this phone is no good - Negative

NO, the phone is good - Positive

	good	phone	no	is	
DOC1	1	1	1	1	→ Neg
DOC2	1	1	1	1	→ Pos

Sequence is really important

I went to germany and people speaks german there

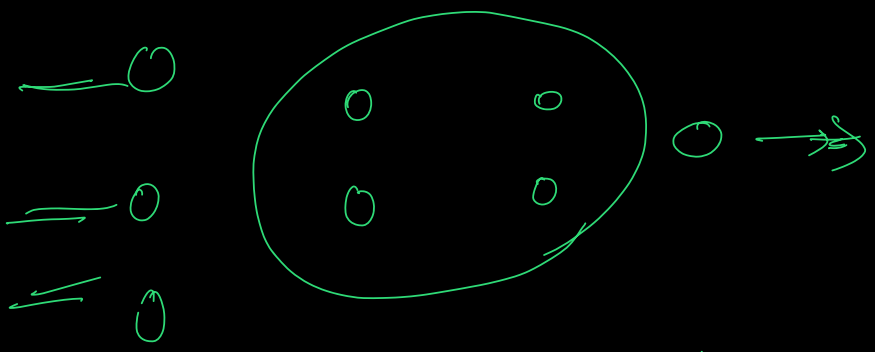
Un Sorted → good like I and that is it ?

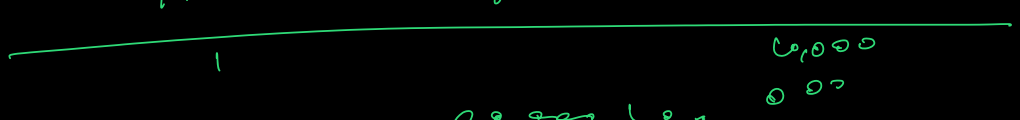
I like that and it is good and ?tasty

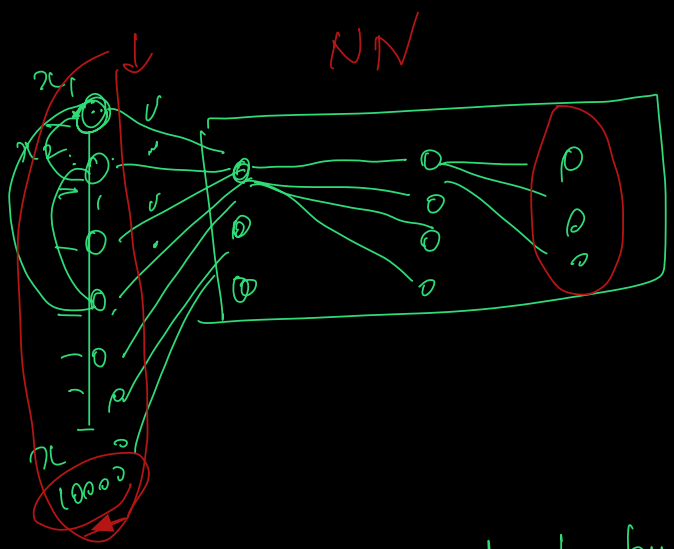
Sequence really matters

gmail → Hi! How are you?

whatsapp → |







I went to market to buy shoes on sunday

On sunday, I went to market to buy shoes

On shoes, I went to sunday to buy market

The feature don't share across different position

Machine translation

i/p → o/p

sentence → sentence

dynamic → dynamic

Input, Outputs can be different length in different example

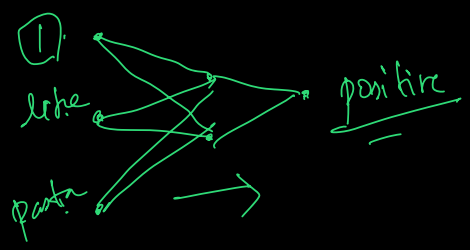
Recurrent Neural Network

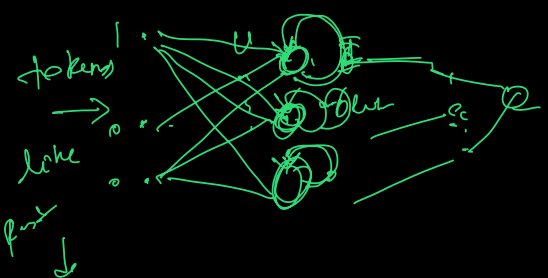
1. Sequencing
2. Memory
3. dynamic length of I/P and O/P

