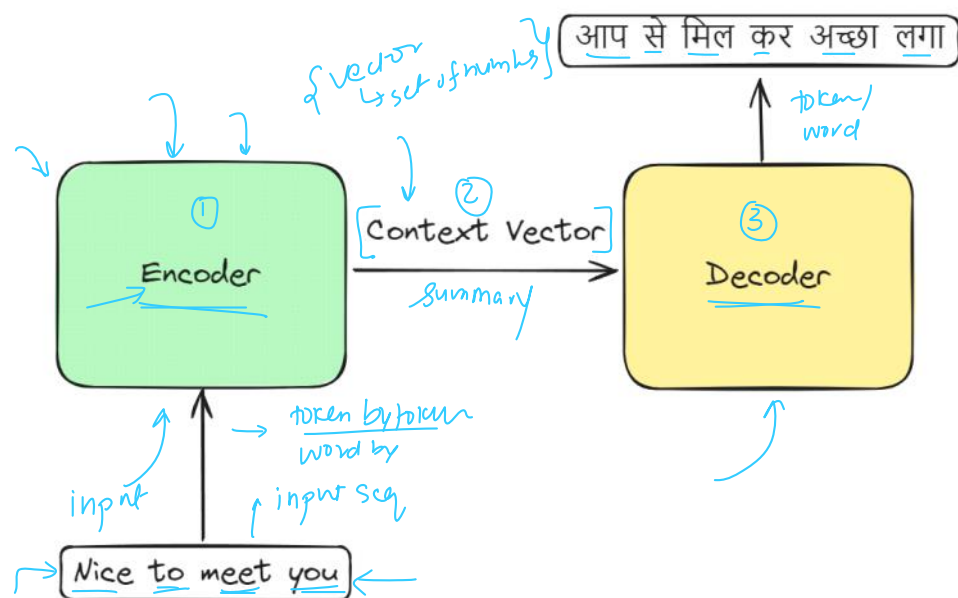


High Level Overview

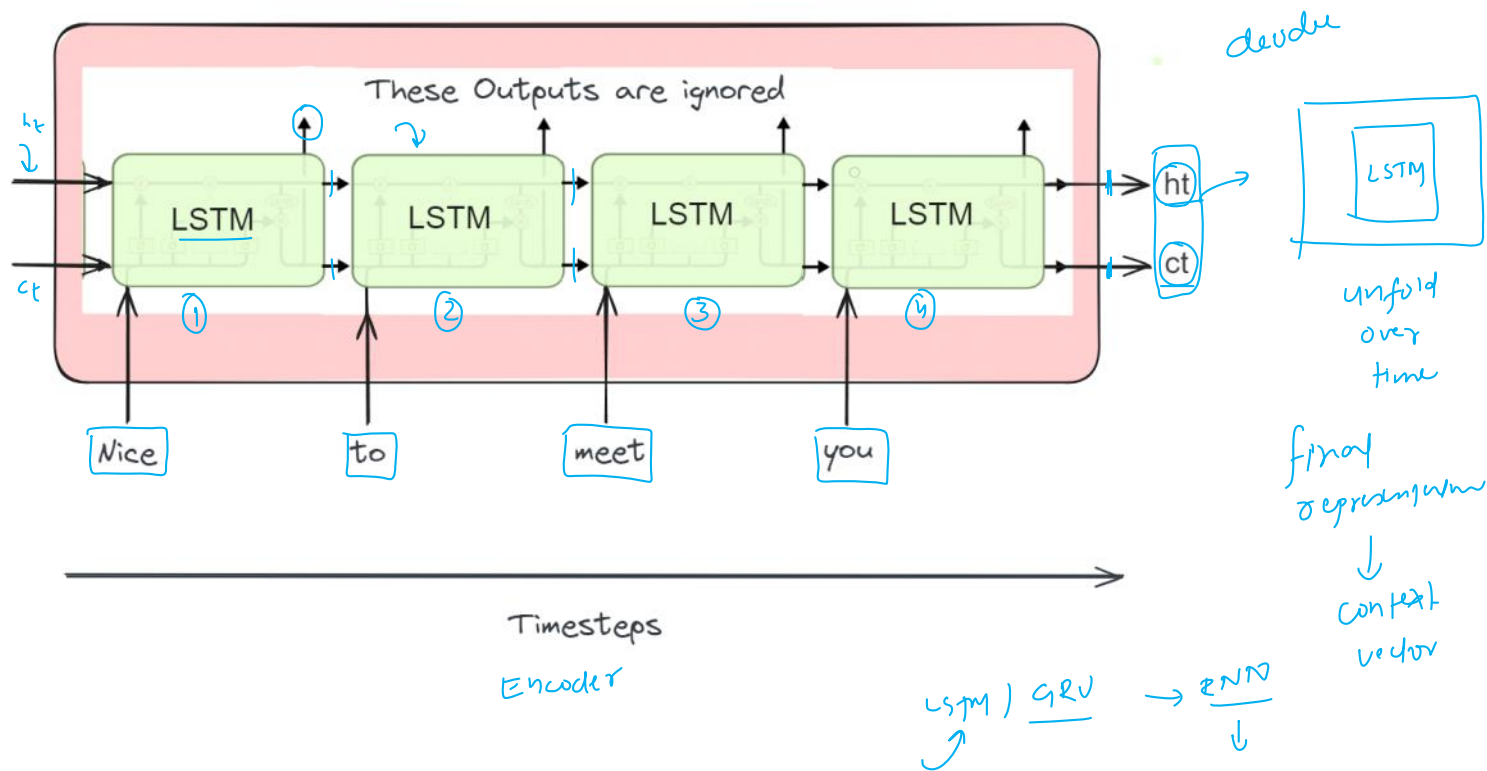
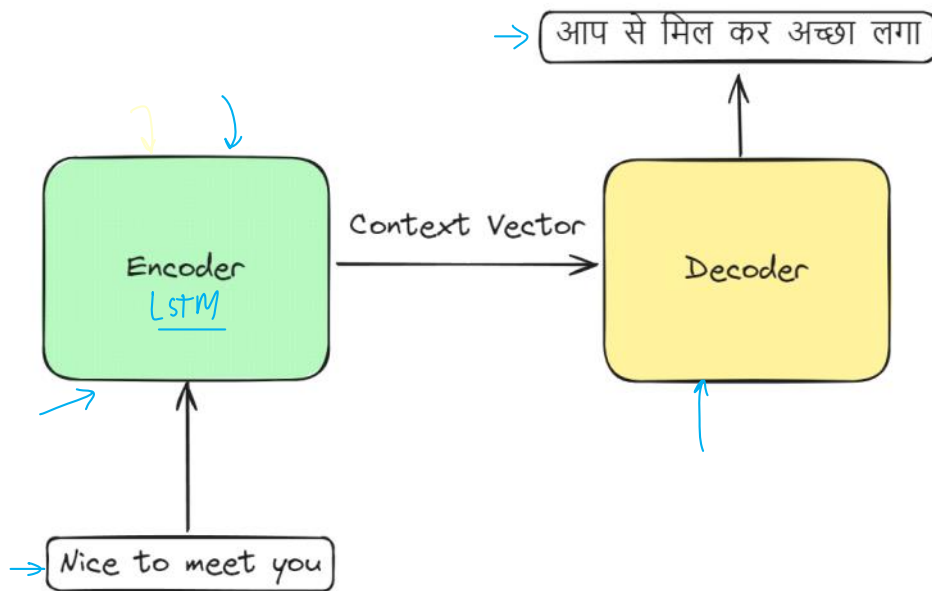
08 December 2023 17:14

