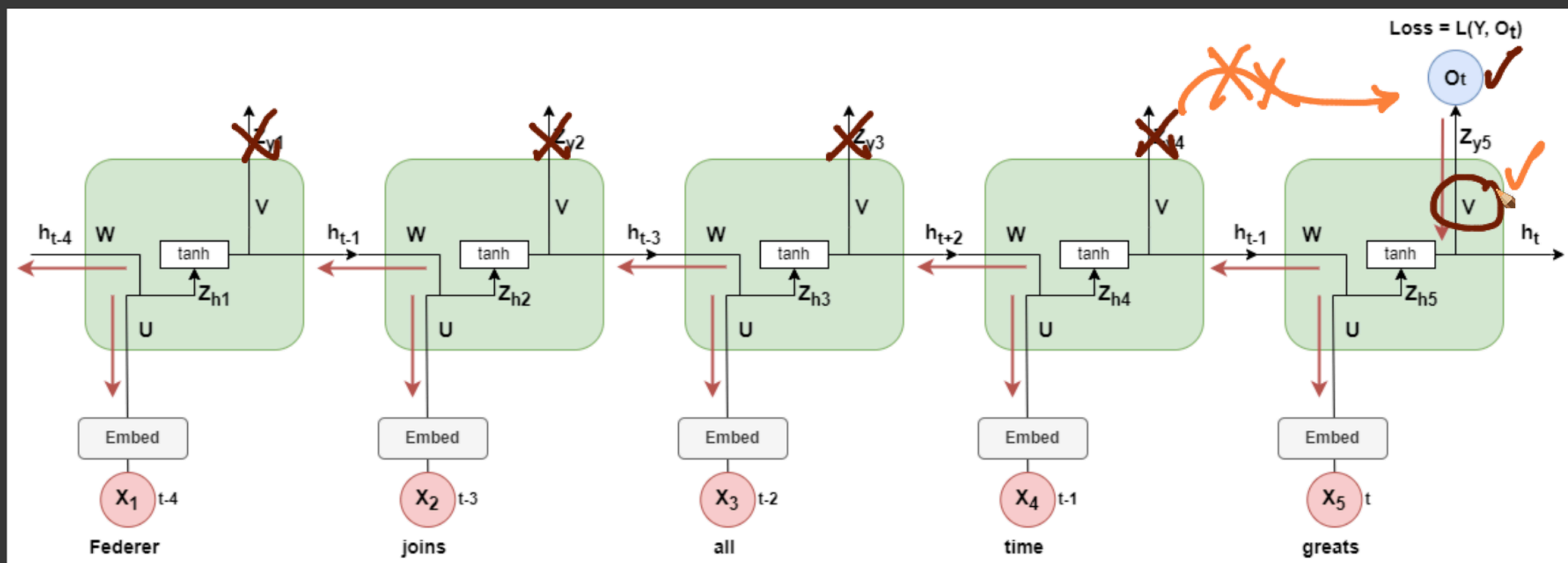
34 / 34

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:17
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

Backward propagation

many -one

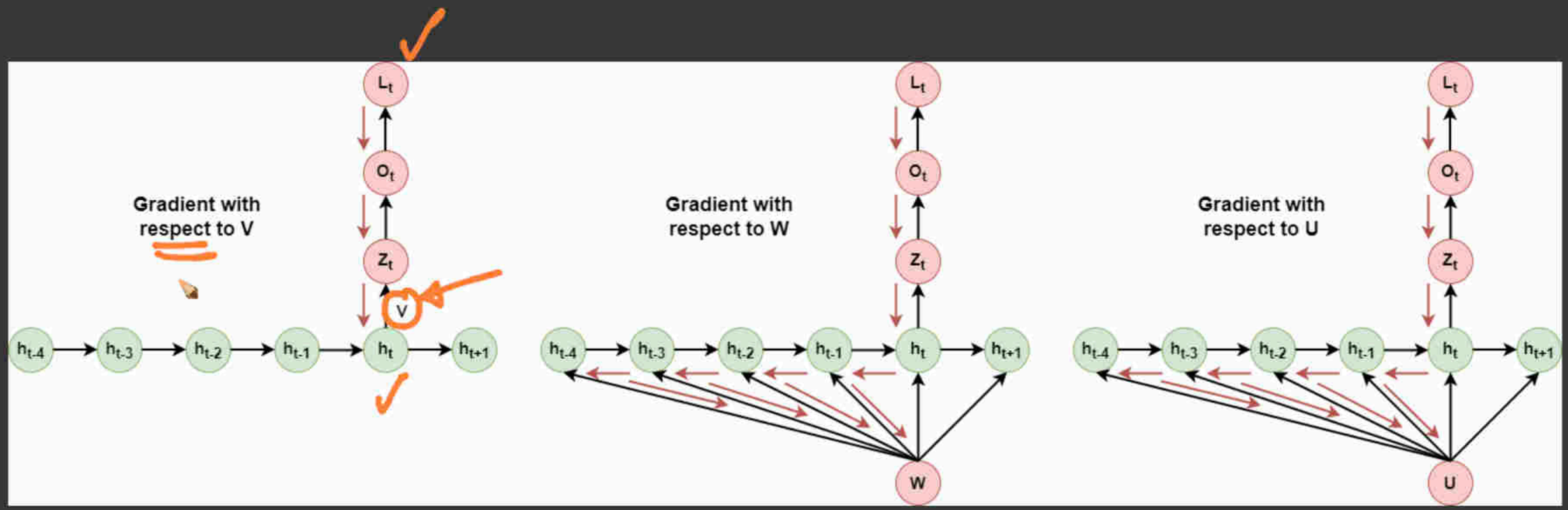
$\frac{\partial L}{\partial V}$



35 / 35

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:18
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect

Simpler representation to understand back propagation through time



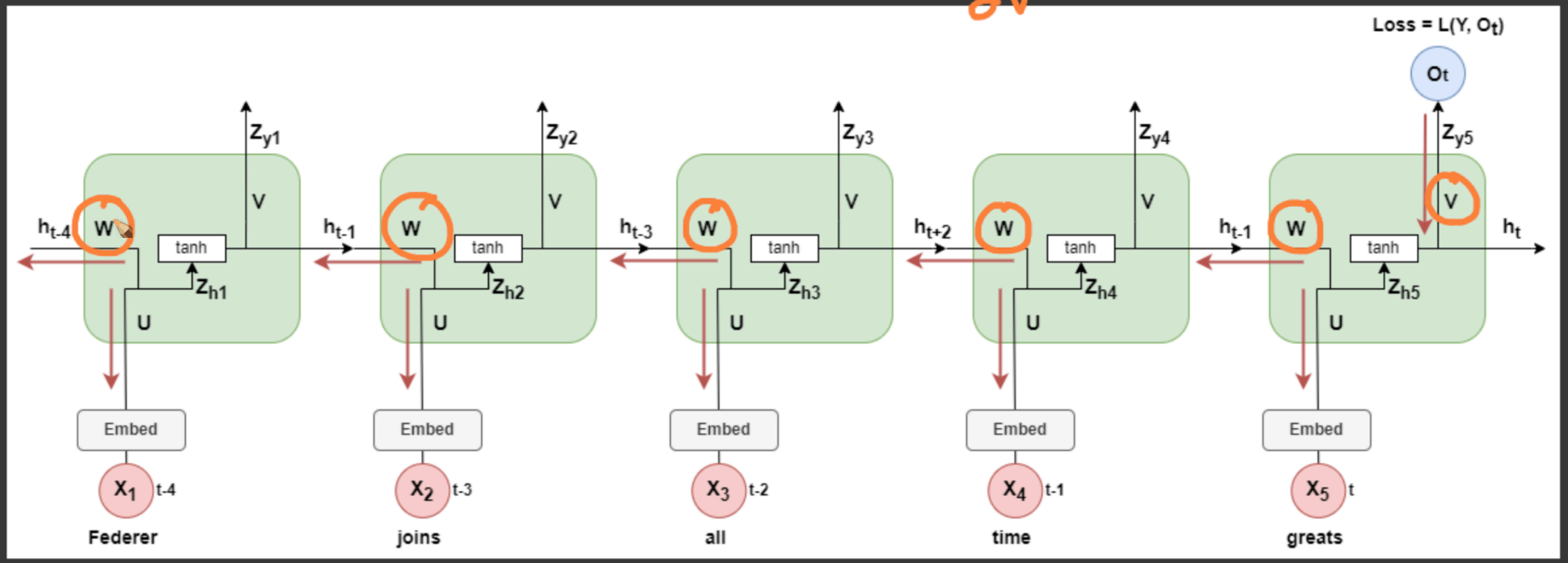
Loss function

36 / 36

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:19
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

Backward propagation

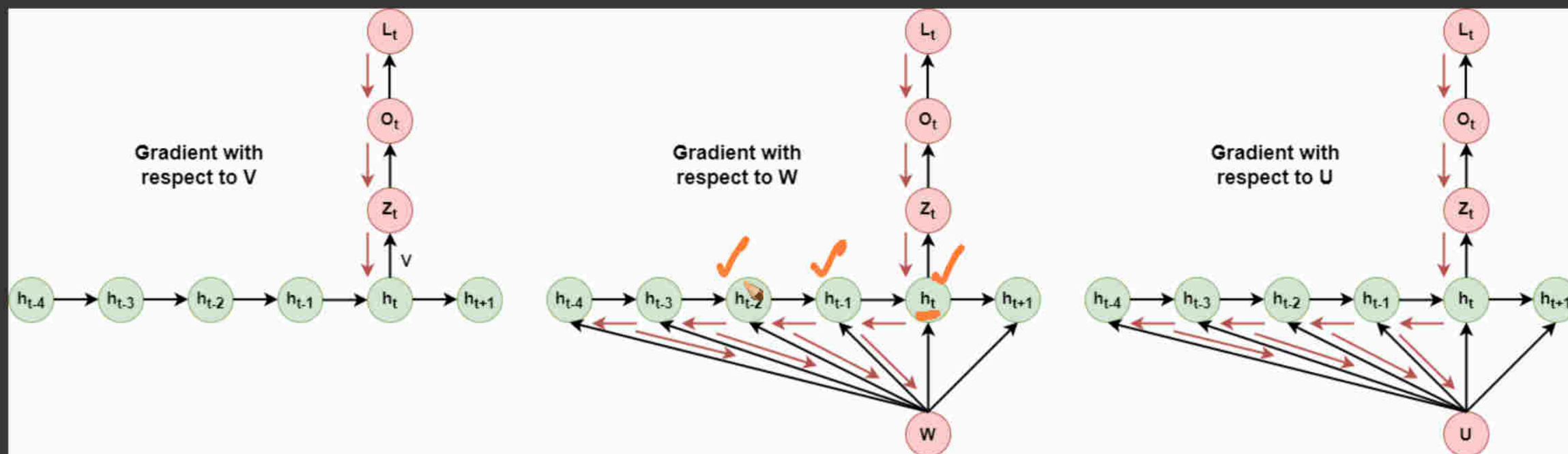
$\frac{\partial L}{\partial V}$



Usual steps involved in updating parameter weights:

