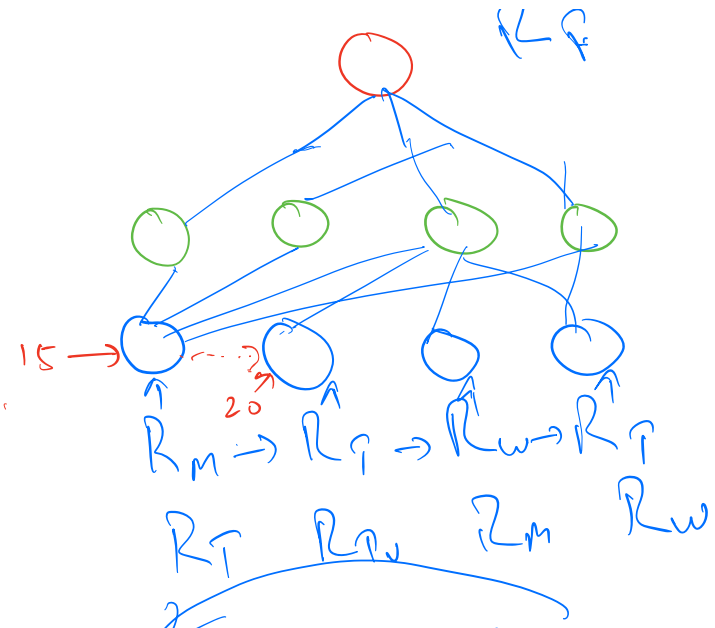
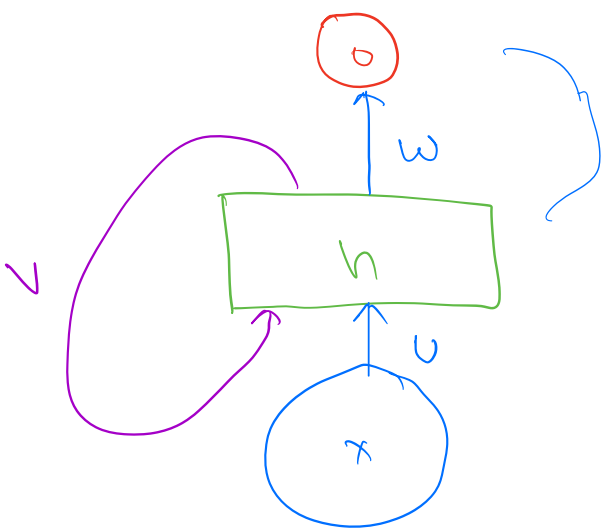


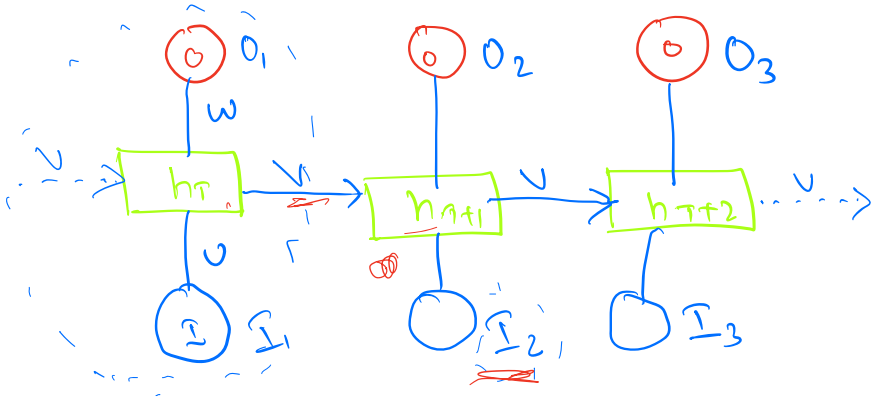
R_M 15cm
R_T 20cm
R_W X
R_T 10

$R_T \; R_W \; R_M \; R_W$

A cat, which is black, is hungry.