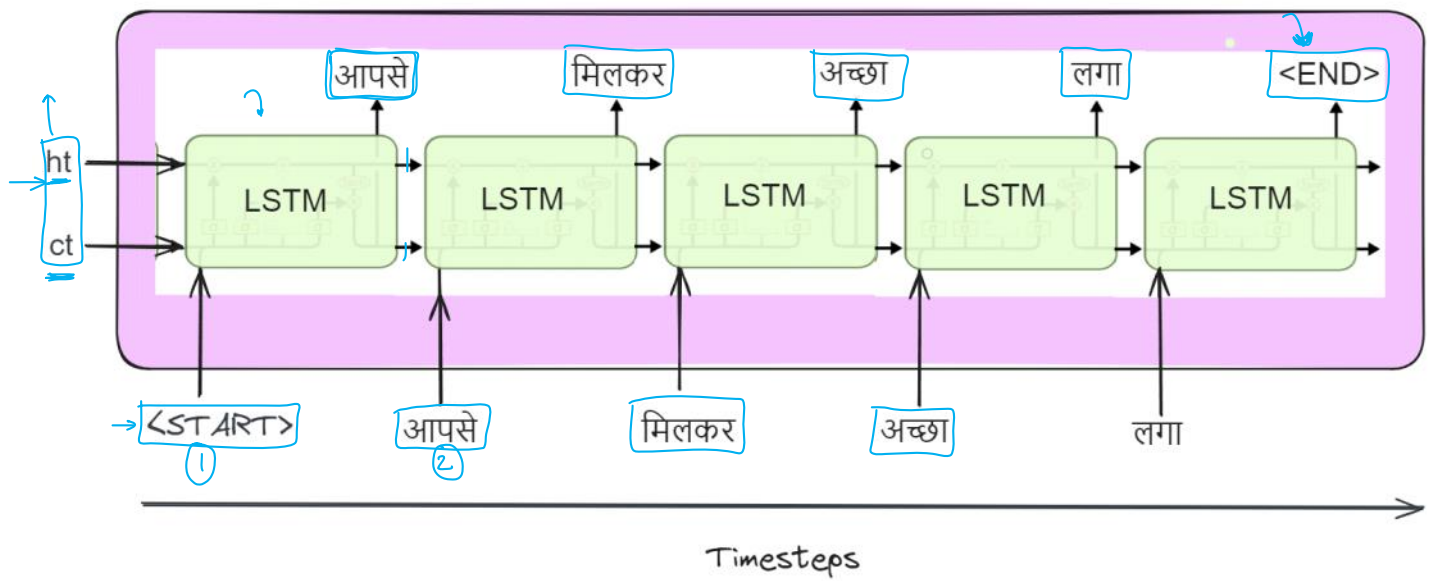


Machine tran → exam

What's under the hood?

