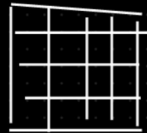


1) Tabular  $\Rightarrow$  Car, Num  $\Rightarrow$  ANN

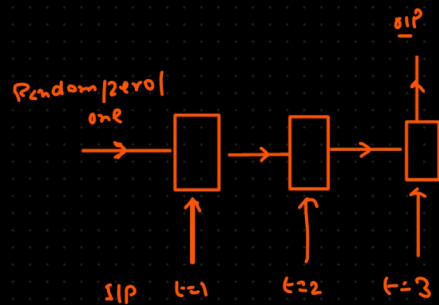


3) Sequence  $\Rightarrow$  Sound, text, speech, time series  
 $\uparrow$   
Input  
 $\times$  Output RNN  $\left\{ \begin{array}{l} \text{LSTM} \\ \text{GRU} \end{array} \right\}$

2) Image Data  $\Rightarrow$   $\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$   
2-D

CNN  $\rightarrow$  fully connected (ANN)

Input  $\Rightarrow$  RNN / LSTM / GRU  
mapping



Seq to Seq  $\xrightarrow{\text{(many-many)}}$  RNN / LSTM  $\Rightarrow$  (fix length vector) (long vector)  
Random/zero one  $\rightarrow$   $\left\{ \begin{array}{l} \text{Seq - input} \\ \text{Seq - output} \end{array} \right\}$  Machine translation

2014  $\Rightarrow$  Encoder-Decoder { seq-seq learning }

RNN / LSTM (seq to seq) fix-length

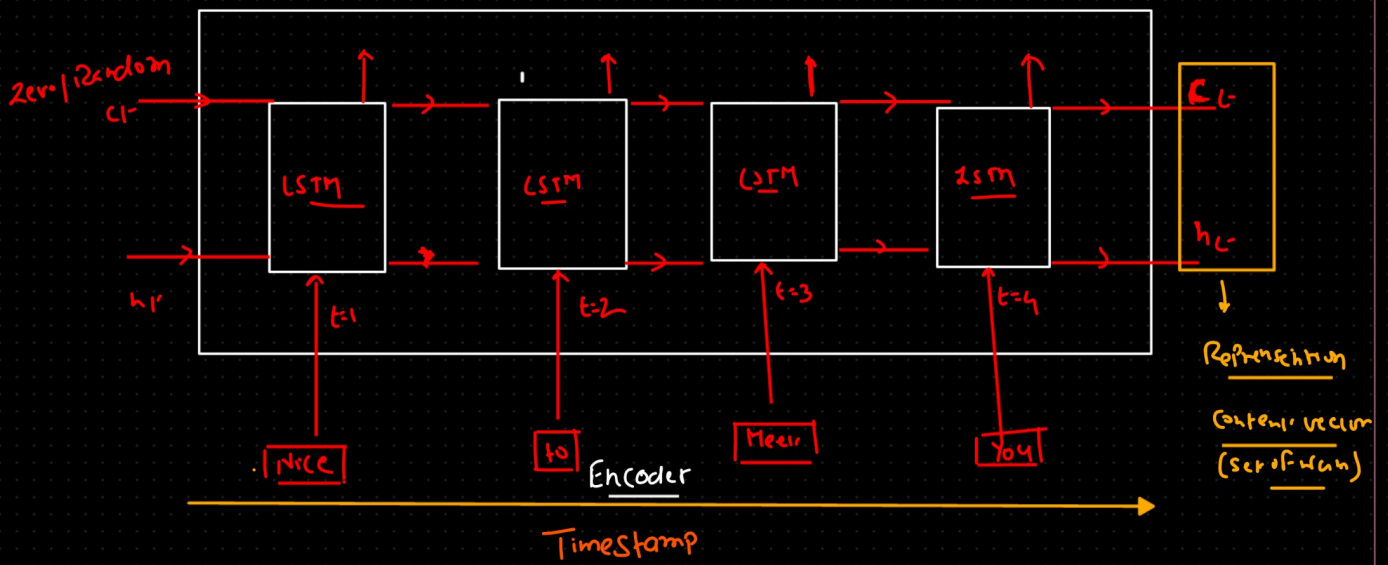
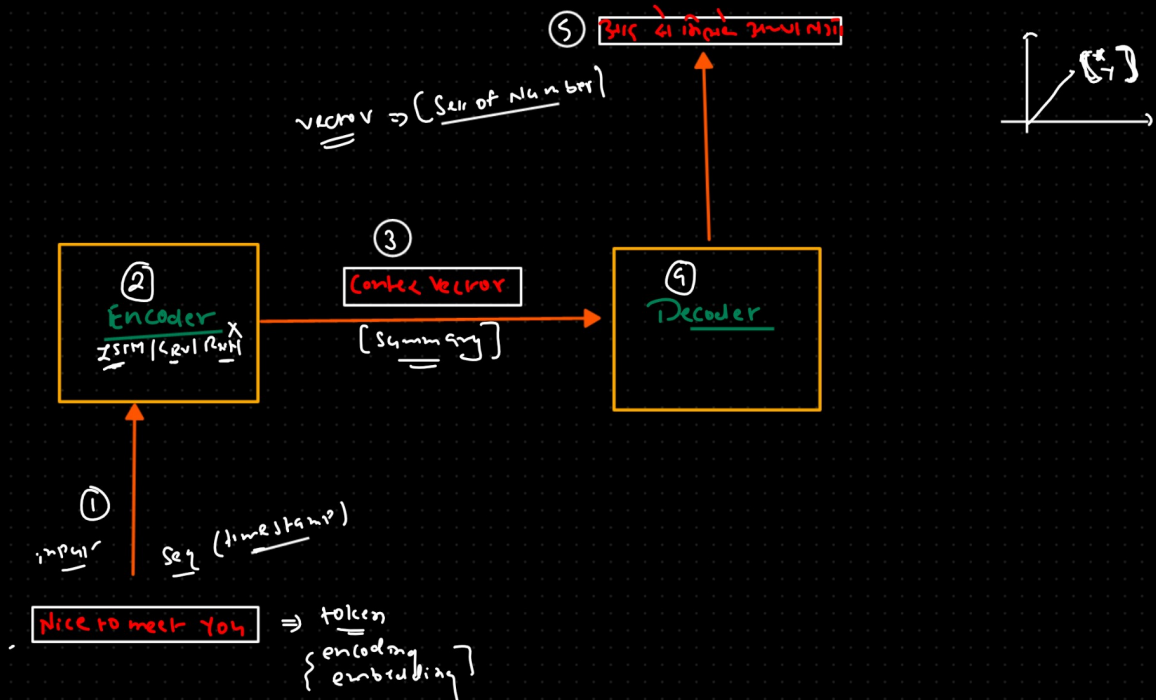
Machine Trans