37 / 37

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:20
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect



Loss function

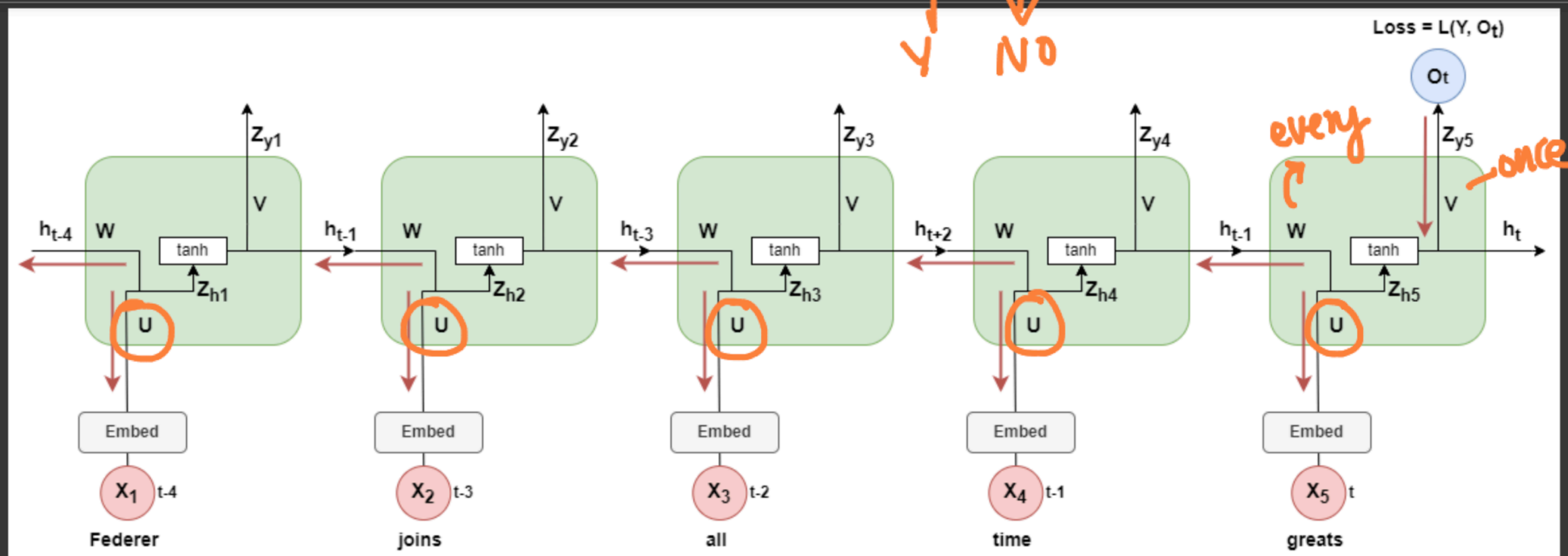
What is a Loss function and why we need?

38 / 38

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:21
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

every once

Y NO

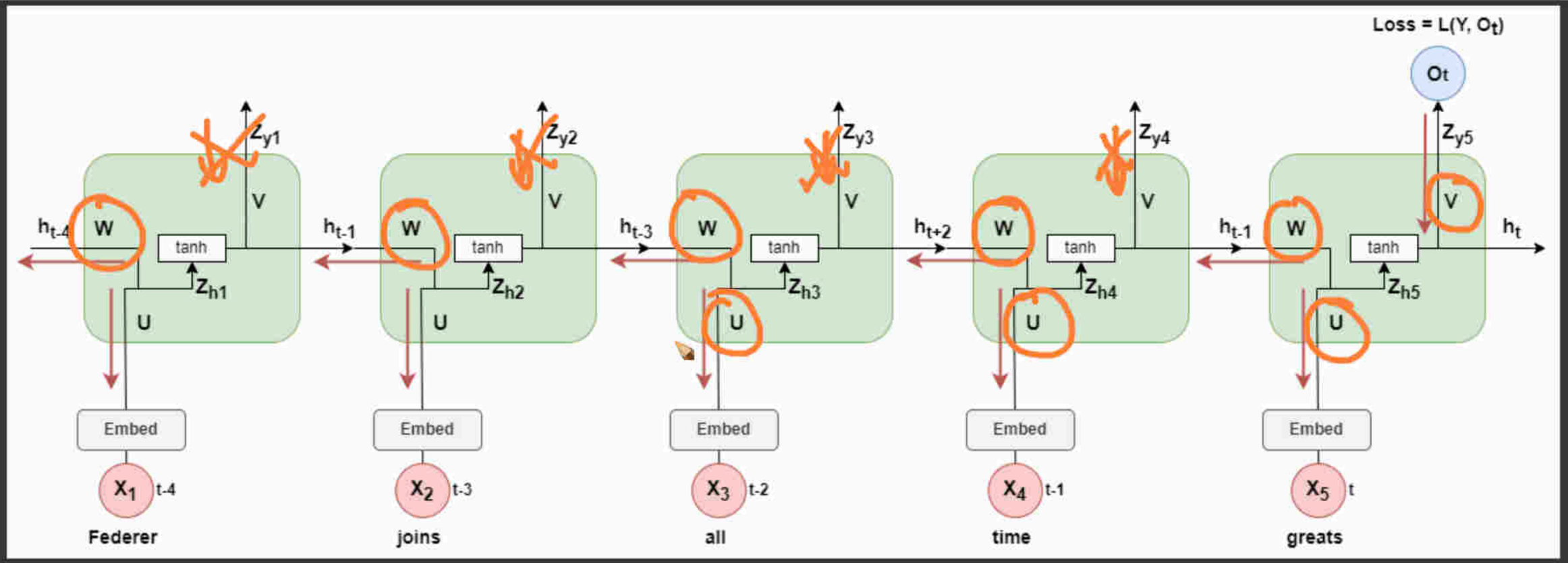


Usual steps involved in updating parameter weights:

- Calculate the gradients of the loss with respect to the parameters
- Multiply it with the Learning rate

39 / 39

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:22
Vanilla-RNN_V1.ipynb - Colab 1506.00019.pdf Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect



Usual steps involved in updating parameter weights:

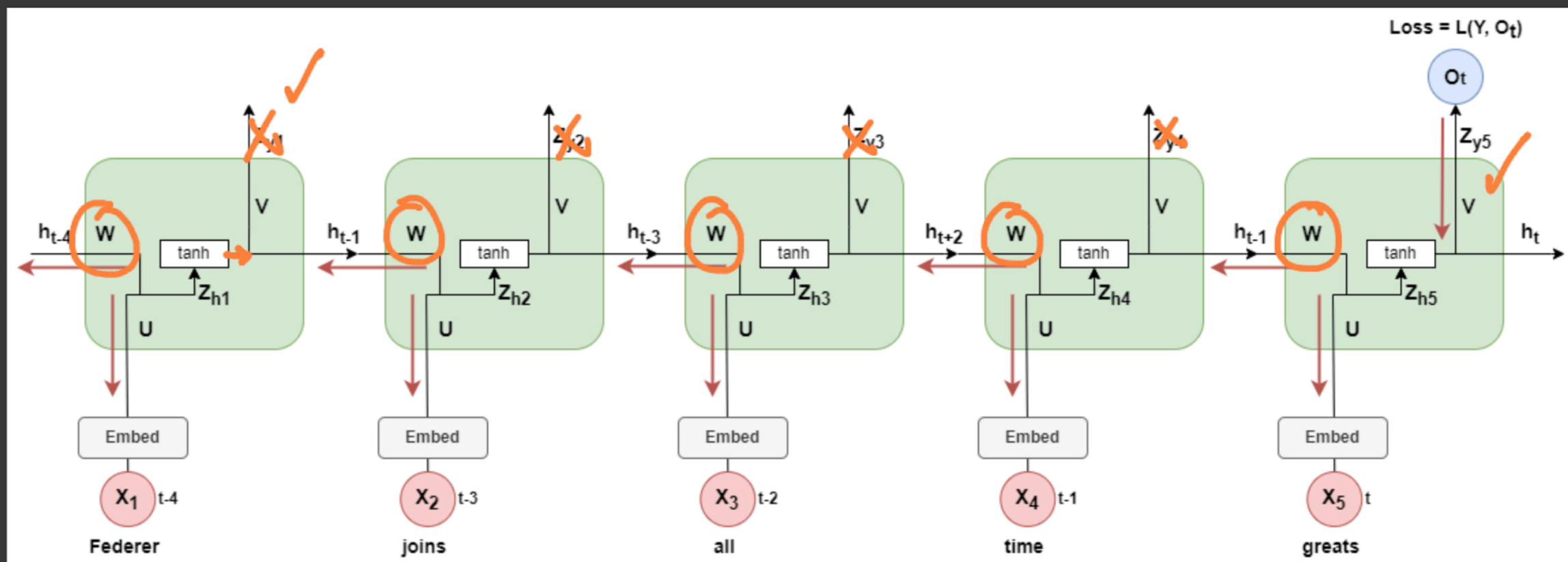
- Calculate the gradients of the loss with respect to the parameters
- Multiply it with the Learning rate

40 / 40

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:24
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