Unfolded RNN

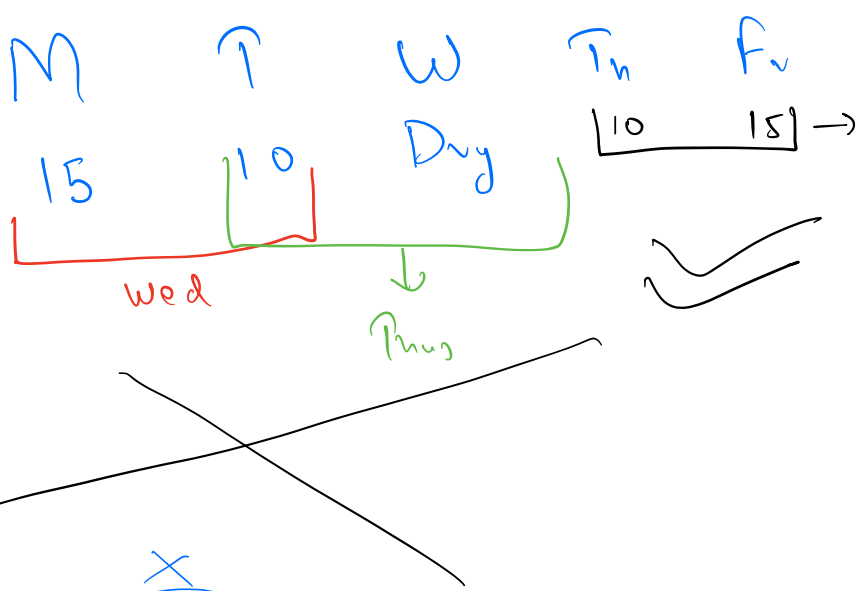


Output is a function of hidden state

hidden state is a func of input

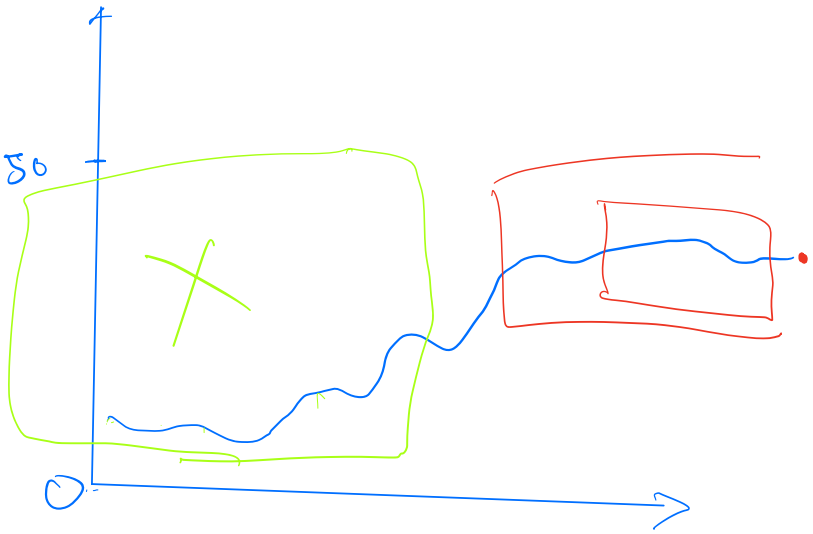
$O_1 \rightarrow fun(h_t)$ $O_2 \rightarrow fun(h_t + h_{t+1})$

$O_3 \rightarrow f(h_{t+1} + h_{t+2})$



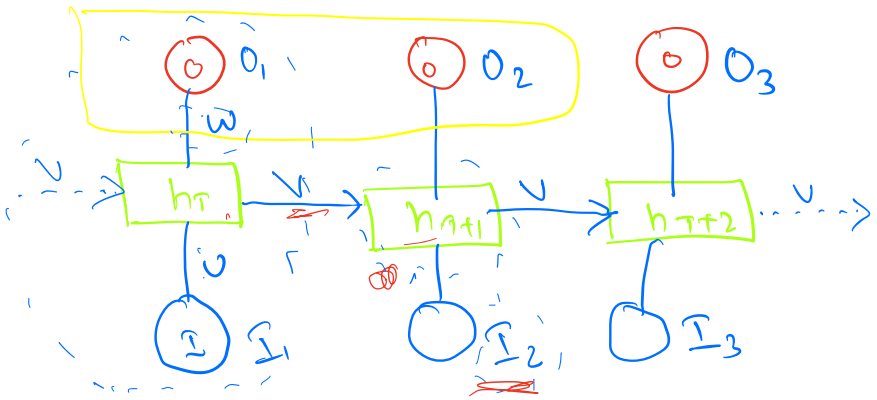
Ram is a male. Sita is a female.