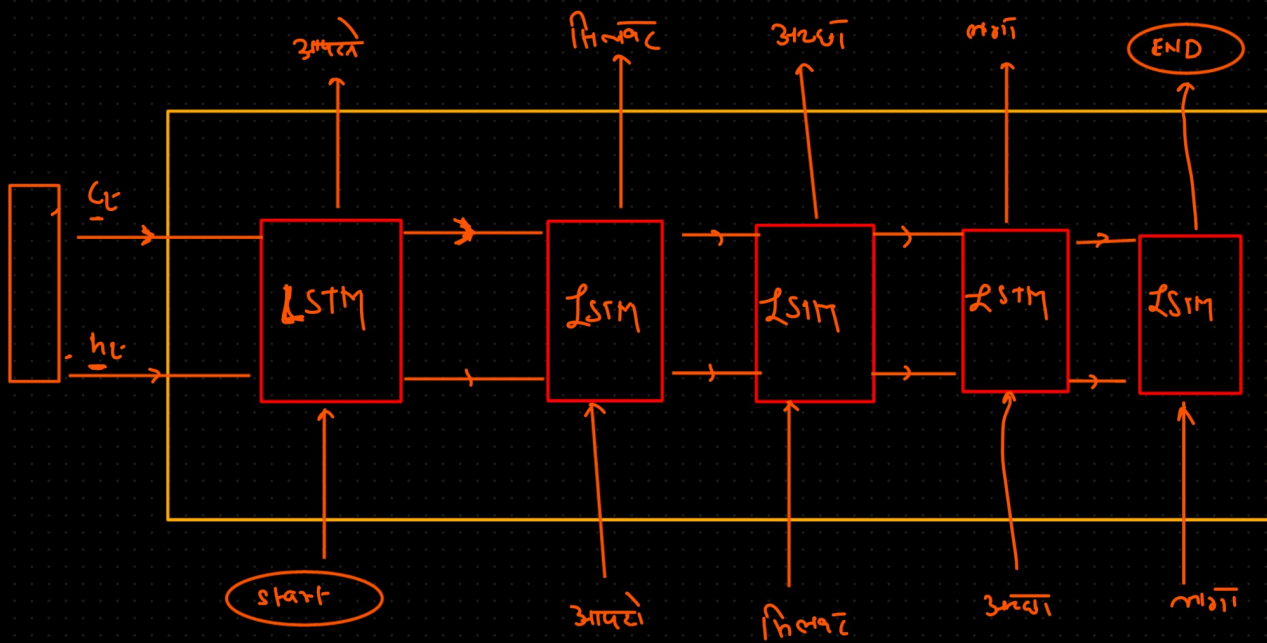
$\left\{ \begin{array}{l} \text{how to meet you} \\ \text{meis} \leftarrow \text{meis} \text{ mei} \text{ mei} \end{array} \right\}$