Backward propagation

many-many



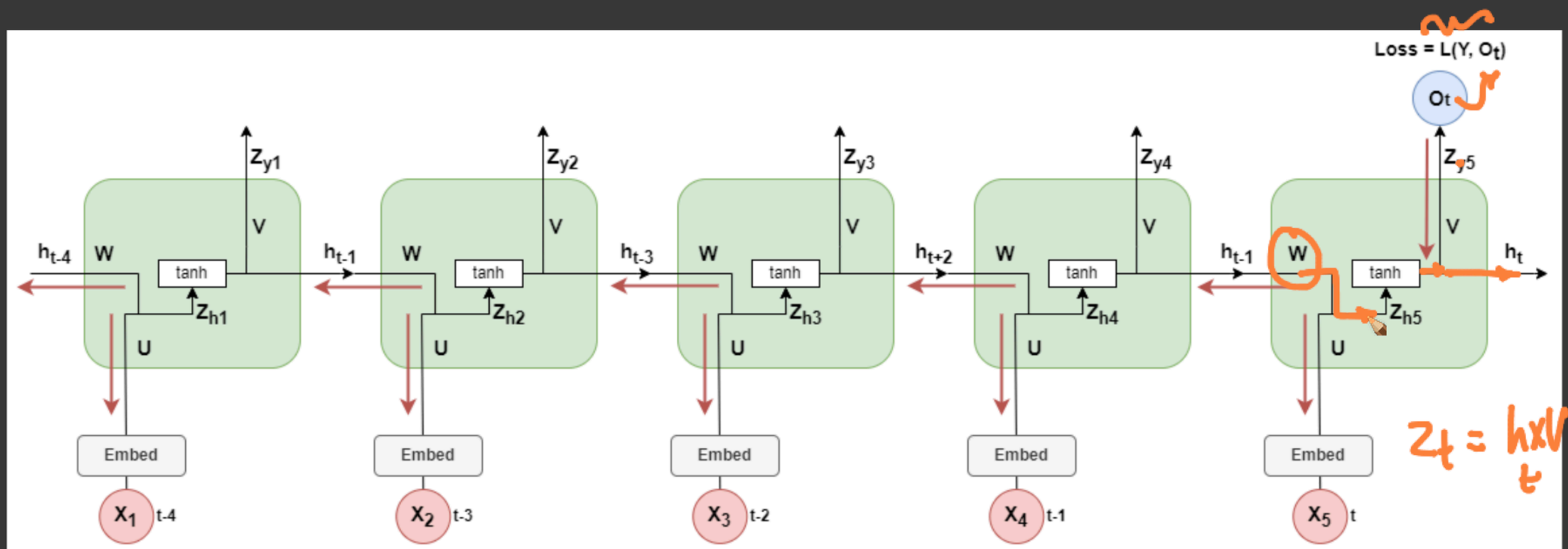
41 / 41

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:27
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

$L = L(O_t, Y_t)$

Backward propagation

$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t} \times \frac{\partial O_t}{\partial Z_t} \times \frac{\partial Z_t}{\partial h_t} \times \frac{\partial h_t}{\partial W}$$



$Z_t = h_t \times V$

42 / 42

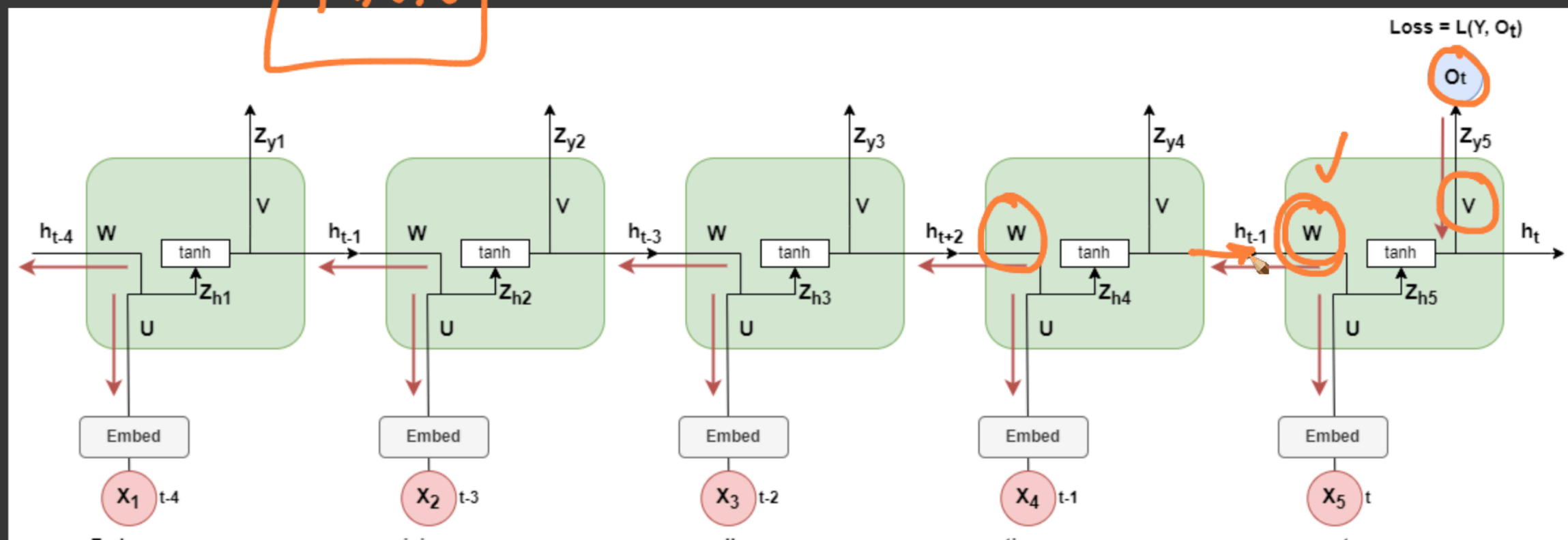
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:28
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

$L = L(O_t, Y_t)$

Backward propagation

$\frac{\partial L_t}{\partial W}$

RNN



43 / 43

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:28
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t} * \frac{\partial O_t}{\partial Z_t} * \frac{\partial Z_t}{\partial h_t} * \frac{\partial h_t}{\partial W}$$

Gradients of loss with respect to W at different t

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

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

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

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

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

Bringing all together

44 / 44

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:29
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

time step using the chain rule in calculus

* A very important thing to note here is the weight matrix **W** is common across all the time steps

$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t} * \frac{\partial O_t}{\partial Z_t} * \frac{\partial Z_t}{\partial h_t} * \frac{\partial h_t}{\partial W}$$

$$\frac{\partial h_t}{\partial \vec{W}}$$

Gradients of loss with respect to W at different t

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

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

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

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

45 / 45

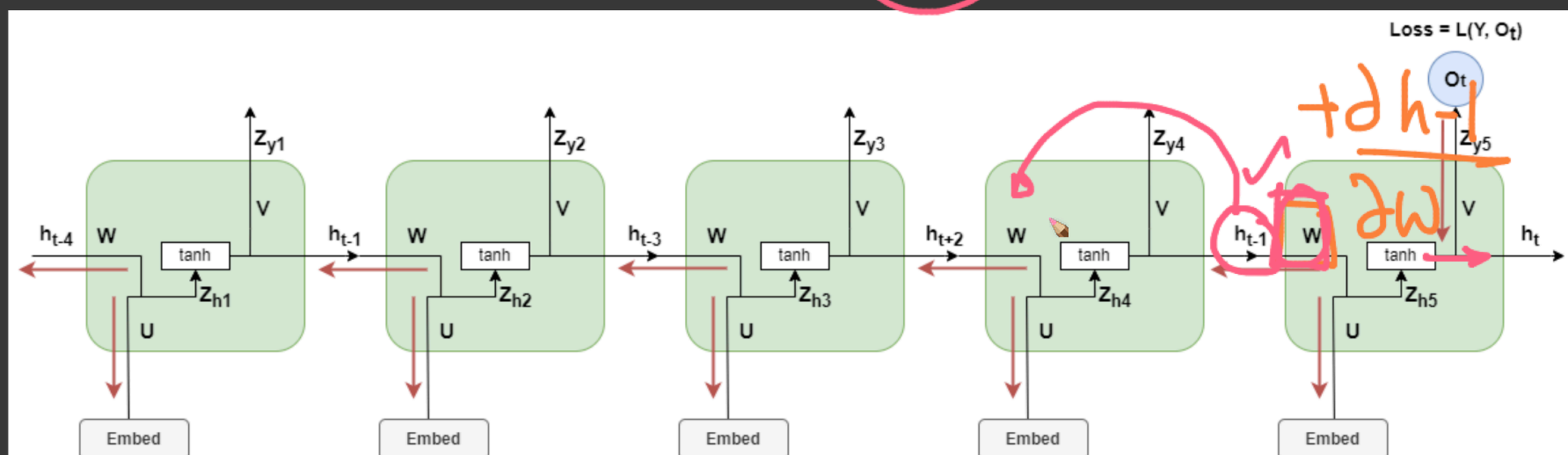
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:30
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

$O_t = \sigma_y(Z_{yt})$

$L = L(O_t, Y_t)$

Backward propagation

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



46 / 46

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:32
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t} * \frac{\partial O_t}{\partial Z_t} * \frac{\partial Z_t}{\partial h_t} * \boxed{\frac{\partial h_t}{\partial W}}$$

$h_{t-2} \xrightarrow{W} \cdots \longrightarrow h_{t-1} \xrightarrow{W} \cdots h_t$

Gradients of loss with respect to W at different t

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

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

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

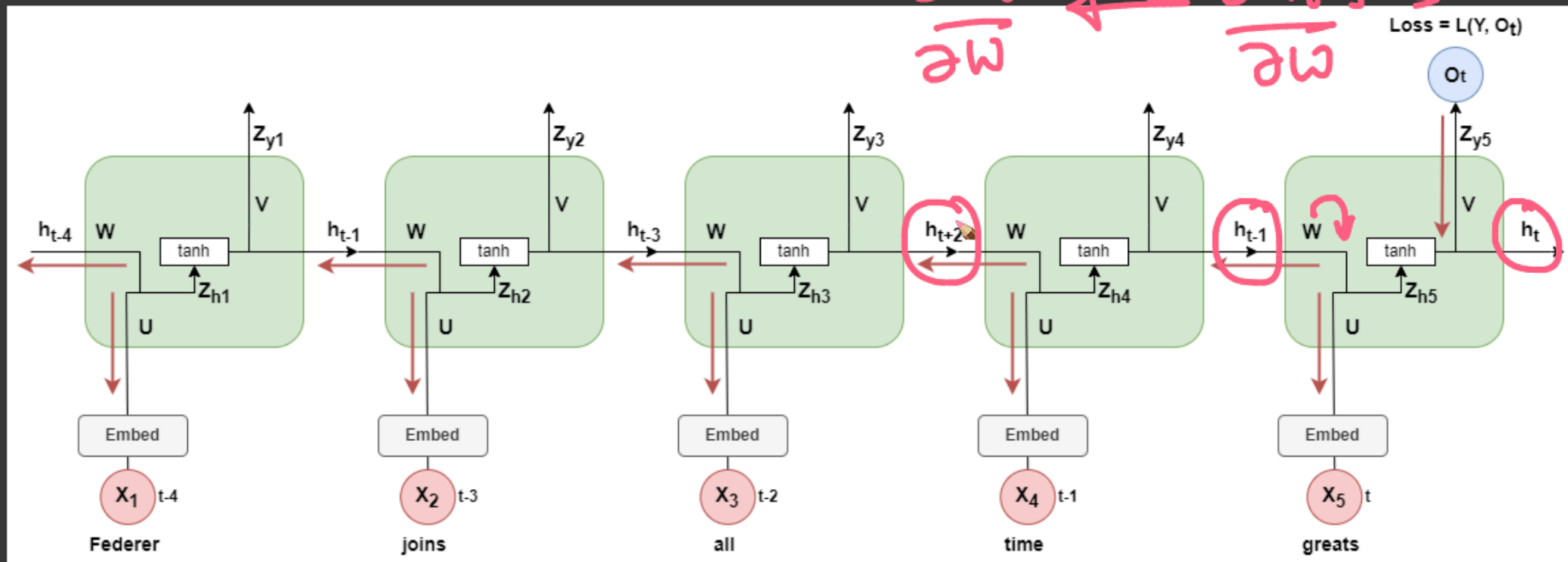
$$\frac{\partial h_{t-3}}{\partial W} = \frac{\partial h_{t-3}}{\partial W} + \frac{\partial h_{t-3}}{\partial h_{t-4}} * \frac{\partial h_{t-4}}{\partial W}$$

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

Bringing all together

47 / 47

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:32
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect



Usual steps involved in updating parameter weights:

- Calculate the gradients of the loss with respect to the parameters

48 / 48

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:33
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t} * \frac{\partial O_t}{\partial Z_t} * \frac{\partial Z_t}{\partial h_t} * \frac{\partial h_t}{\partial W}$$

Gradients of loss with respect to W at different t

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

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

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

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

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

49 / 49

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:35
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial W} +$$

$$\frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial W} +$$

$$\frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial W} +$$

$$\frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial W} +$$

$$\frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial W} +$$

$$\frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial h_{t-5}}\frac{\partial h_{t-5}}{\partial W}$$

BPTT

Chain rule

algebra

Problem?