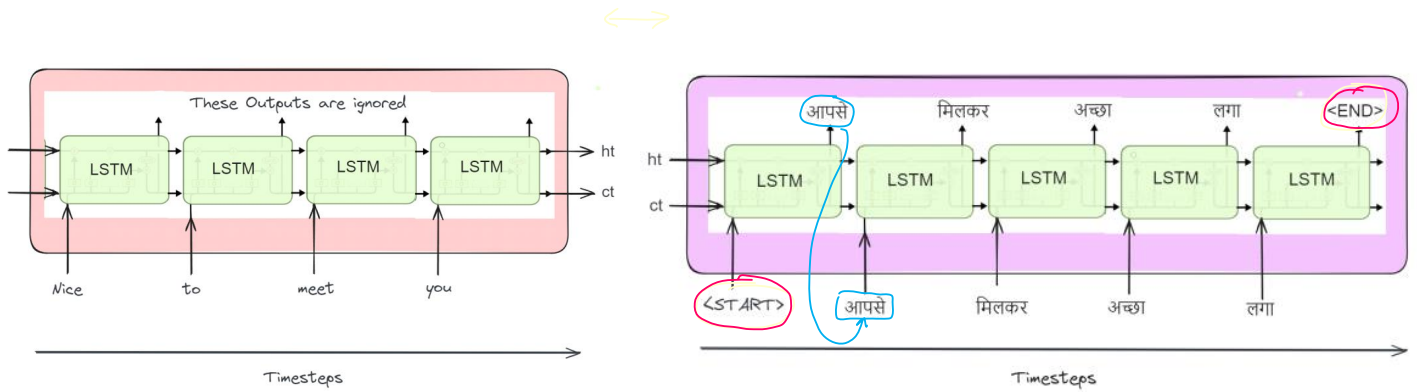
08 December 2023 17:14





Training the Architecture using Backpropagation

08 December 2023 17:15



2) Dataset \rightarrow machine Translation

eng	hindi
1) \rightarrow Think about it	सोचो सो
2) \rightarrow Come in	अंदर आ जाओ

\rightarrow numbers \rightarrow NLP

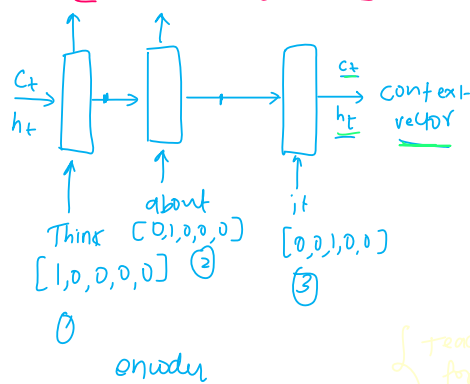
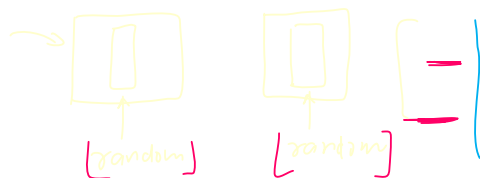
[¹think, ²about, ³it]
[⁴come, ⁵in]
 \rightarrow [सोचो, सो]
 \rightarrow [अंदर, आ [जाओ]]

one-hot encoding

eng	hindi
think \rightarrow [1, 0, 0, 0, 0]	<start>
about \rightarrow [0, 1, 0, 0, 0]	<end>
	[1, 0, 0, 0, 0, 0, 0]
	\rightarrow [0, 1, 0, 0, 0, 0, 0]
	0.2 0.15 0.5 [0.3] 0.07 0.02 0.1

अंदर \rightarrow [0, 0, 0, 1, 0, 0, 0]

row 1 \rightarrow [Think about it \rightarrow सोचो सो \rightarrow end]



loss

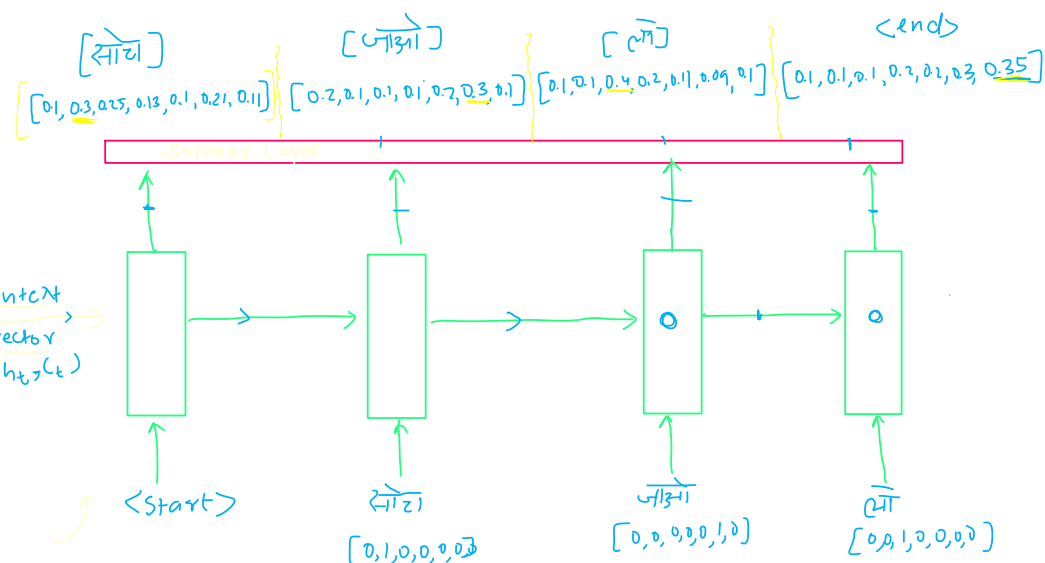
eng	hindi
सोचो	[0, 0, 1, 0, 0, 0, 0] (सो)
सो	[0, 0, 0, 0, 0, 1] <end>
	[0, 1, 0, 0, 0, 0, 0] (आ)
	[0, 0, 0, 0, 0, 0, 1] <end>

avg \rightarrow 0.7

Teacher forcing

eng	hindi
\hat{y}_{true} [0, 1, 0, 0, 0, 0]	सोचो
\hat{y}_{pred} [0, 0, 1, 0, 0, 0]	सो
	[0, 0, 0, 0, 0, 1] <end>
	[0, 1, 0, 0, 0, 0, 0] (आ)
	[0, 0, 0, 0, 0, 0, 1] <end>