$\left\{ \begin{array}{l} \underline{\text{Input}} \Rightarrow \text{seq} \Rightarrow \text{enq} \Rightarrow \underline{\text{variable length}} \\ \underline{\text{Output}} \Rightarrow \text{seq} \Rightarrow \text{hnd.} \Rightarrow \underline{\text{variable length}} \end{array} \right\}$

Basic  $\rightarrow \{ \underline{\text{RNN}}, \underline{\text{LSTM}}, \underline{\text{GRU}} \}$

Bigger picture of encoder and decoder





- 1 Training (FP) 1 for encoder and decoder the training will happen together
  - 2 Loss (cal. cross entropy)
  - 3 Optimization (BP) (weight)
- Encoder-Decoder Machine Trans.
- 2 (Dataset)  
= Parallel

OHE

<u>Eng</u>	<u>Hindi</u>
1 think about it	आपके विचार
2 <u>come in</u>	आपके आने पर

↓ tokenization

<u>Eng</u>	<u>Hindi</u>
[ think, about, it ]	[ आपके, विचार ]
[ come, in ]	[ आपके, आने, पर ]

↓ Decoder

<u>Eng</u>	<u>Hindi</u>
1 start	start
2 = 7	end

Row  $\Rightarrow 1$

[think about it]  $\Rightarrow$  [चिंतन करो]

चिंतन करो

high

why I should wrong

end

multiclass (loss)

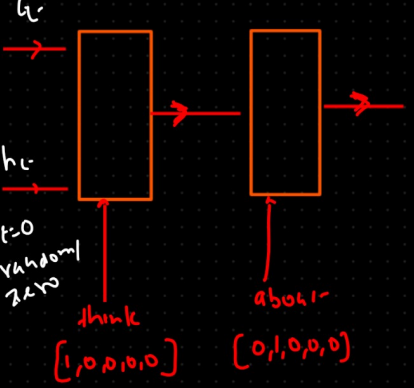
$y_{true}$  [0, 1, 0, 0, 0, 0, 0]  
 $y_{pred}$  [0, 2, 0, 0, 5, 0, 25, 0, 1, 0, 0]

[0, 0, 1, 0, 0, 0, 0]  
[0, 1, 0, 15, 0, 3, 0, 0, 0, 0, 4]