Let's generalize this to a single formula

$$\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\sum_{r=1}^{t}\frac{\partial h_t}{\partial h_r}\frac{\partial h_r}{\partial W}$$

50 / 50

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:36

Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(

google.com/search?q=tanh+derivative&rlz=1C5CHFA_enIN958IN958&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjl0YrZyqH7AhXm4jgGHav8...

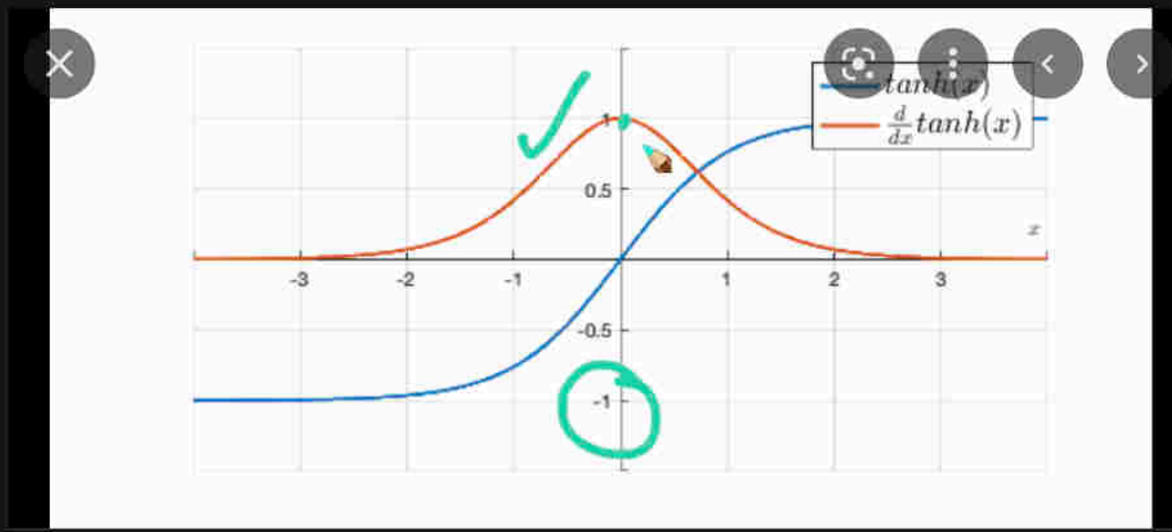
Update

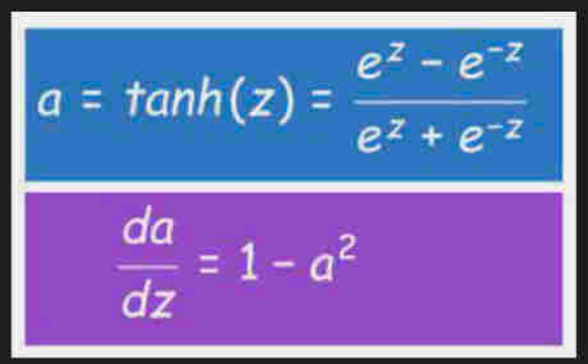
Google

tanh derivative

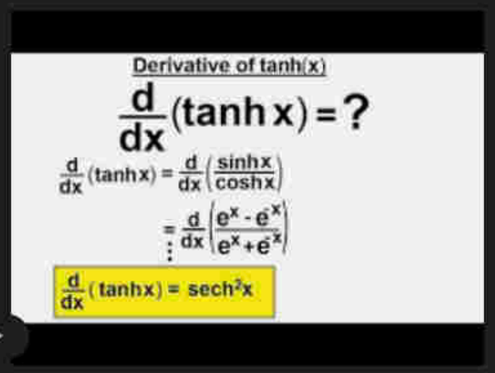
$$\begin{aligned}\frac{d}{dz}\left(\frac{e^z-e^{-z}}{e^z+e^{-z}}\right) &= \frac{e^z+e^{-z}}{(e^z+e^{-z})^2}d\left(e^z-e^{-z}\right) - \frac{e^z-e^{-z}}{(e^z+e^{-z})^2}d\left(e^z+e^{-z}\right)\\ &= \frac{(e^z+e^{-z})(e^z+e^{-z})}{(e^z+e^{-z})^2} - \frac{(e^z-e^{-z})(e^z-e^{-z})}{(e^z+e^{-z})^2}\\ &= \frac{(e^z+e^{-z})^2-(e^z-e^{-z})^2}{(e^z+e^{-z})^2}\\ &= 1-\left(\frac{e^z-e^{-z}}{e^z+e^{-z}}\right)^2\\ &= 1-tanh(z)^2\end{aligned}$$

Derivative of Tanh Function |
blogs.cuit.columbia.edu





Ronny Restrepo
ronny.rest



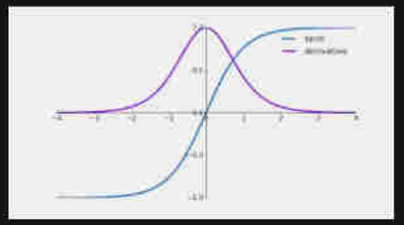
Calculus 2: Hyperbolic Functions (21 of ...
youtube.com

Quora
What is the derivative of tanh(x)? - Quora
Visit

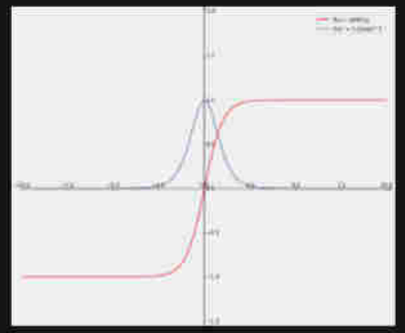
Images may be subject to copyright. Learn More

Related images

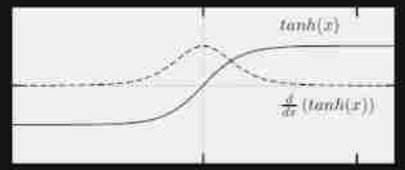
See more



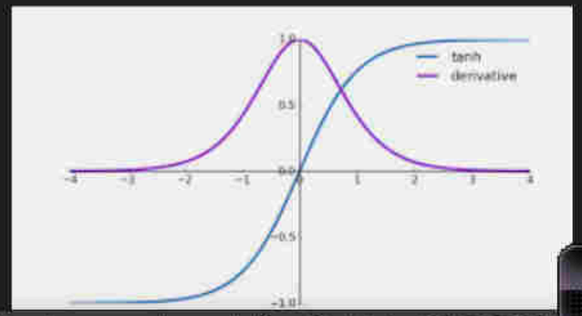
Ronny Restrepo
ronny.rest



Understand tanh(x) Activation F...
le.com



The threshold function, tanh(x) (...
researchgate.net



51 / 51

https://www.google.com/url?sa=i&url=https%3A%2F%2F...

...urce=images&cd=vfe&ved=0CA0QjRxqFwoTCMjQ6trKofsCF...

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:37
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

$$
\begin{aligned}
\frac{\partial L_t}{\partial W} = {} & \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial h_{t-5}}\frac{\partial h_{t-5}}{\partial W}
\end{aligned}
$$

5 - len sentence

100 - len sentence

Let's generalize this to a single formula

$$
\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\sum_{r=1}^{t}\frac{\partial h_t}{\partial h_r}\frac{\partial h_r}{\partial W}
$$

52 / 52

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:38
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

$$
\begin{aligned}
\frac{\partial L_t}{\partial W} = {} & \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial h_{t-5}}\frac{\partial h_{t-5}}{\partial W}
\end{aligned}
$$

(*) Con of simple RNNs

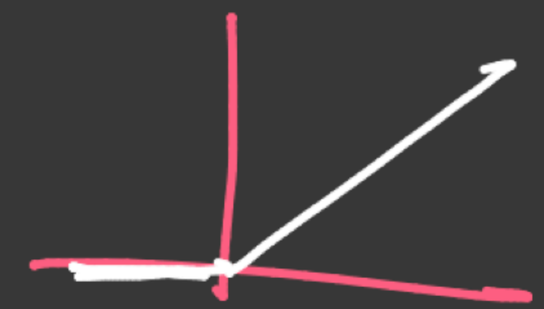
Let's generalize this to a single formula

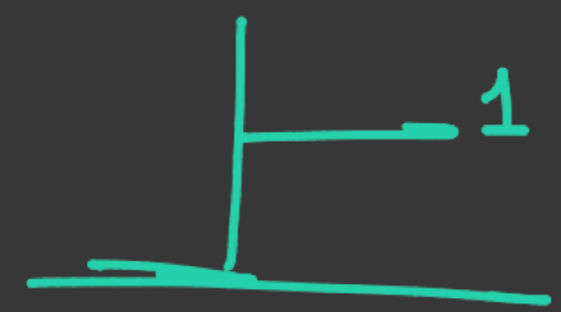
$$
\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\sum_{r=1}^{t}\frac{\partial h_t}{\partial h_r}\frac{\partial h_r}{\partial W}
$$

53 / 53

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:39
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

$$
\begin{aligned}
\frac{\partial L_t}{\partial W} = {} & \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial h_{t-5}}\frac{\partial h_{t-5}}{\partial W}
\end{aligned}
$$





Let's generalize this to a single formula

$$
\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\sum_{r=1}^{t}\frac{\partial h_t}{\partial h_r}\frac{\partial h_r}{\partial W}
$$