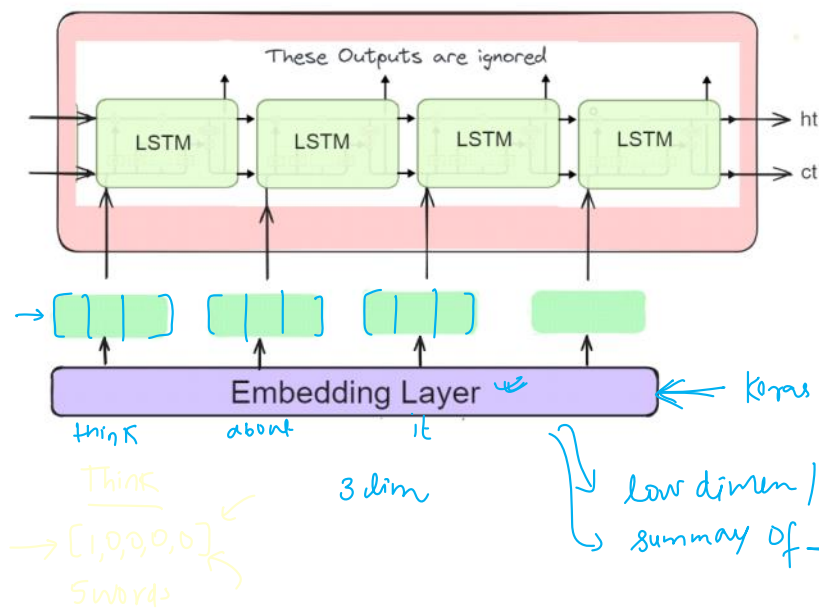
avg \rightarrow 0.7



Improvement 1 - [Embeddings]

