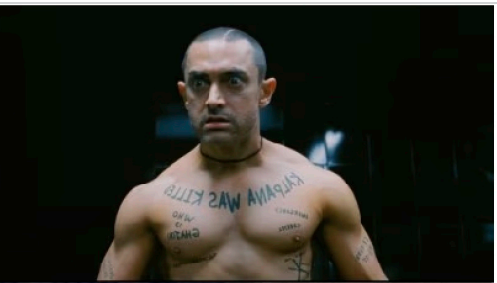
Session 07

LSTM

April 26, 2024



RNN

LSTM

Agenda

→ 11:15

1. Data Shape in RNNs, multi-layer RNN
2. Encoder Decoder Architecture
3. LSTM
4. GRU

→ Ready used in industry

Transformer

Data Shape into RNN

text → ~ 100 words
~ 1K words
~ 10 words

target → multi-class

[Batch-size, no-of-words] → Fixed value

Product name → 6000 character level

[n rows, 6000] → [n, 6000, 128] (dimension of embed.)

200

64 gb Ram

128 gb Ram

median length → loose some data

loss → minimal

99.7%

Relay + hindi En

↓

V.V good Ans

LSTM

<PAD>

[I love lasagna <PAD> <PAD>]

Before → [1000, 10, 128]

word → 128 emb

[1 , 10, 128]