54 / 54

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:41
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

$$
\begin{aligned}
\frac{\partial L_t}{\partial W} = {} & \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial W} + \\
& \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}\frac{\partial h_{t-1}}{\partial h_{t-2}}\frac{\partial h_{t-2}}{\partial h_{t-3}}\frac{\partial h_{t-3}}{\partial h_{t-4}}\frac{\partial h_{t-4}}{\partial h_{t-5}}\frac{\partial h_{t-5}}{\partial W}
\end{aligned}
$$

Leaky Relu

Let's generalize this to a single formula

$$
\frac{\partial L_t}{\partial W} = \frac{\partial L_t}{\partial O_t}\frac{\partial O_t}{\partial Z_t}\frac{\partial Z_t}{\partial h_t}\sum_{r=1}^{t}\frac{\partial h_t}{\partial h_r}\frac{\partial h_r}{\partial W}
$$

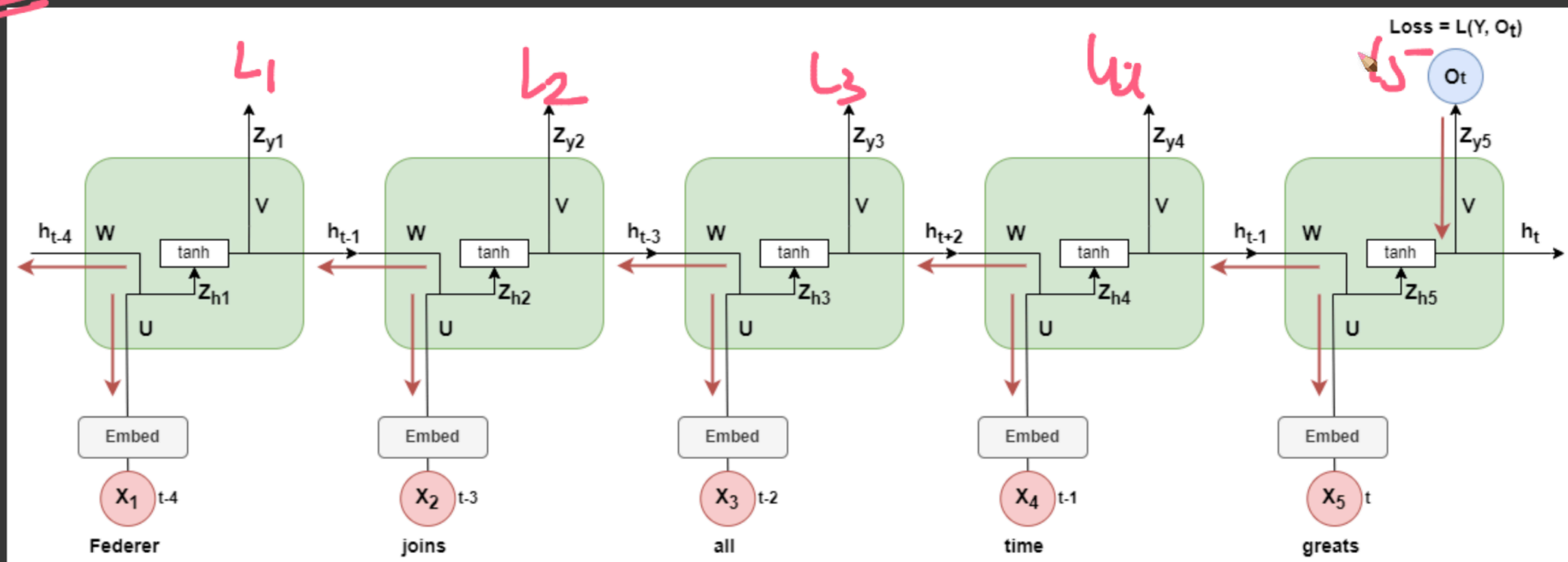
Update the weight matrix W

55 / 55

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:42
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

Backward propagation

POS



Usual steps involved in updating parameter weights:

56 / 56

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:44
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect

the Captain

Scenarion 1: Sachin Tendular was _ _ _ _

Scenarion 2: Sachin Tendular was _____ of Rajya Sabha

Answer: Sachin Tendular was MP of Rajya Sabha

In the scenario 2, it's easier to predict the word "MP" if we had access to input from both the left and the right of the current word instead of only the words from the left.

Difference between UniDirectional and BiDirectional RNN

	UniDirectional RNN	Bidirectional RNN
Information flow	Left to Right	Left to Right & Right to Left
Output depends	h_{t-1}, X_t	h_{t-1}, X_t, h_{t+1}
Weights	W, U, V	W, U, V, W', U', V'

Code Implementation

57 / 57

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:45
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

Scenarion 1: Sachin Tendular was _ _ _ _

Scenarion 2: Sachin Tendular was _ _ _ of Rajya Sabha

Answer: Sachin Tendular was MP of Rajya Sabha

In the scenario 2, it's easier to predict the word "MP" if we had access to input from both the left and the right of the current word instead of only the words from the left.

Difference between UniDirectional and BiDirectional RNN

	UniDirectional RNN	Bidirectional RNN
Information flow	Left to Right	Left to Right & Right to Left
Output depends	h_{t-1}, X_t	h_{t-1}, X_t, h_{t+1}
Weights	W, U, V	W, U, V, W', U', V'

Code Implementation

58 / 58

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:45
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

Scenarion 1: Sachin Tendular was _ _ _ _

Scenarion 2: Sachin Tendular was _ _ of Rajya Sabha

Answer: Sachin Tendular was MP of Rajya Sabha

many-one

In the scenario 2, it's easier to predict the word "MP" if we had access to input from both the left and the right of the current word instead of only the words from the left.

Difference between UniDirectional and BiDirectional RNN

	UniDirectional RNN	Bidirectional RNN
Information flow	Left to Right	Left to Right & Right to Left
Output depends	h_{t-1}, X_t	h_{t-1}, X_t, h_{t+1}
Weights	W, U, V	W, U, V, W', U', V'

Code Implementation

59 / 59

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:46
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect

Scenarion 1: Sachin Tendular was ___ _ _ _

Scenarion 2: Sachin Tendular was _____ of Rajya Sabha

Answer: Sachin Tendular was MP of Rajya Sabha

In the scenario 2, it's easier to predict the word "MP" if we had access to input from both the left and the right of the current word instead of only the words from the left.

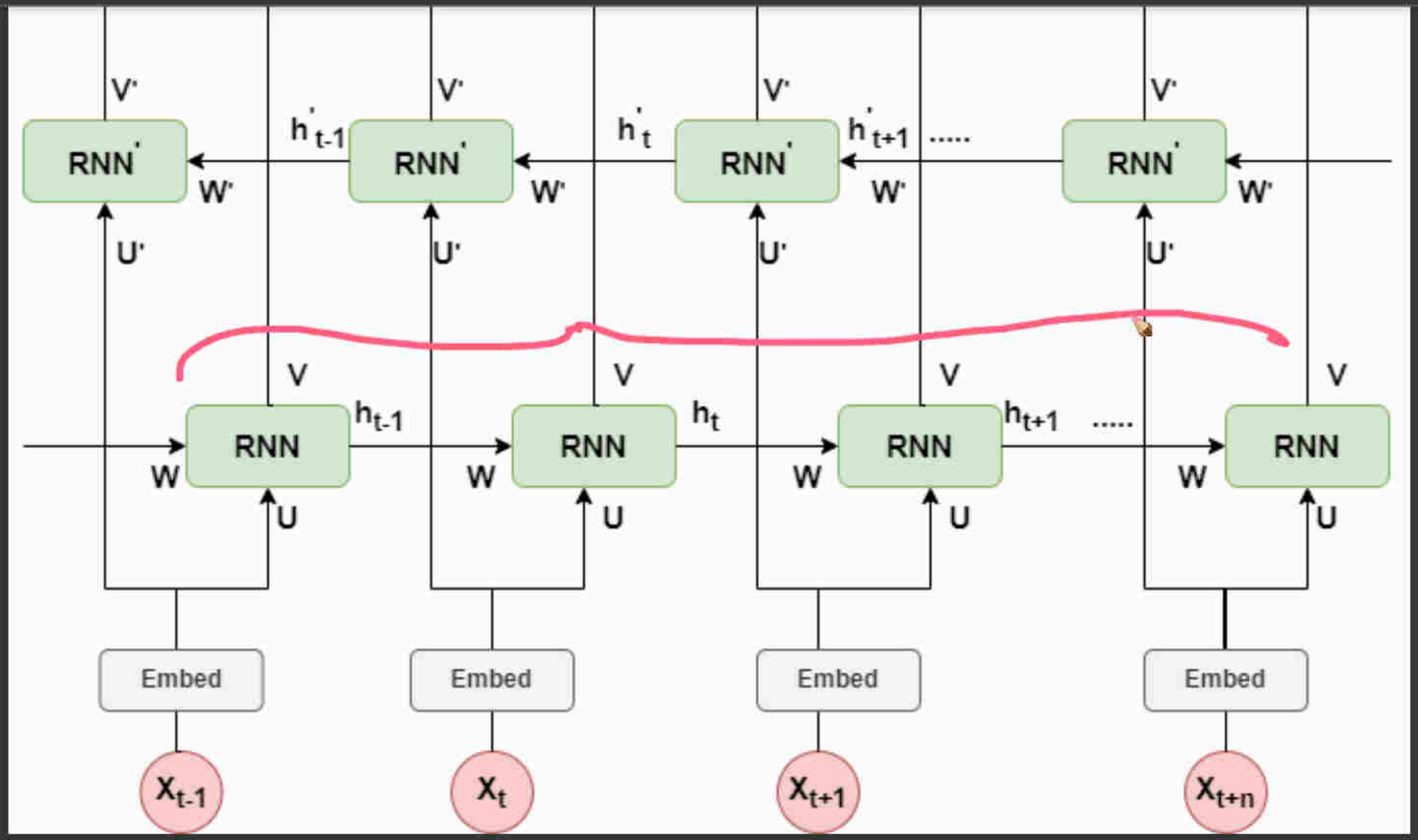
Difference between UniDirectional and BiDirectional RNN

	UniDirectional RNN	Bidirectional RNN
Information flow	Left to Right	Left to Right & Right to Left
Output depends	h_{t-1}, X_t	h_{t-1}, X_t, h_{t+1}
Weights	W, U, V	W, U, V, W', U', V'