08 December 2023 17:16

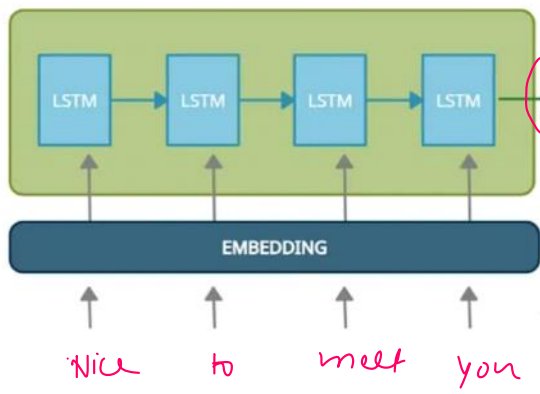
encoder



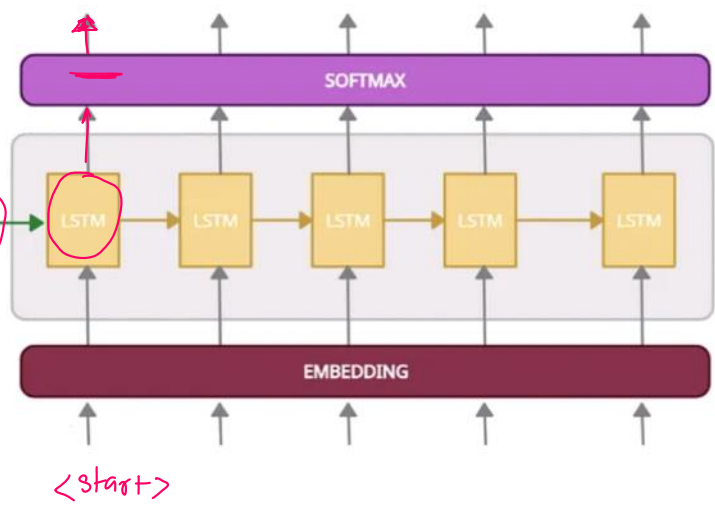
encoder
decoder

<start>

Encoder



decoder

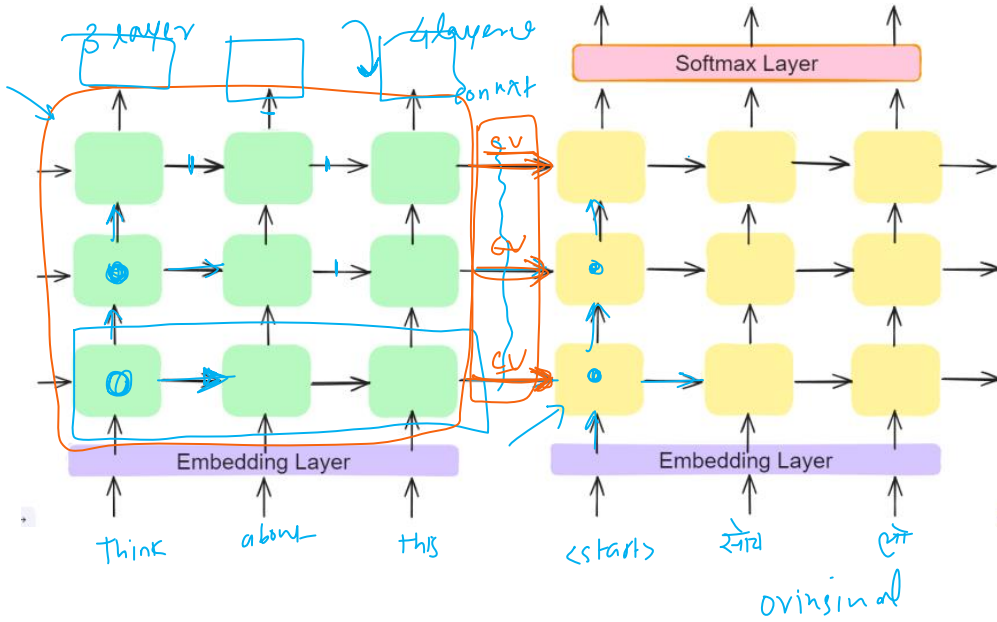


Improvement 2 - Deep LSTMs

08 December 2023 17:16

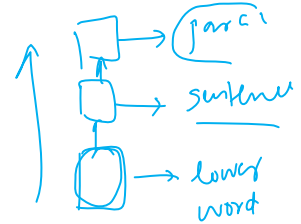
long para

single layer



3) deep learning NN learning

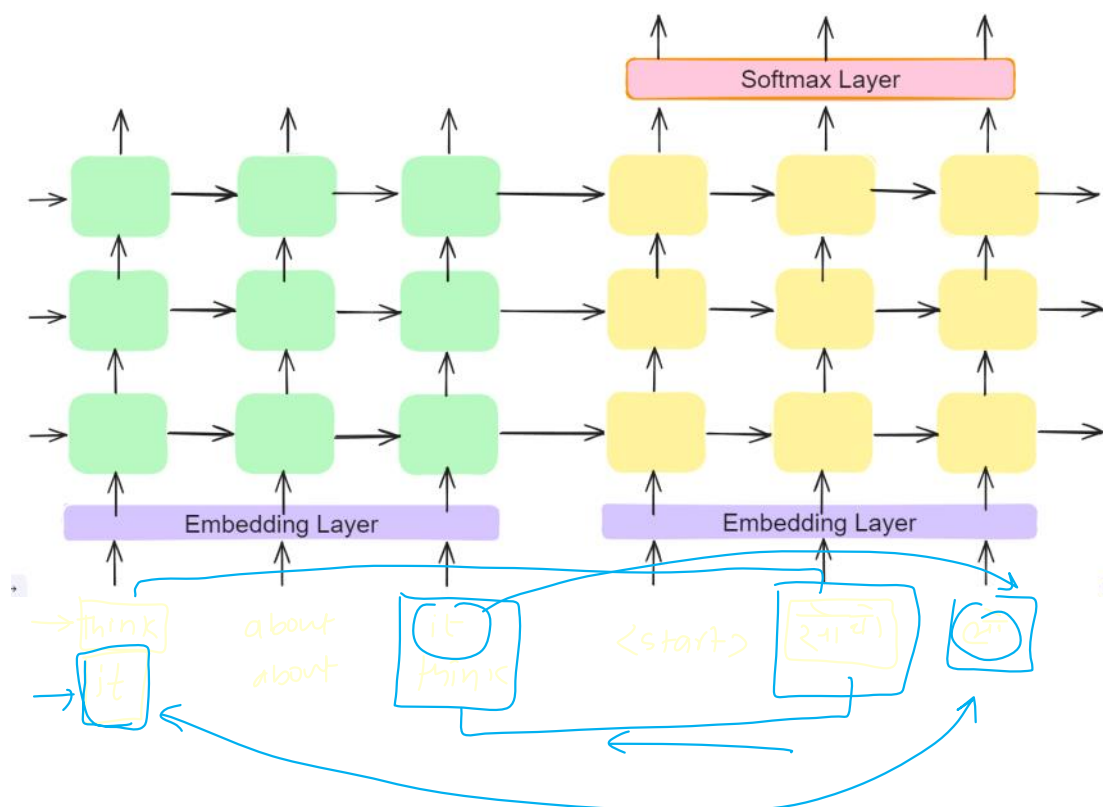
- 1) long term dependencies
- 2) layered representation



the phone battery is bad but the output is great

Improvement 3 - Reversing the Input

08 December 2023 17:16



language
↓
initial context
eng - french

The Sutskever Architecture

08 December 2023 17:18

<start> <end>

Application to Translation: The model focused on translating English to French, demonstrating the effectiveness of sequence-to-sequence learning in neural machine translation.

Special End-of-Sentence Symbol: Each sentence in the dataset was terminated with a unique end-of-sentence symbol ("<EOS>"), enabling the model to recognize the end of a sequence.

→ **Dataset:** The model was trained on a subset of 12 million sentences, comprising 348 million French words and 304 million English words, taken from a publicly available dataset.

Vocabulary Limitation: To manage computational complexity, fixed vocabularies for both languages were used, with 160,000 most frequent words for English and 80,000 for French. Words not in these vocabularies were replaced with a special "UNK" token.

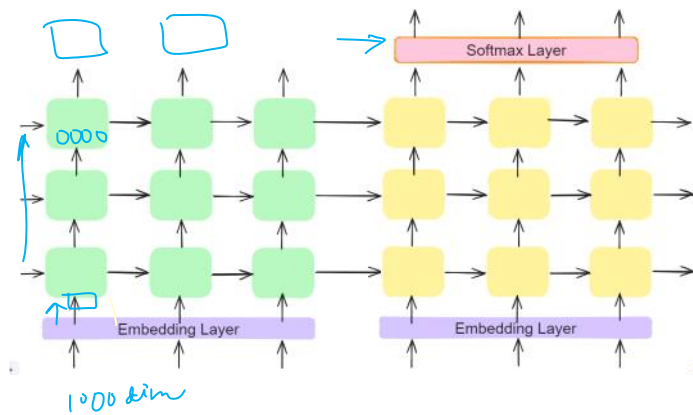
Reversing Input Sequences: The input sentences (English) were reversed before feeding them into the model, which was found to significantly improve the model's learning efficiency, especially for longer sentences.