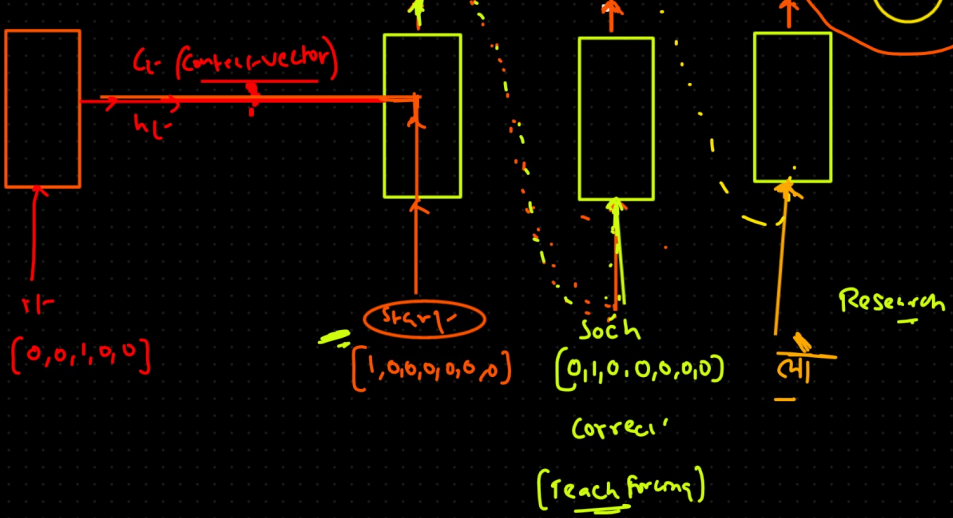
[0, 0, 1, 0, 0, 0, 0]  
[0, 1, 0, 15, 0, 3, 0, 0, 0, 0, 4]

SOFTMAX FUNCTION

(3.5)



[ENCODER]



[Teach forcing]

Eng

Hindi

$y_{true}$  [0, 0, 0, 0, 0, 0, 0, 1]  
 $y_{pred}$  [0, 1, 0, 2, 0, 2, 0, 1, 0, 3, 0, 1, 0, 4]

[0, 0, 0, 0, 0, 0, 0, 1]  
[0, 1, 0, 2, 0, 2, 0, 1, 0, 3, 0, 1, 0, 4]

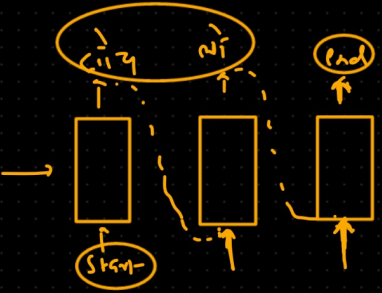
loss = 0

- 1 think about it
- 2 come in

चिंतन करो  
आइए आ जाइए

Forward Propagation

Loss



$y_{true}$  [0, 1, 0, 0, 0, 0, 0] सोच  
 $\hat{y}_{pre}$  [0, 2, 0, 1, 0, 5, 0, 2, 0, 1, 0, 3]

[0, 0, 1, 0, 0, 0, 0] नो  
[0, 1, 0, 15, 0, 3, 0, 5, 0, 0, 5, 0, 4]

[0, 0, 0, 1, 0, 0, 0]  
[0, 1, 0, 2, 0, 2, 0, 1, 0, 3, 0, 1, 0, 4]

Loss  $\Rightarrow$  Categorical cross entropy

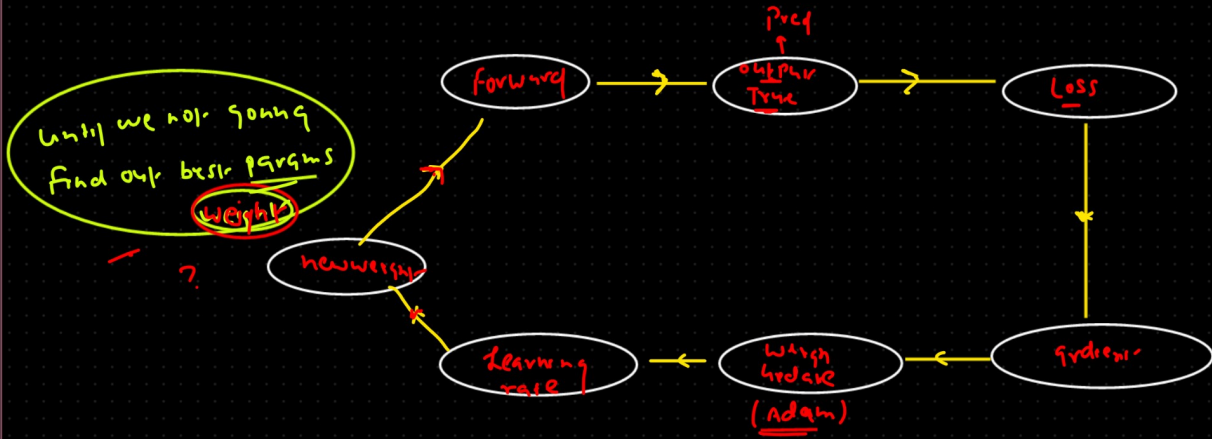
$$= - \sum_{i=1}^{n-1} y_i^{true} * \log(y_i^{pred})$$