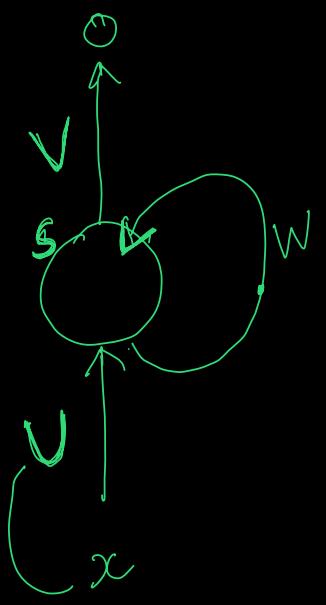
ANN / NN

FFNN

RNN



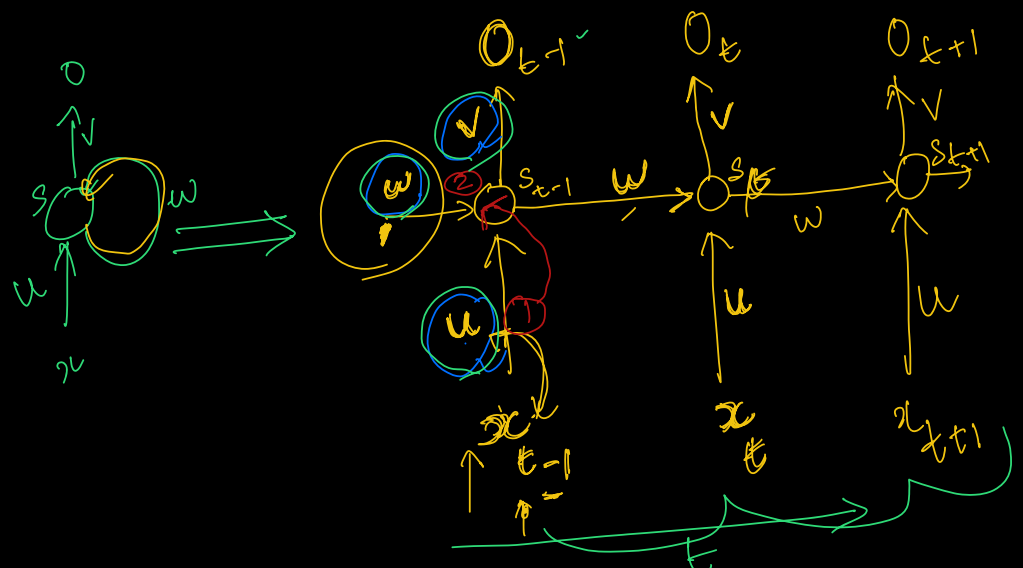




x - Input
O - Output

U → Weights from Input
V - Weights for Output
W - Weights for previous hidden state