1	15
2	20
3	12
.	.
.	.
.	.
.	.
.	.
.	.
27	30
28	45
29	43
30	38



Ram, who is walking in a green park, is thirsty.

Ram is thirsty.



$I_{(t)}$ = Given

$O_{(t)} = Softmax(W\, h_{(t)} + b_o)$

$h_{(t)} = f(U \cdot X_{(t)} + V\, h_{(t-1)} + b_h)$

Input — hidden state of previous t.s.

Loss functions

Reg → MAE, MSE, RMSE

Classification → Binary Cross Entropy

" Category "

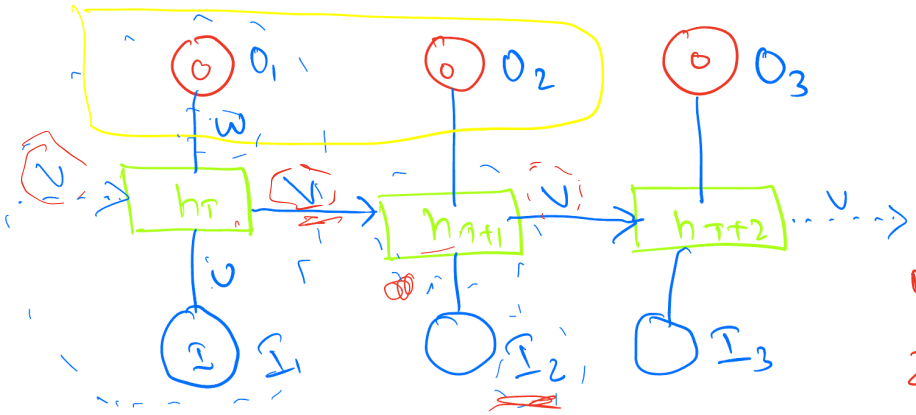
$$U = U - \frac{\partial O}{\partial U}$$

$$V = V - \frac{\partial O}{\partial V}$$

$$W = W - \frac{\partial O}{\partial W}$$

$$0 = \sum_{i=0}^{n} O_n$$

Drawbacks of Rnn



Weight with $h_t \rightarrow V$

$V = 2$

Network 50 ts

2^{50} = Extremely Large

0.2^{50} = too low

1) Exploding Gradients
2) Vanishing Gradients

Purpose of weight (V).

In simple language, is to make sure dependencies between data is learned properly.