b.size

embedding

max length

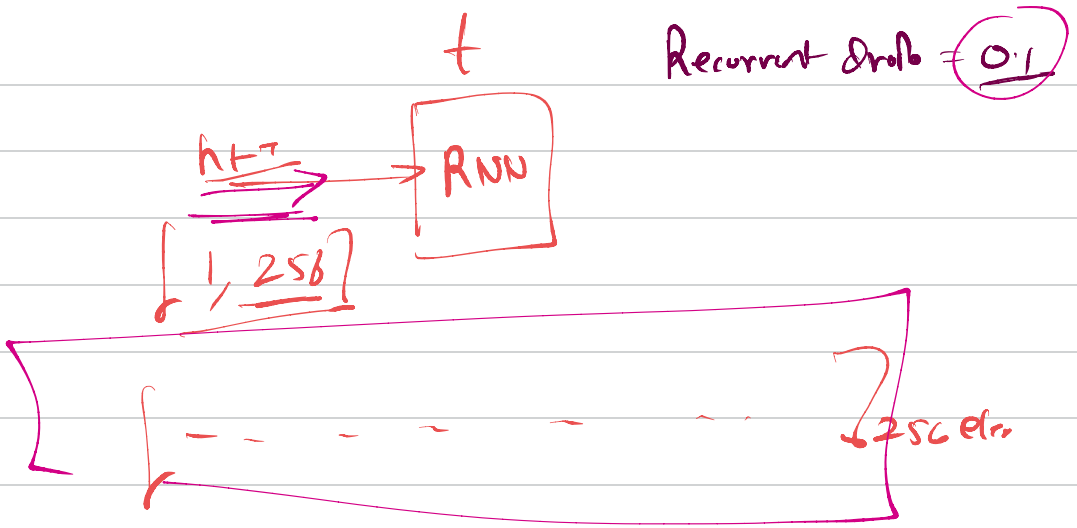
t=1 → t=2 → t=10

↑ 128d

↓ 128

27

↓ 128

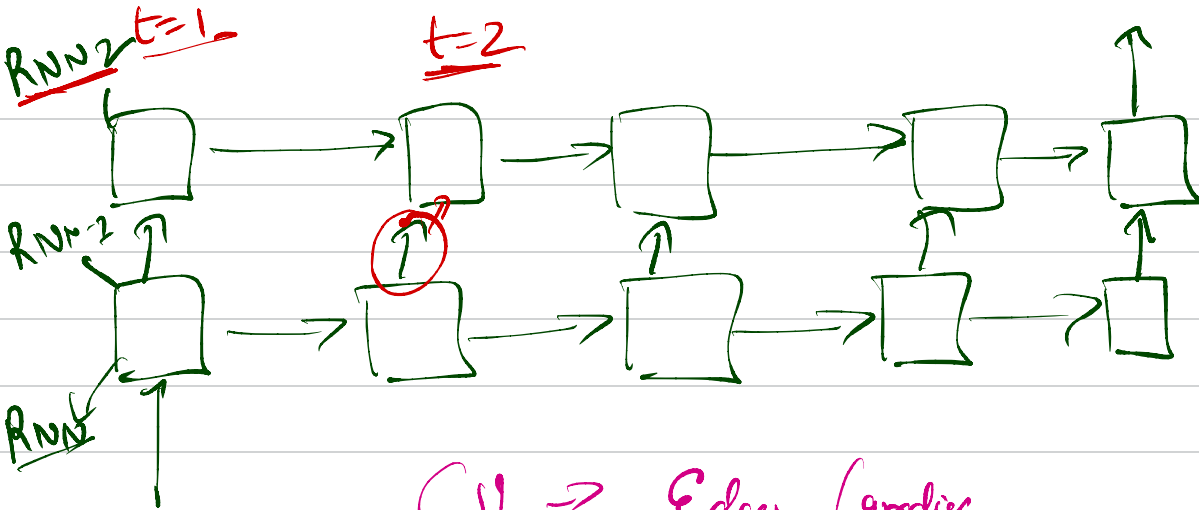


Recurrent dropout = 0.1

For each time step → the same random 10% of elements in hidden state will → dropout

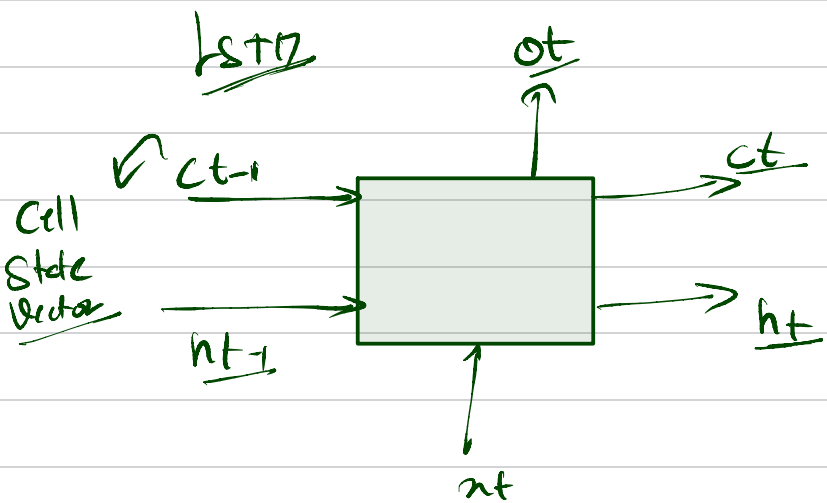
R-d = 0.1

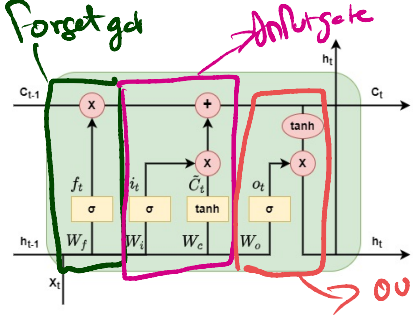
t = 1 ... t

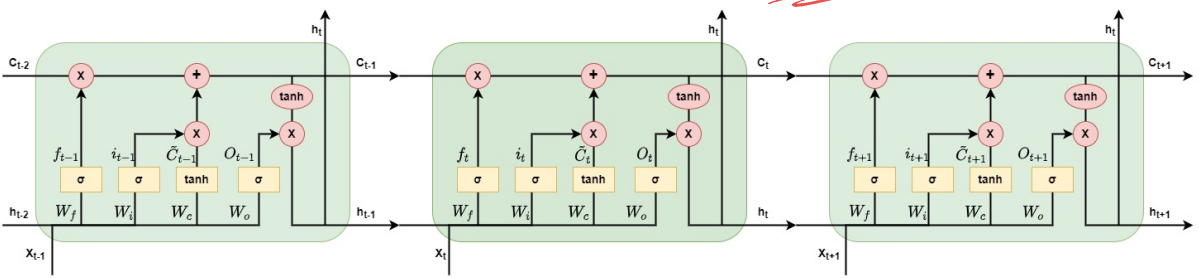


CV -> Edges / gradient
Parts of objects

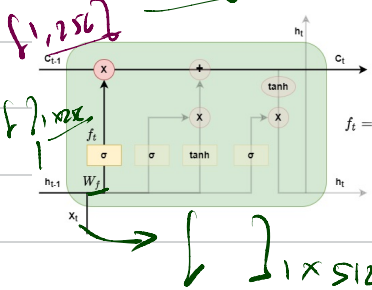
RNN -> L1 - Small range depends
L2 -> Bigger Phrase interest