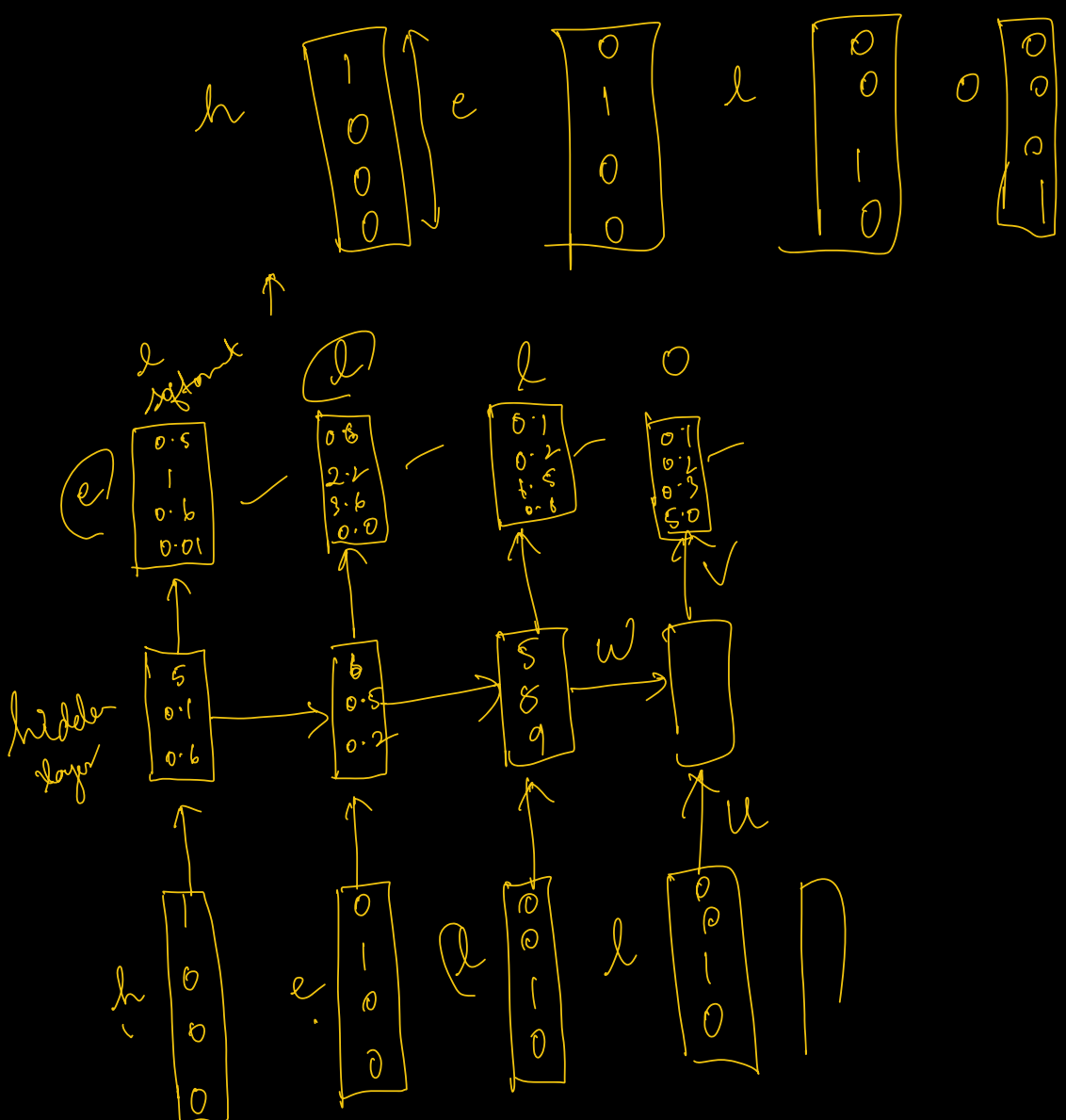
S - State of hidden Unit



Prediction of next character

'h', 'e', 'l', 'l', 'o'

(4 vocab) h e l o


h
e
l
o
softmax
hidden layer
W
V
U

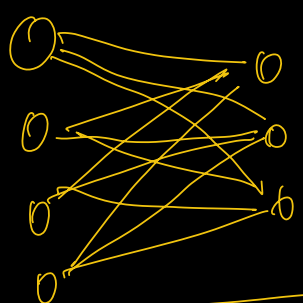
Math behind RNN

1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
h	e	l	o

x_{t-1} x_t x_{t+1} x_{t+2}

input neuron multiply by weights of randomly allocated, (3) neuron

U $\Rightarrow$ 3 × 4



Step 1

$$x_t = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

U			
0.2	0.8	0.5	0.6
0.1	0.08	0	0.5
0.5	0.8	0.02	0.1

3 × 4 · 4 × 1

x_t = [1, 0, 0, 0]