$$Loss = (-1 * \log(0.1)) + (-1 * \log(0.15)) + (-1 * \log(0.4))$$

$$= 0.39 \text{ Loss value}$$



## Backpropagation



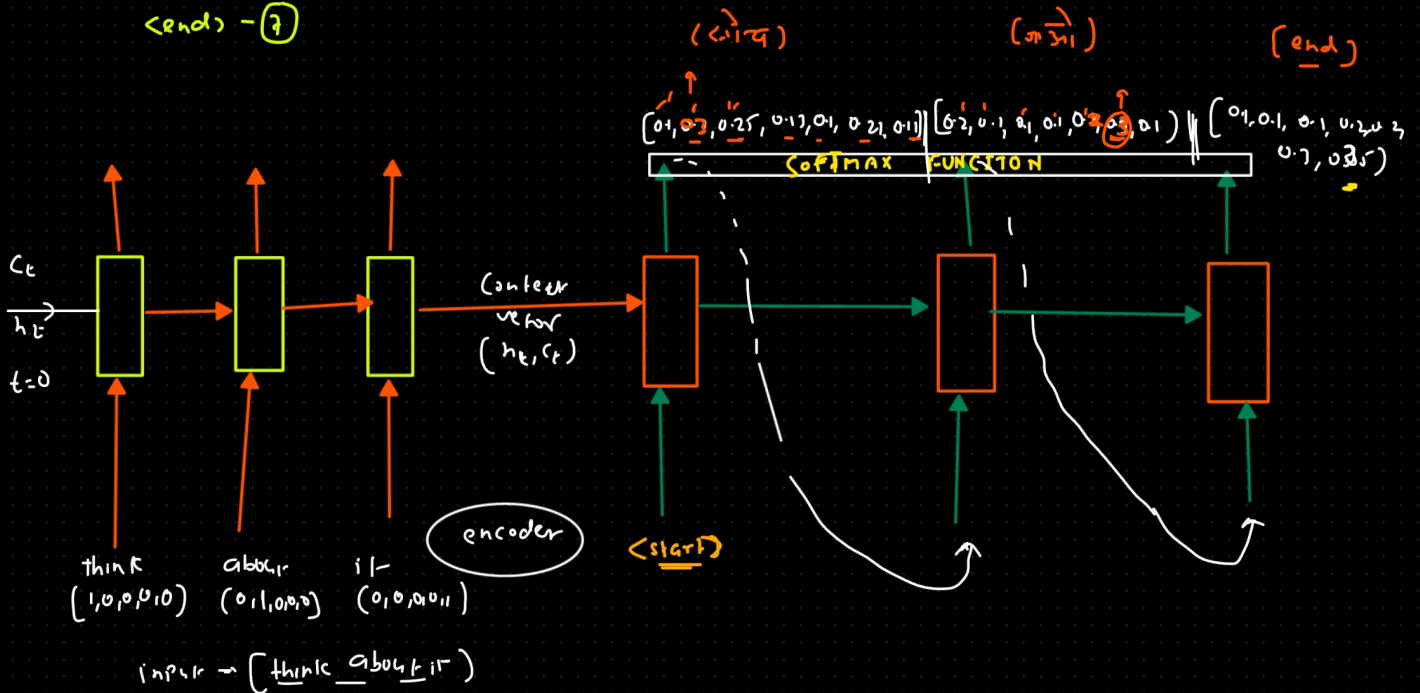
Optimize weights  $\Rightarrow$  Encoder-Decoder

## fully trained

## Prediction

$(start) \rightarrow (1)$      $z \rightarrow (2)$      $z \rightarrow (3)$   
 $z \rightarrow (4)$      $z \rightarrow (5)$      $z \rightarrow (6)$   
 $(end) \rightarrow (7)$

Weight  $L$   $\Rightarrow$  all the weight (pred)



think about it  $\Rightarrow$  dim  $u_{\mathbb{R}^n}^1$ ?