Forget gate



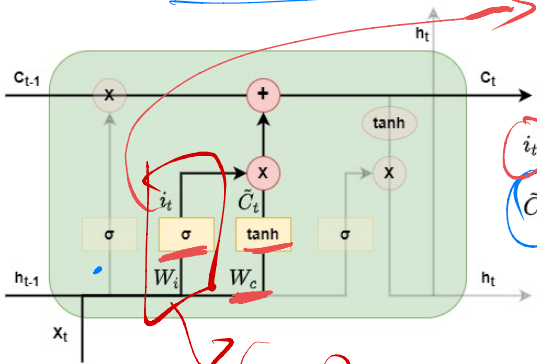
what information in "cell state" should retain or discard

[1, 256]
[1, 512] ↓ axis = 0
= [1, 768] × [768, 256] = [1, 256]
or

[5, 6, 8, 10] ⊗ [0, 0, 1, 2] = [0, 0, 8, 20]

[1, 256] → 0 ~ 1

INPUT GATE



controlling how much of $\tilde{C}_t$ to be added to C_t

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$$

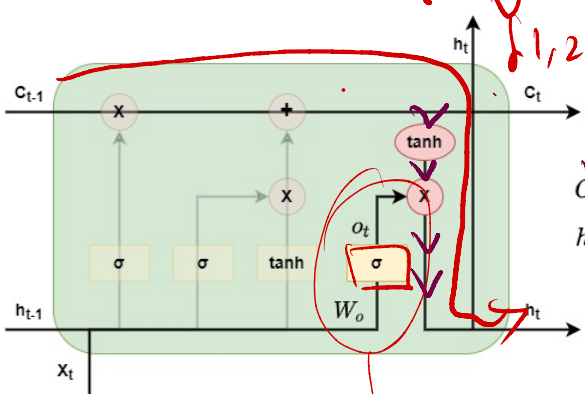
$$\tilde{C}_t = tanh(W_C * [h_{t-1}, x_t] + b_c)$$

Potential cell state vector

(0-1)

on/off switch

Output Gate



what information should be passed to next hidden state

$[1,256] \times \theta^5 \rightarrow [256,5]$

on/off switch

$$O_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$

$$h_t = O_t * tanh(C_t)$$

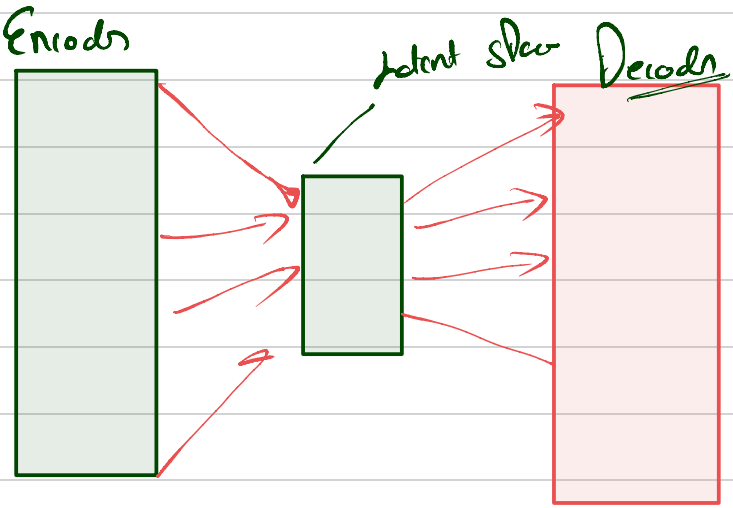
(0-1)

on output of cell state vector

"how much"

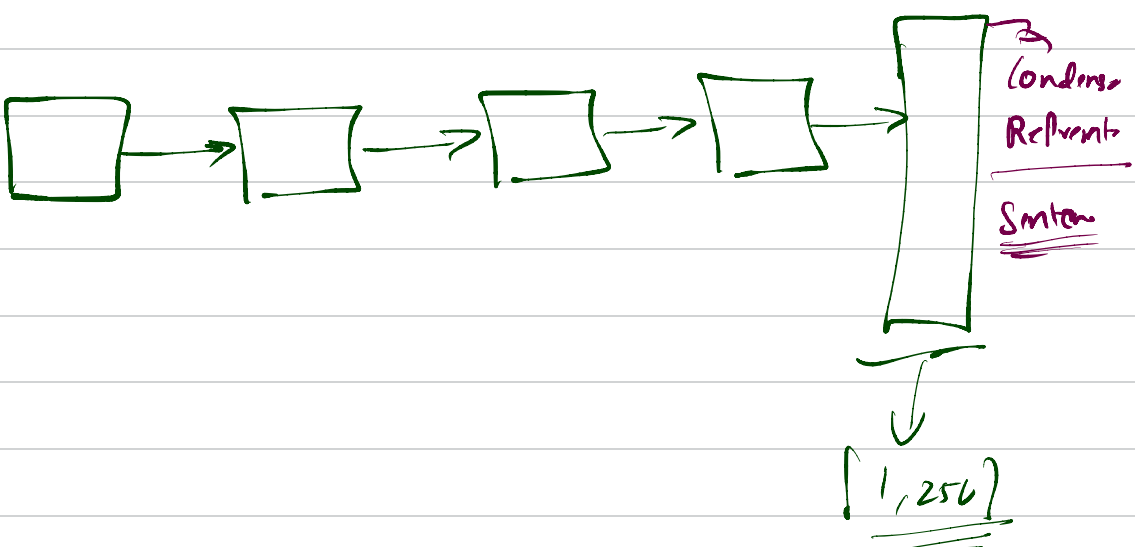
$$C_t = F_t \times C_{t-1} + i_t \times \tilde{C}_t$$