$U * x_{t-1} =$ [0.2, 0.8, 0.1]

W ⇒ 0.4893 1 X bias = 0.5

S_{t-1} initialize ⇒

0
0
0

S_{t-1}

Step 2

W b

0.4893 + 0.5 X

0
0
0

⇒

0.5
0.5
0.5

$W * S_{t-1} + bias$

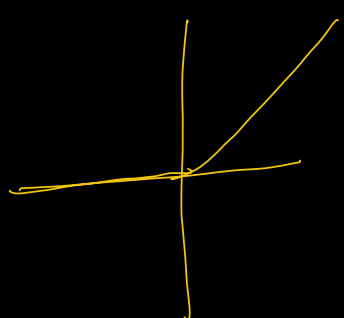
S_{t-1}

$$S_t = \tanh\left(W * S_{t-1} + U x_t\right)$$

Step-3

tanh (| 0.5 | + | 0.2 |
|---|---|---|
| 0.5 | | 0.8 |
| 0.5 | | 0.5 |

)



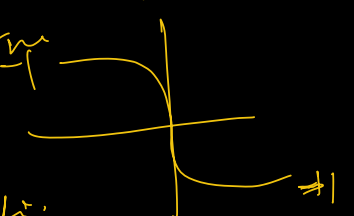
S_t ⇒

0.6
0.8
0.2

Past info

Karna

hidden info.



$V =$

6.5	6.2	1.2
0.2	0.1	0.0
2.6	1.5	0.6
5.0	0.9	0.8

Step-k

3 neuron output 4

$$O_k = V \cdot S_k$$

$=$ (4×3) × (3×1)

0.6
0.8
0.2

$=$ Softmax $\Rightarrow$

1.00
2.5
0.9
1.2

0.1
0.4
0.2
0.3

e

e

$u \times$

0
1
0
0

$\Rightarrow$

3×1

Back Propagation

↓

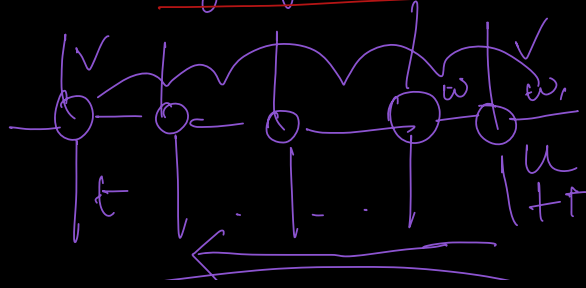
Adjust the weights (W, V, U)

Time Information (Sequence)

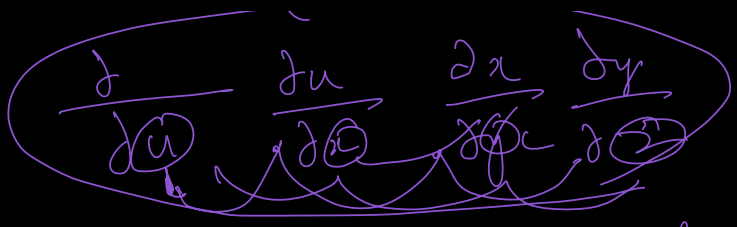
Back Propagation Through time (BPTT)

Problem with RNN

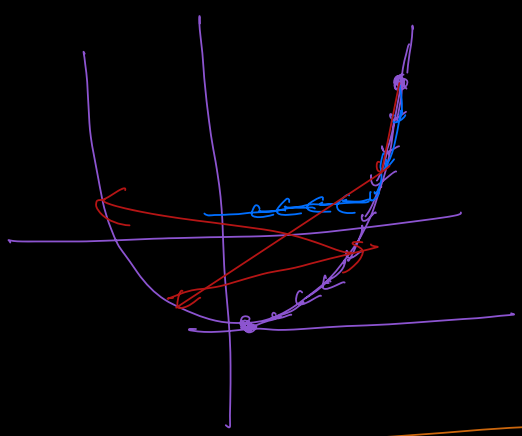
1. Time & Resource Intensive
2. Not effective in Long sequence (Limited memory)
   may forget some important information in the long history
3. Vanishing gradient descent