LSTM

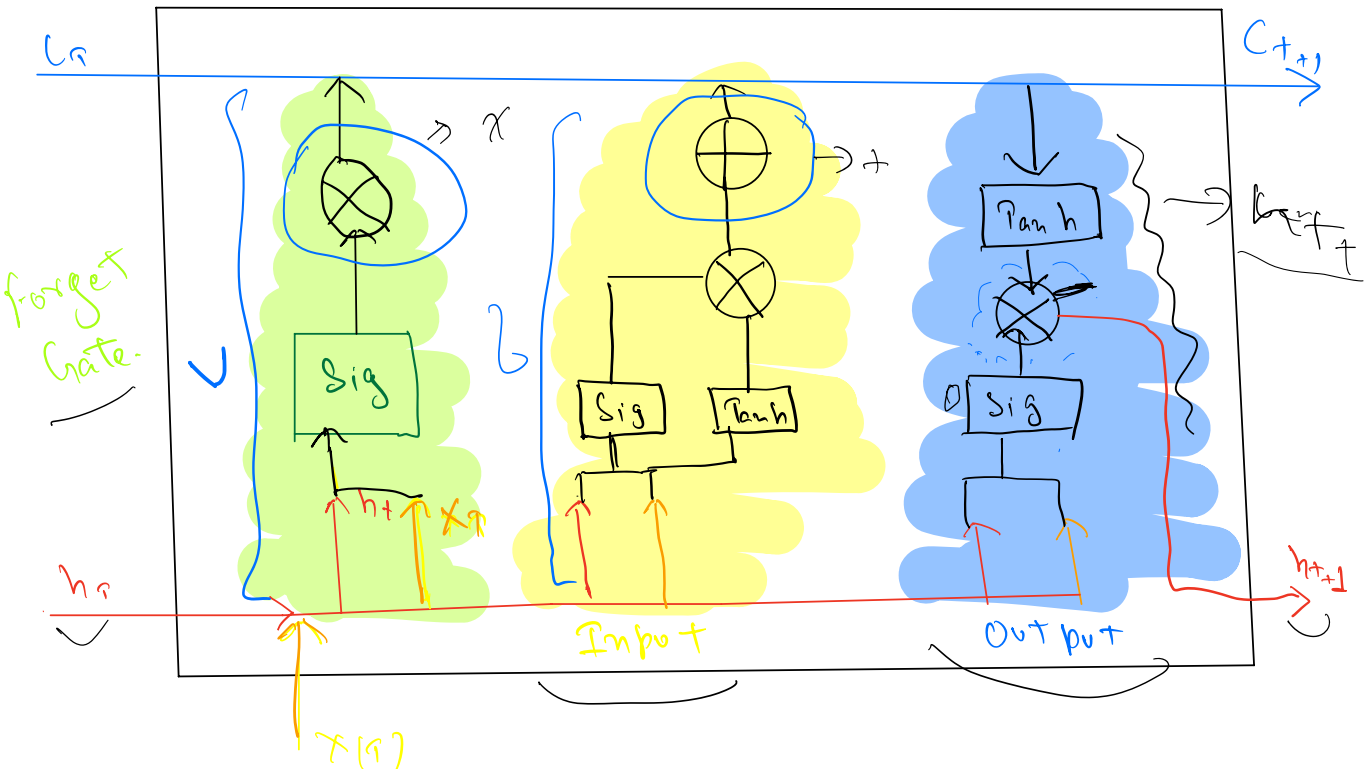
Long Short Term Memory.

Long Term memory ⇒ Taken care by cell state
⇒ Cell state updated by GATES

Short Term memory ⇒ Similar to what is in RNN
⇒ Updated by GATES.

Albert Einstein (/ˈ~~aɪnstaɪn/ EYEN-styne;[4] German: [ˈalbɛɐt ˈʔaɪnʃtaɪn] (listen); 14 March 1879 – 18 April 1955~~) was a German-born theoretical physicist,[5] widely acknowledged to be one of the greatest and most influential physicists of all time. Einstein is best known for developing the theory of relativity, but he also made important contributions to the development of the theory of quantum mechanics. Relativity and quantum mechanics are the two pillars of modern physics.[1][6] His mass–energy equivalence formula E = mc2, which arises from relativity theory, has been dubbed "the world's most famous equation".[7] His work is also known for its influence on the philosophy of science.[8][9] He received the 1921 Nobel Prize in Physics "for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect",[10] a pivotal step in the development of quantum theory.

MR. A was good friend of Albert. He actually taught him ML.

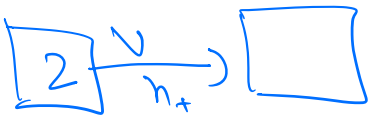


Input gate

Tanh varies between -1 to 1.

~~(~~$(x_t + h_t)$ is fed to create a potential long term memory.

Sigmoid (0, 1)



$h_t \cdot V + b \Rightarrow 2 \times V + b$

→ Explosion

LSTM ?

→ RNN

→ Can handle long term dependencies better.

→ Retain info for longer.

LSTM Architecture

> Designed for long term dep.

> Introducing Memory state / Cell state

> h_t there to handle short Term dep.

> Gates to update these long & short states.

3) How cell state is updated

By multiplying by output of F.G and adding input of I.G.

Rnn Vs LSTm

> Rnn cant hand long term dep in data as good as LSTM.

> When handling LTD ~~is~~ RNN encounters Vanishing & Exploding Gradients, while LSTM dont.

> LSTM is a RNN.

What does F.G do?

Decides how much long term memory to remember.

what does I G do?

Decides ~~How~~ what needs to be added or updated.

What is output rate?

OG generate new hidden state and makes sure it is in sync with LSTM.