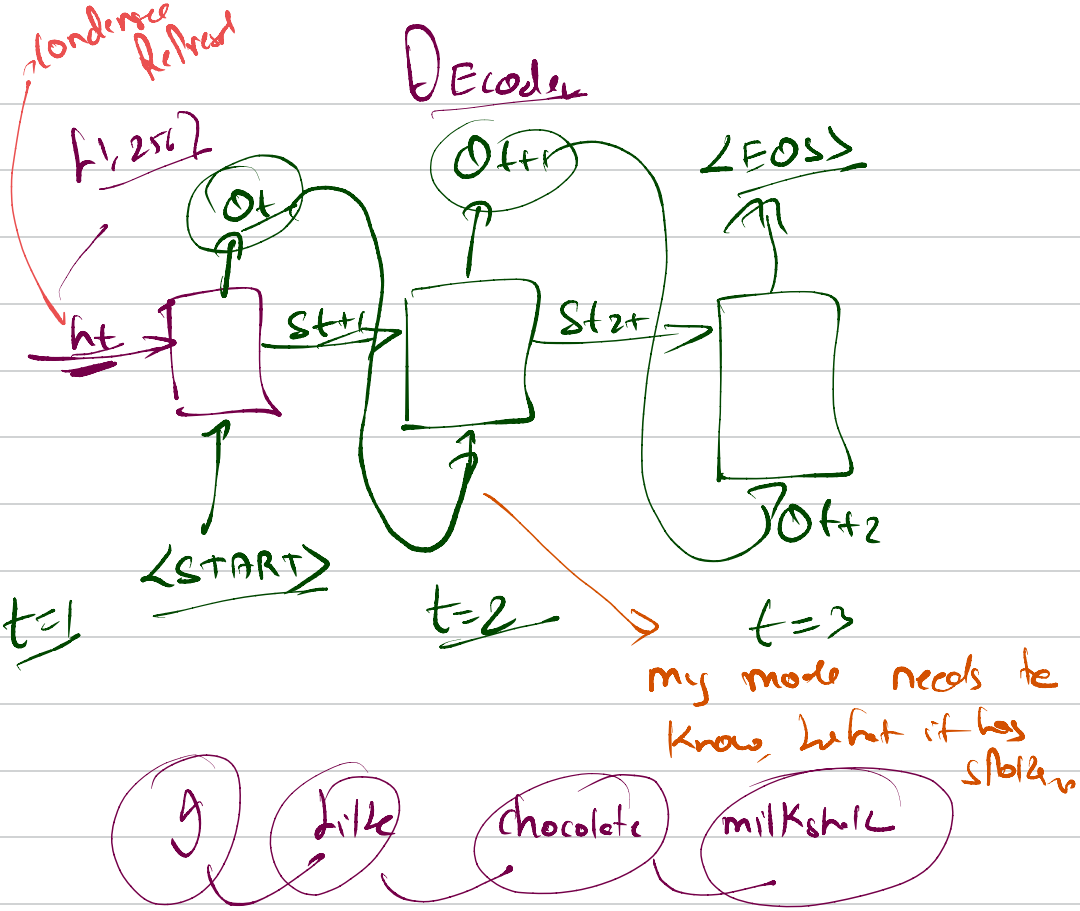
Encoder - Decoder



"I loved the Roller-coaster at Wonderla"

t1 t2 t3 t4 t5 t6




condense Refresh
Decoder
[1, 256]
Ot
Ot+1
<EOS>
ht
St+1
St2+
Ot+2
<START>
t=1
t=2
t=3
my model needs to know what it has
I
like
chocolate
milkshake