Code Implementation

60 / 60

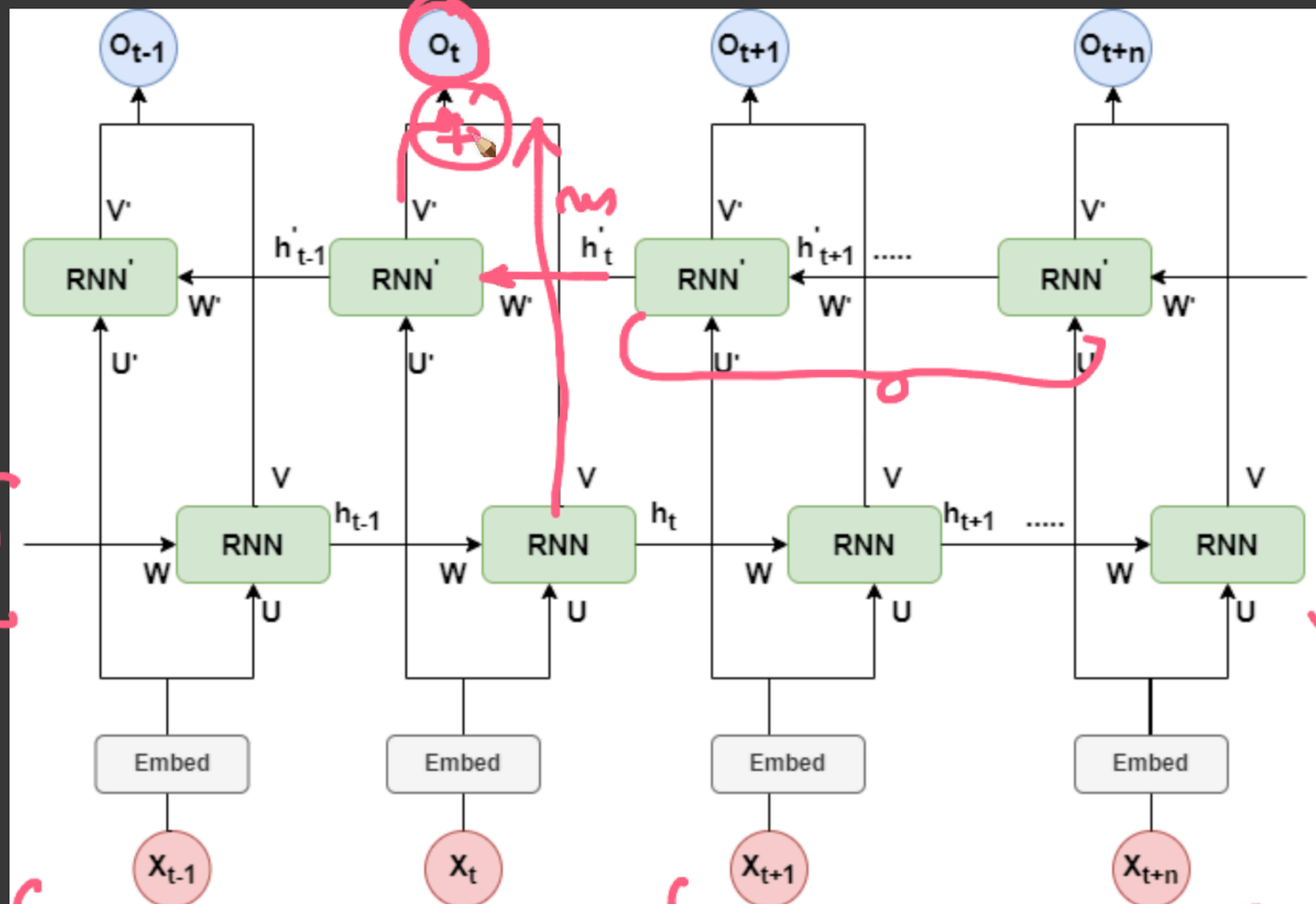
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:46
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect



Scenarion 1: Sachin Tendular was

61 / 61

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:49
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
+ Code + Text Last saved at 17:10 Connect



} → Bi-direc
first-last

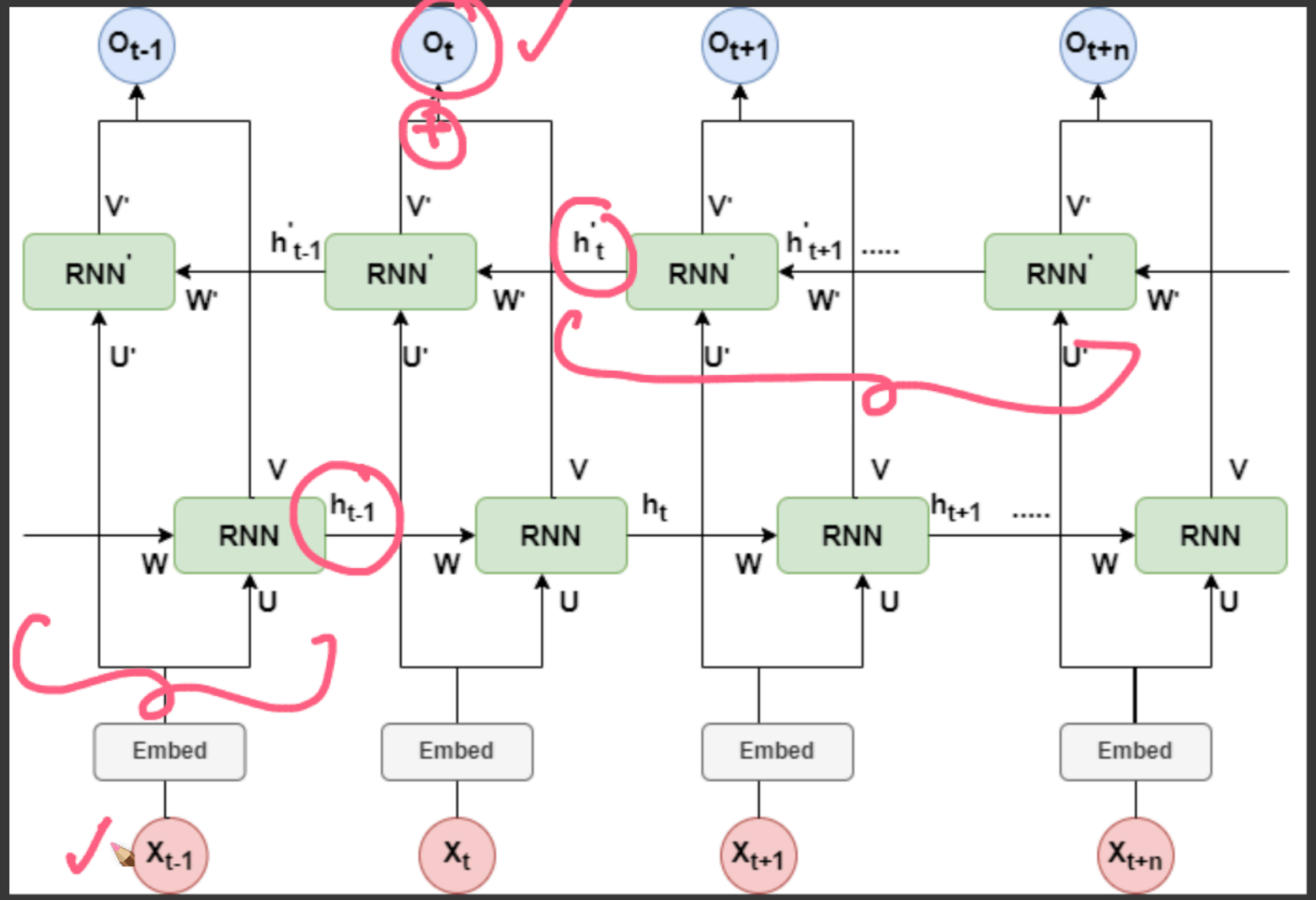
} → Uni-direc
last → first

→ time

Scenarion 1: Sac

62 / 62

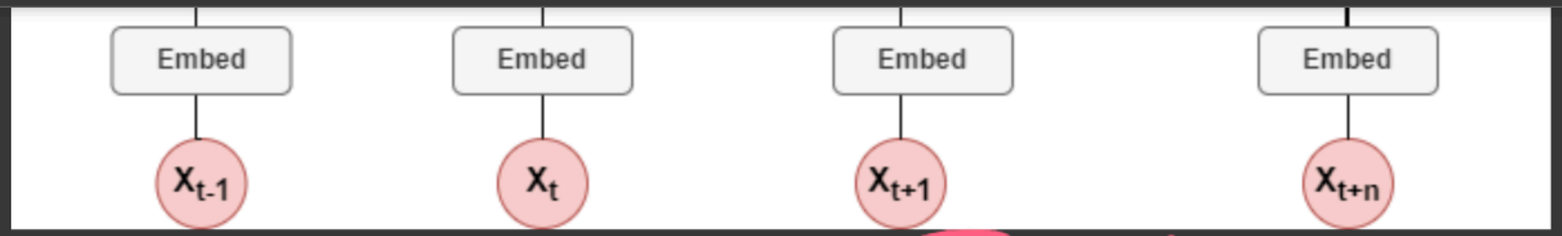
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:50
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65
Update
+ Code + Text Last saved at 17:10 Connect



Scenarion 1: Sac

63 / 63

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:51
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=kK9_LqJC4Y65 Update
+ Code + Text Last saved at 17:10 Connect



OHE

Scenarion 1: Sachin Tendular was _ _ _ _

Scenarion 2: Sachin Tendular was _____ of Rajya Sabha

Answer: Sachin Tendular was MP of Rajya Sabha

1 2 3 4 5 6 7

In the scenario 2, it's easier to predict the word "MP" if we had access to input from both the left and the right of the current word instead of only the words from the left.