Word Embeddings: The model used a 1000-dimensional word embedding layer to represent input words, providing dense, meaningful representations of each word.

Architecture Details: Both the input (encoder) and output (decoder) models had 4 layers, with each layer containing 1,000 units, showcasing a deep LSTM-based architecture.

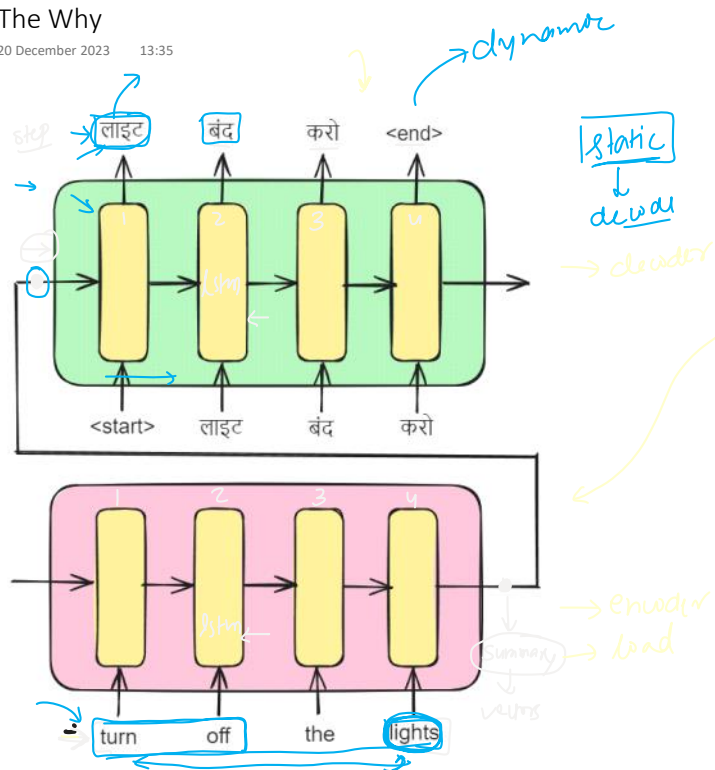
Output Layer and Training: The output layer employed a Softmax function to generate the probability distribution over the target vocabulary. The model was trained end-to-end with these settings.

Performance - BLEU Score: The model achieved a BLEU score of 34.81, surpassing the baseline Statistical Machine Translation (SMT) system's score of 33.30 on the same dataset, marking a significant advancement in neural machine translation.



The Why

20 December 2023 13:35



50 words

Encoder

Once upon a time in a small Indian village, a mischievous monkey stole a turban from a sleeping barber, wore it to a wedding, danced with the bewildered guests, accidentally got crowned the 'Banana King' by the local kids, and ended up leading a vibrant, impromptu parade of laughing villagers, cows, and street dogs, all while balancing a stack of mangoes on its head, creating a hilariously unforgettable spectacle and an amusing legend that the village still chuckles about every monsoon season.