Chain reaction
Chain derivable



4. Exploding gradient descent

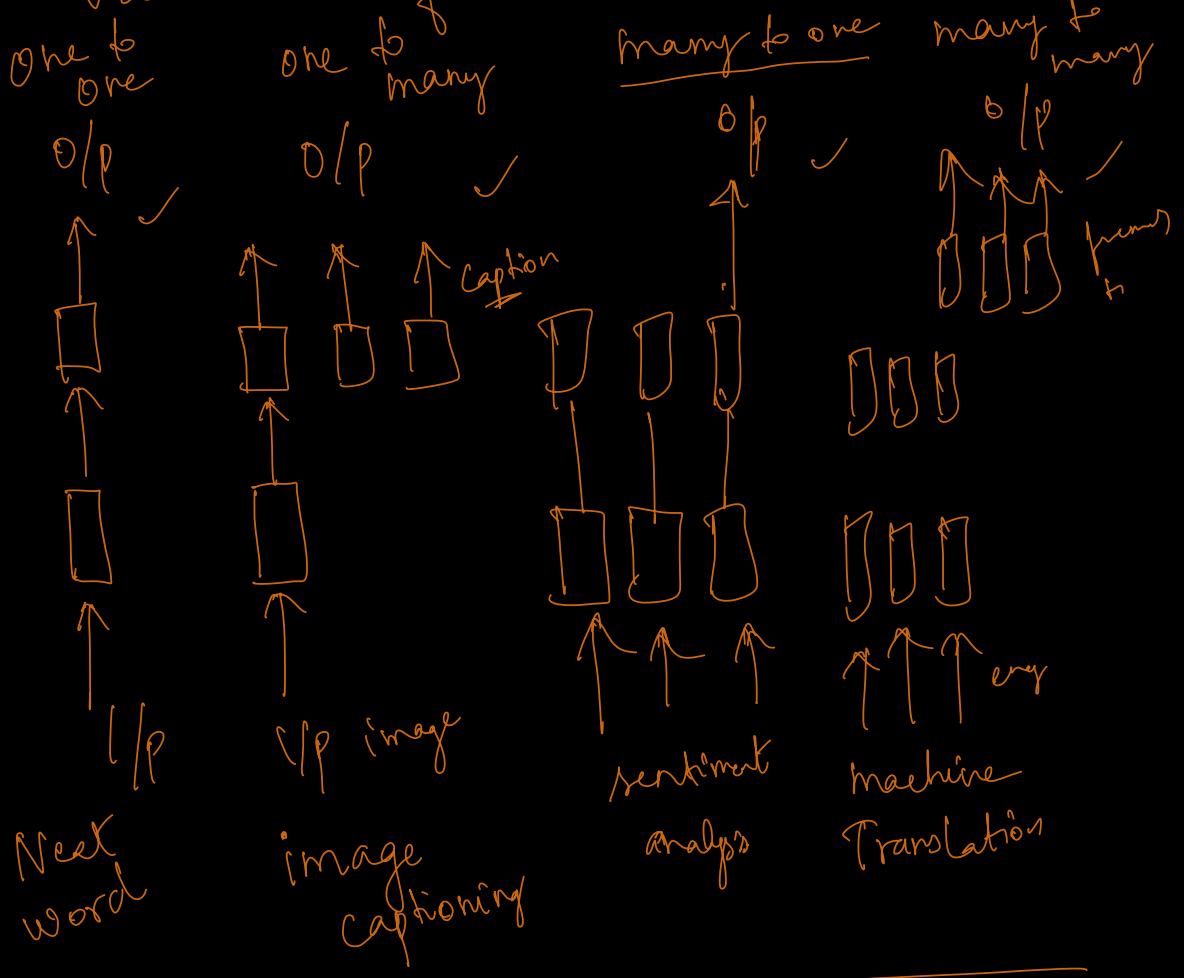


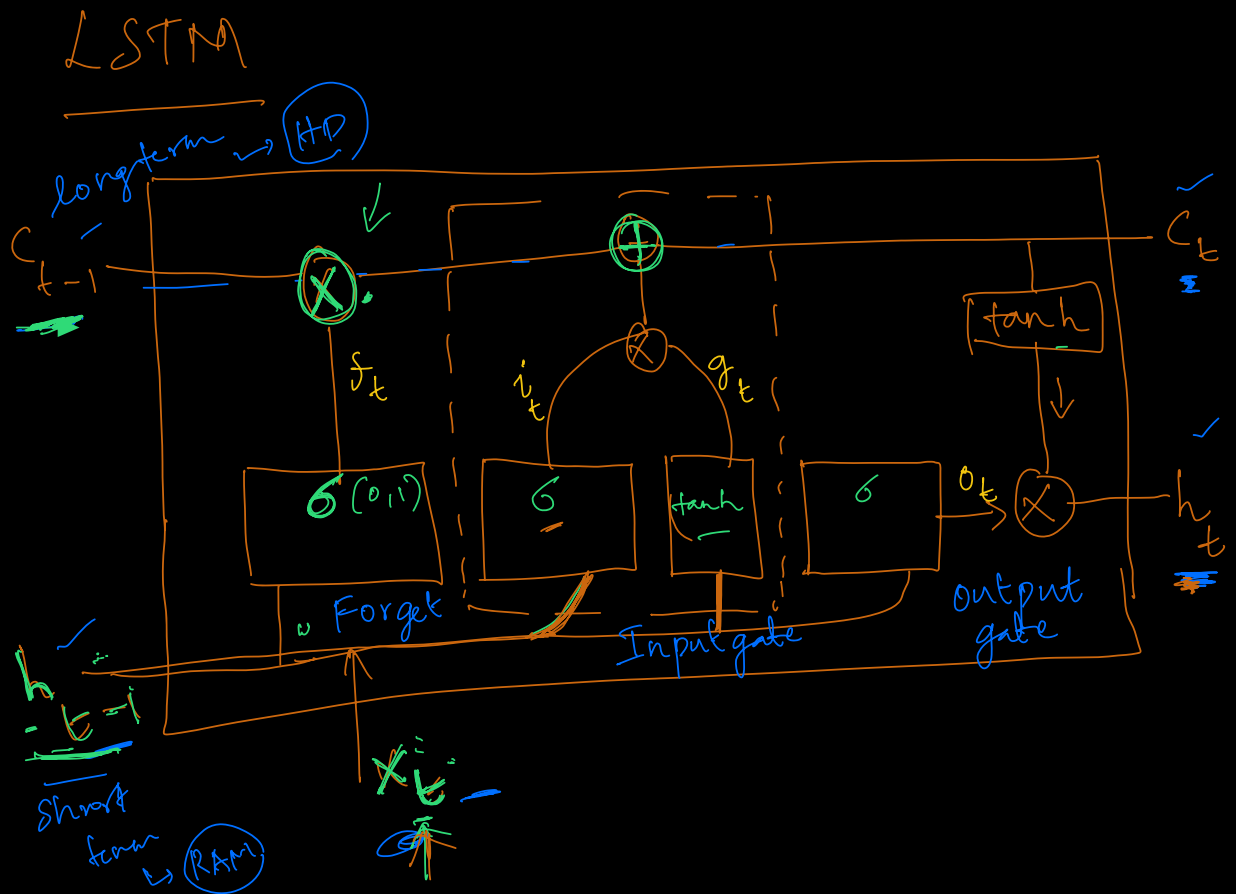
Long Short term Memory (LSTM) ← (RNN)

Gated Recuring Unit (GRU)

Transformer → Attention → GPT

Variants of RNN





C_{t-1} [5 6 7]

f_t [1 0 1]

long term updated [5 0 7]