Difference between UniDirectional and BiDirectional RNN

	UniDirectional RNN	Bidirectional RNN
Information flow	Left to Right	Left to Right & Right to Left
Output depends	h_{t-1}, X_t	h_{t-1}, X_t, h_{t+1}
Weights	W, U, V	W, U, V, W', U', V'

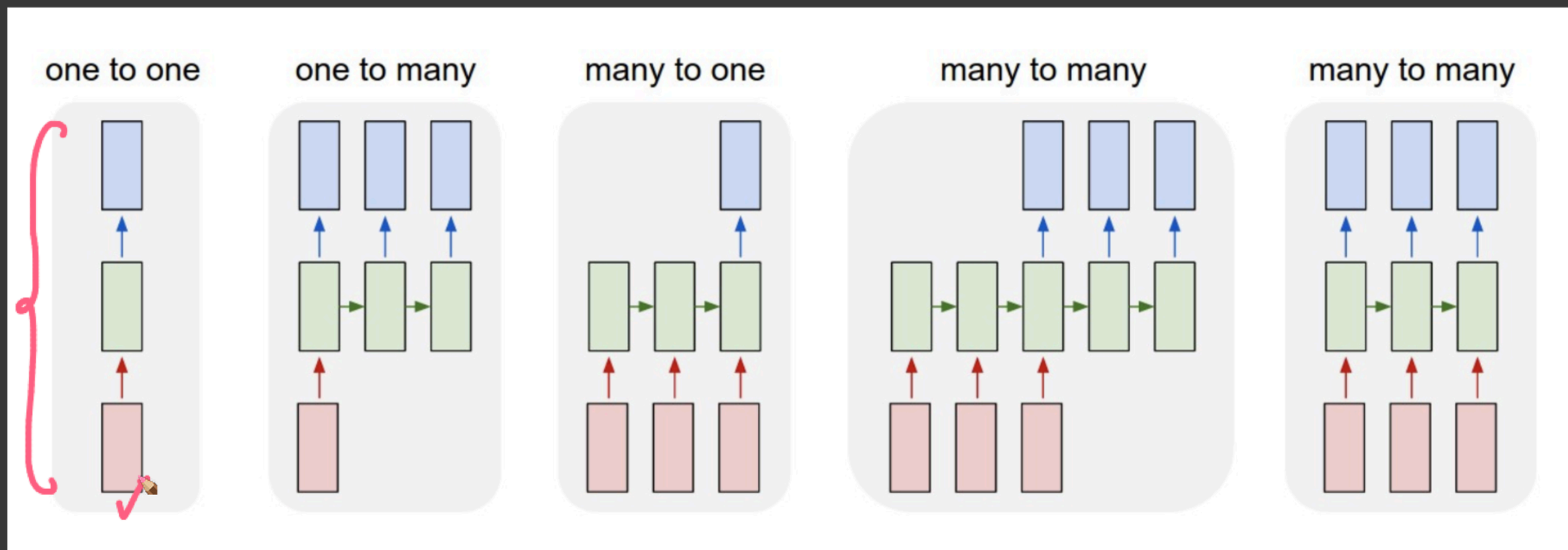
Code Implementation

```
[ ] from tensorflow.keras.preprocessing.sequence import pad_sequences
    from tensorflow.ker                                                    LSTM, GRU, TimeDistributed
```

64 / 64

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:56
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=D-B-sDCI4Y7B
+ Code + Text Last saved at 17:10 Connect

There are couple of varaints of RNN based on the number of inputs and the outputs.

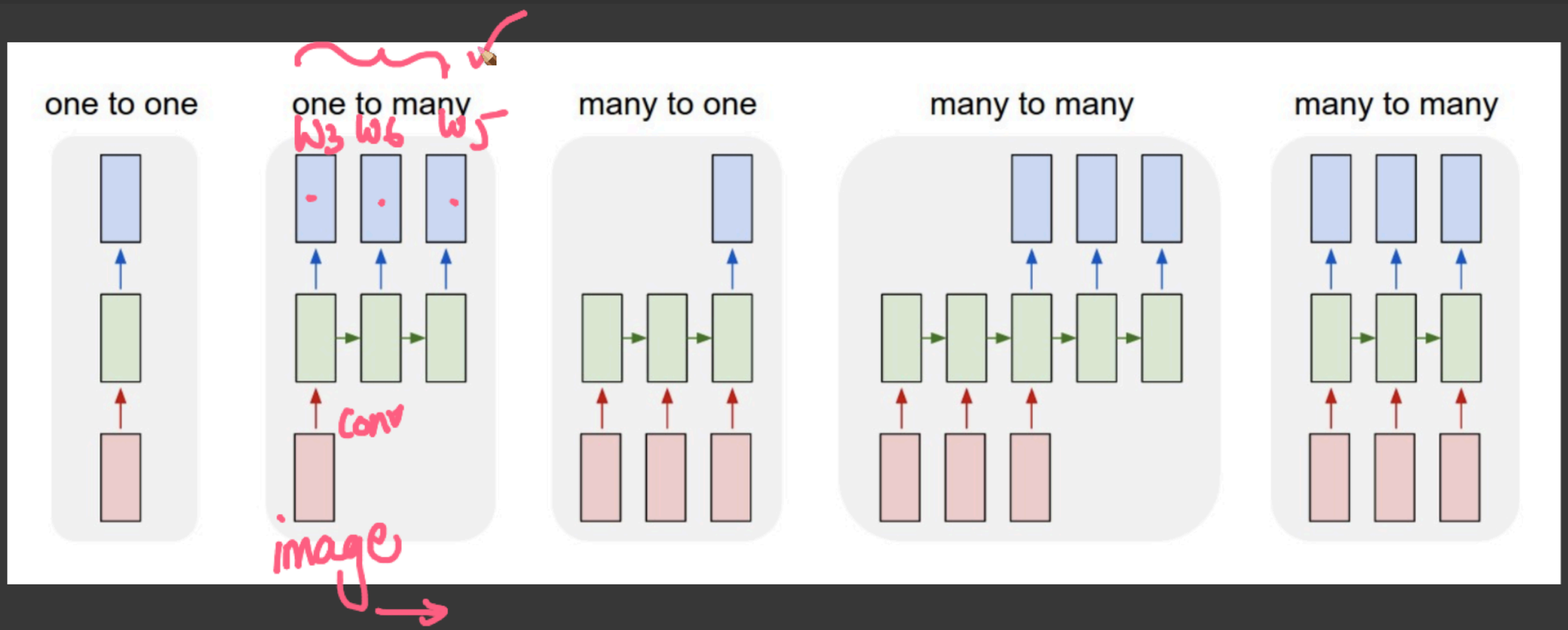


One to One: $T_x=1$, $T_y=1$

Example: Normal Feedforward Neural Network

65 / 65

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:57
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=D-B-sDCI4Y7B
+ Code + Text Last saved at 17:10 Connect



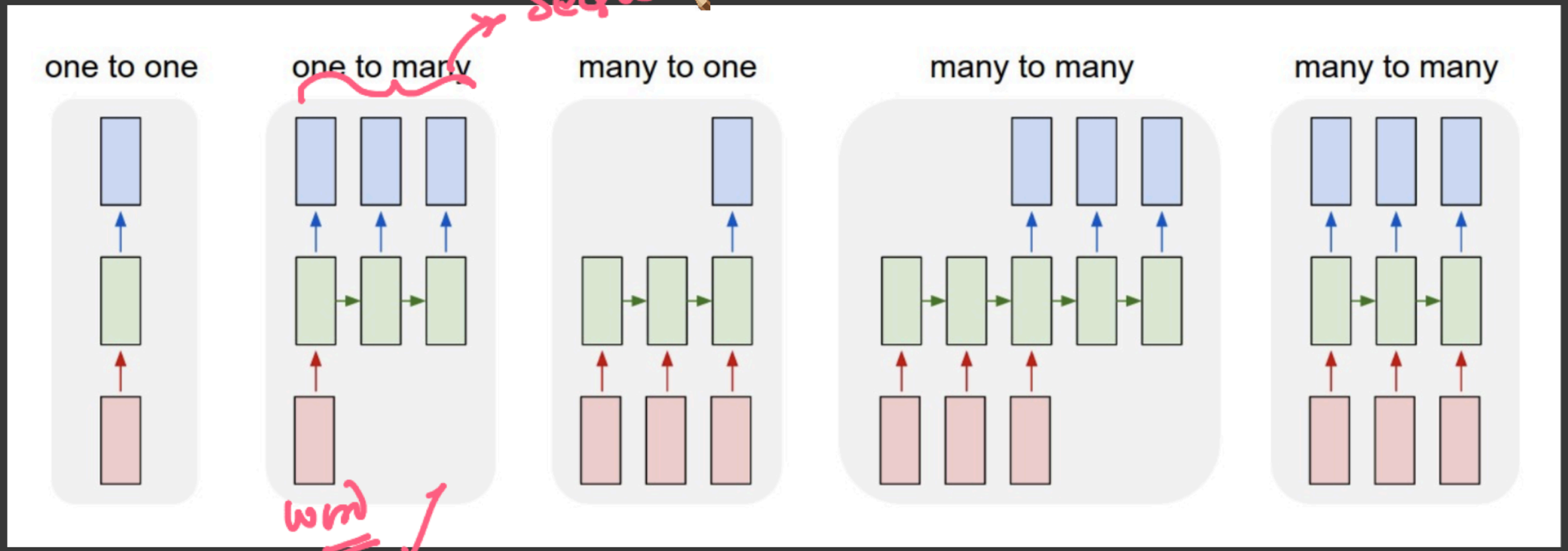
One to One: $T_x=1$, $T_y=1$

Example: Normal Feedforward Neural Network

One to Many: $T_x=1$, $T_y=1$

66 / 66

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:57
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=D-B-sDCI4Y7B
+ Code + Text Last saved at 17:10 Connect