Decoder

> 25 words

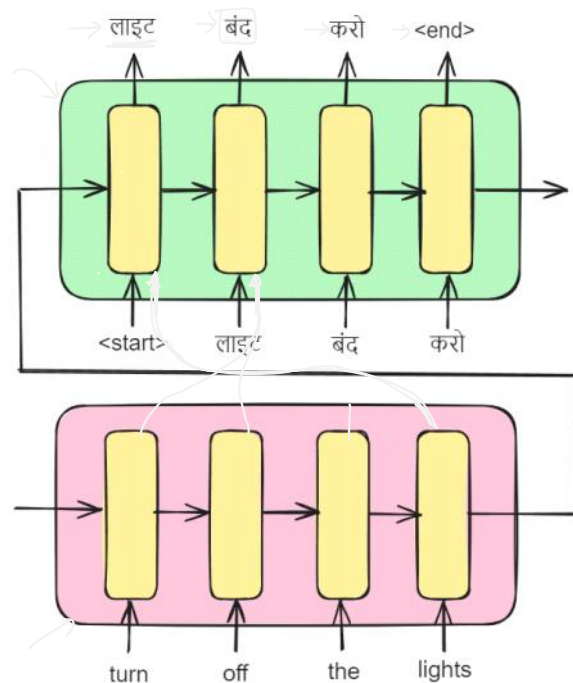
The Solution

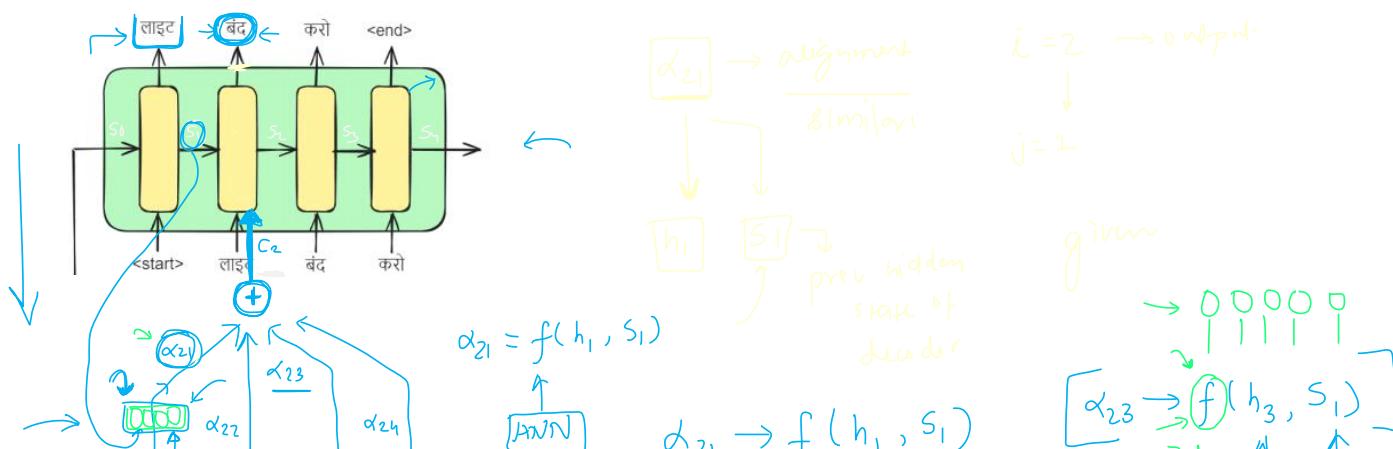
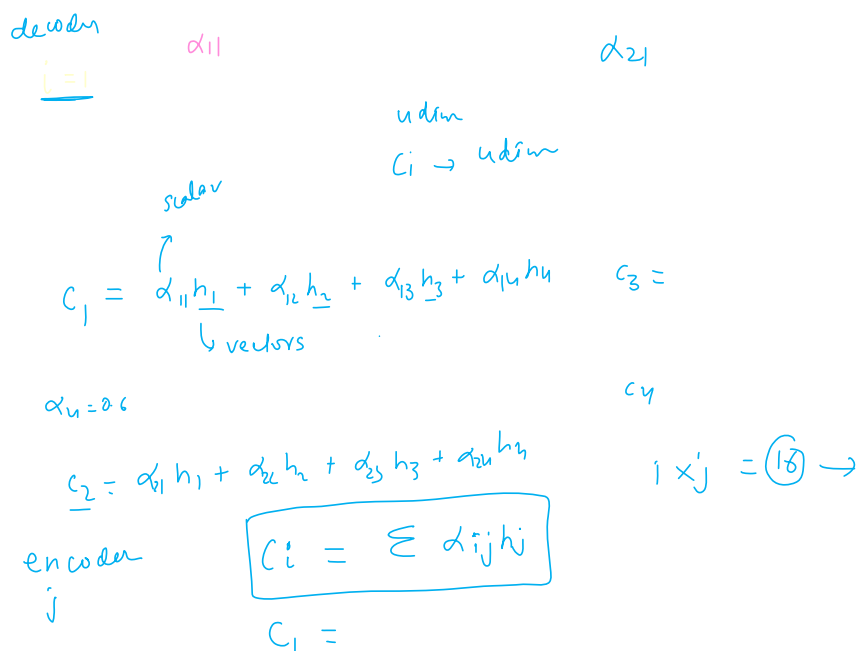
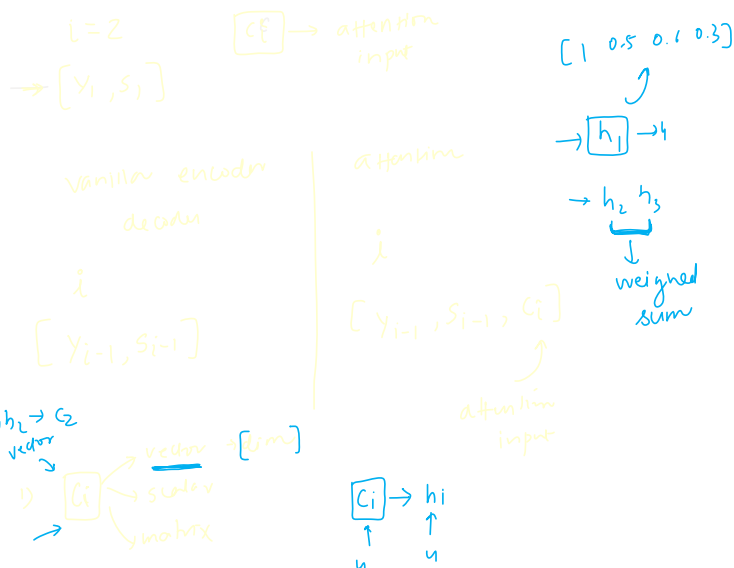
20 December 2023 17:32

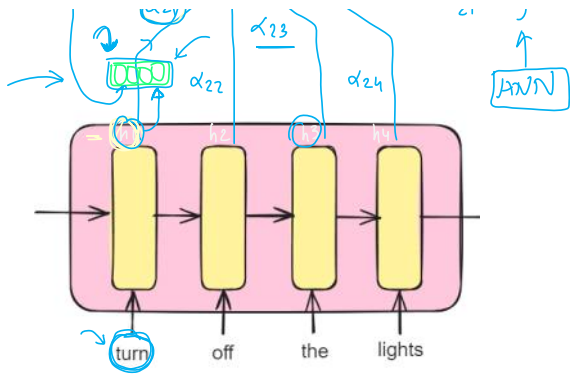


Attention

Once upon a time in a small Indian village, a mischievous monkey stole a turban from a sleeping barber, wore it to a wedding, danced with the bewildered guests, accidentally got crowned the 'Banana King' by the local kids, and ended up leading a vibrant, impromptu parade of laughing villagers, cows, and street dogs, all while balancing a stack of mangoes on its head, creating a hilariously unforgettable spectacle and an amusing legend that the village still chuckles about every monsoon season.