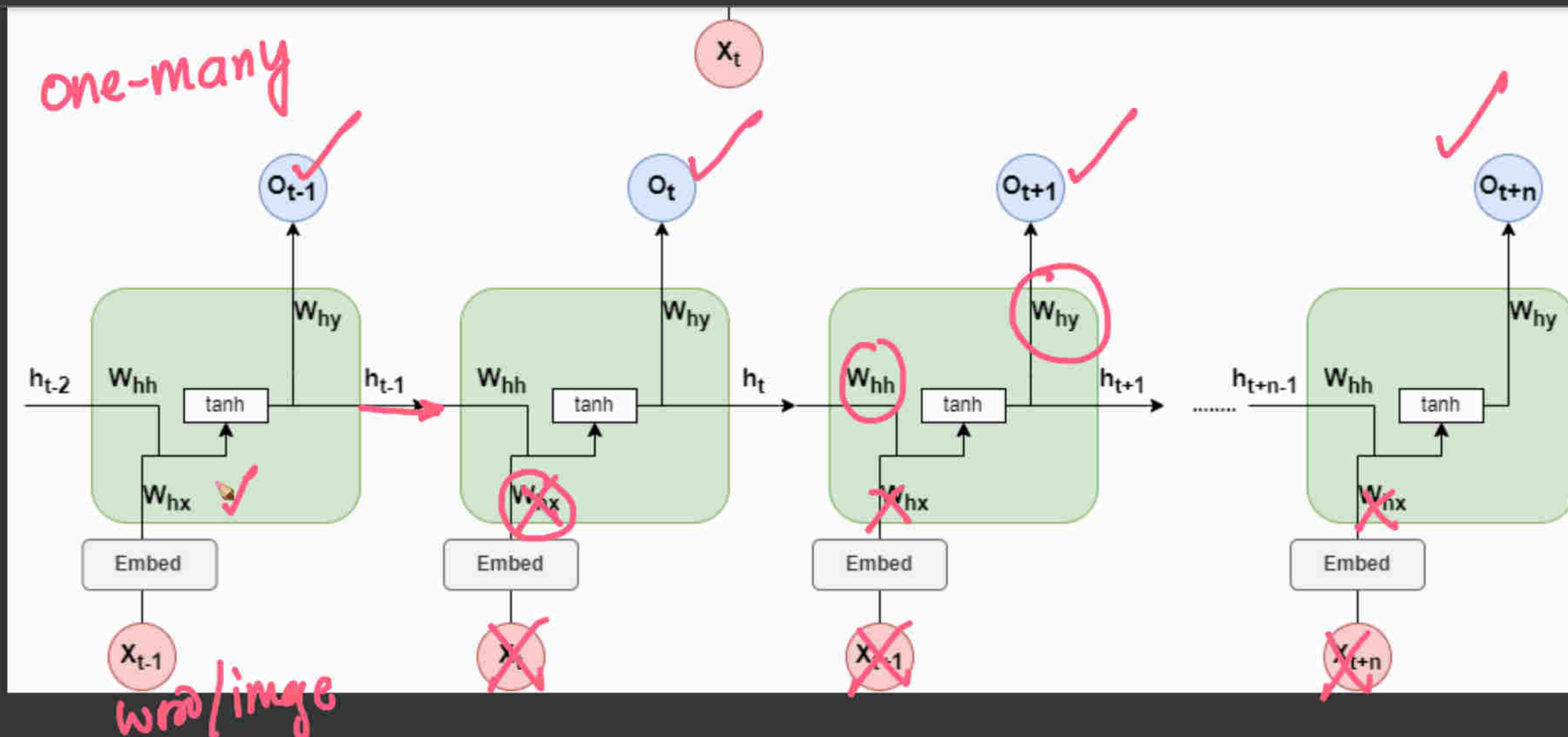
One to One: $T_x=1$, $T_y=1$

Example: Normal Feedforward Neural Network

One to Many: $T_x=1$, $T_y>1$

67 / 67

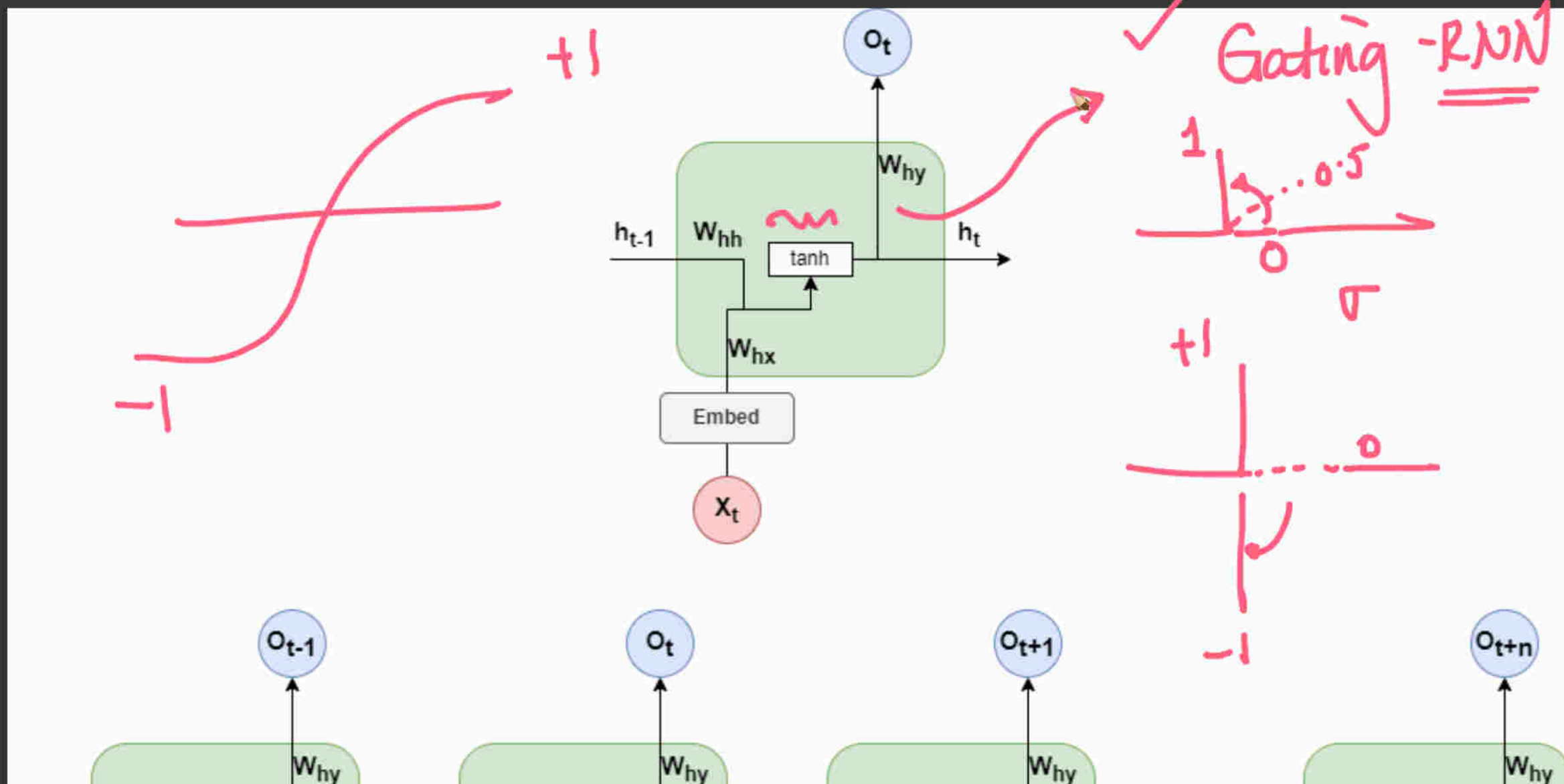
Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:58
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=D-B-sDCI4Y7B
+ Code + Text Last saved at 17:10 Connect



How Forward propagation is different in RNN?

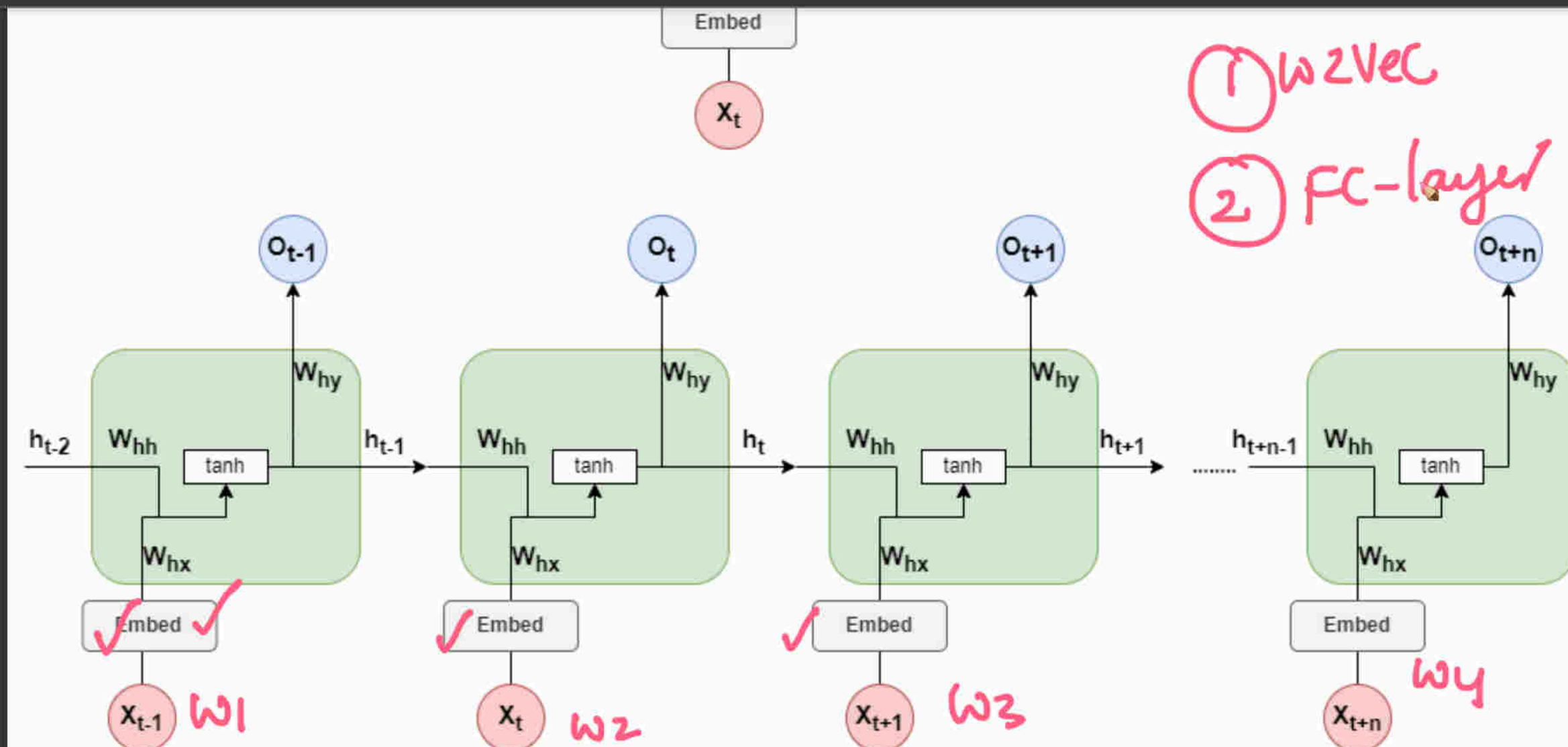
68 / 68

Chrome File Edit View History Bookmarks Profiles Tab Window Help 11:00
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=D-B-sDCI4Y7B
+ Code + Text Last saved at 17:10 Connect



69 / 69

Chrome File Edit View History Bookmarks Profiles Tab Window Help 11:01
Vanilla-RNN_V1.ipynb - Colab | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33IM_MOtMTT6S#scrollTo=D-B-sDCI4Y7B
+ Code + Text Last saved at 17:10 Connect



70 / 70

Chrome File Edit View History Bookmarks Profiles Tab Window Help 11:03
Vanilla-RNN_V1.ipynb - Colabo | 1506.00019.pdf | Fundamentals of Recurrent Ne | tanh derivative - Google Searc | What is the derivative of tanh(x
colab.research.google.com/drive/1PLZJRmC9CwEOSmkrulw33lM_MOtMTT6S#scrollTo=D-B-sDCI4Y7B
Update
+ Code + Text Last saved at 17:10 Connect

```
Title: Federer joins all-time greats
Category: sport
```

Why MLP would not perform well for this problem?

Before moving to MLP, let's understand what is temporal/Sequential data

Temporal component: Temporal means related to time. As we speak or write we do it in a sequence over time and it may not make sense when we change the order. Such type of data is called sequential/temporal data.

Other examples: Daily Stock price, Video

MultiLayer Perceptron

Let's first build a MultiLayer Perceptron to train a FeedForward NeuralNetwork to classify the given sentence into any of the 5 categories.

For the sake of understanding let's look at one example

71 / 71