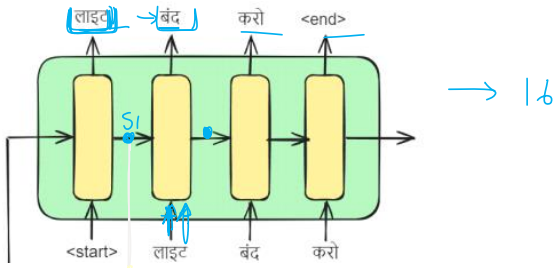
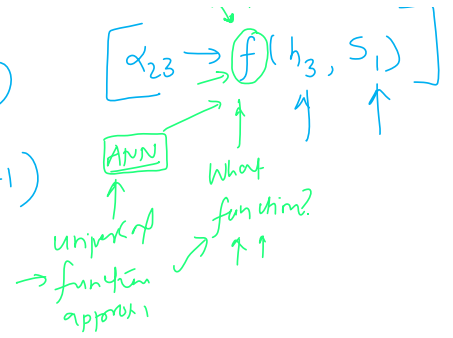






$$\alpha_{21} \rightarrow f(h_1, s_1)$$

$$\alpha_{ij} = f(h_j, s_{i-1})$$

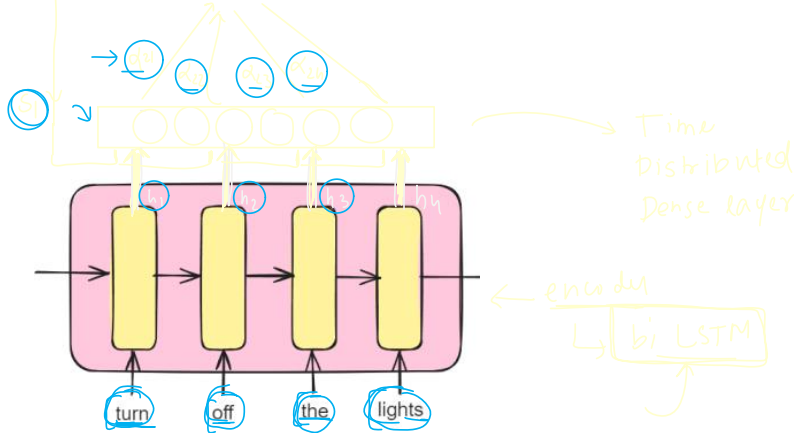


$$\alpha_{11}$$

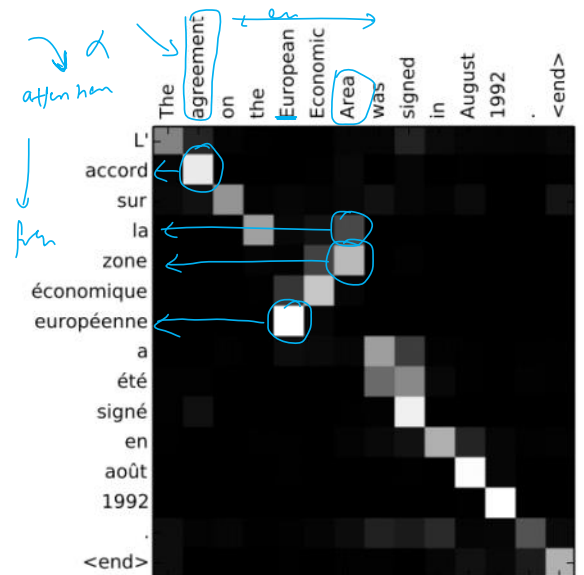
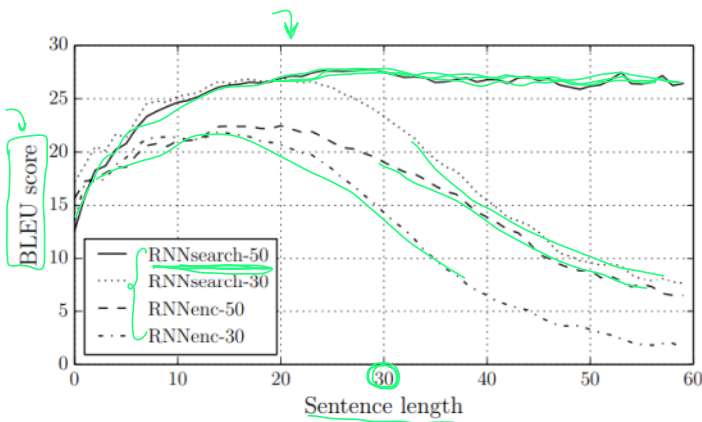
$$\alpha_{12}$$

turn off the lights
 α_{11} α_{12} α_{13} α_{14}
 end

$$c_2 = \alpha_{21}h_1 + \alpha_{22}h_2 + \alpha_{23}h_3 + \alpha_{24}h_4$$



eng-fran



(a)

sentence > 30 words
paragraph
document

bottleneck \rightarrow Attention mechanism


$$4 \times 4 = 16$$

$\alpha \rightarrow$ alignment score

$$c_2 = \alpha_{21}h_1 + \alpha_{22}h_2 + \alpha_{23}h_3 + \alpha_{24}h_4$$