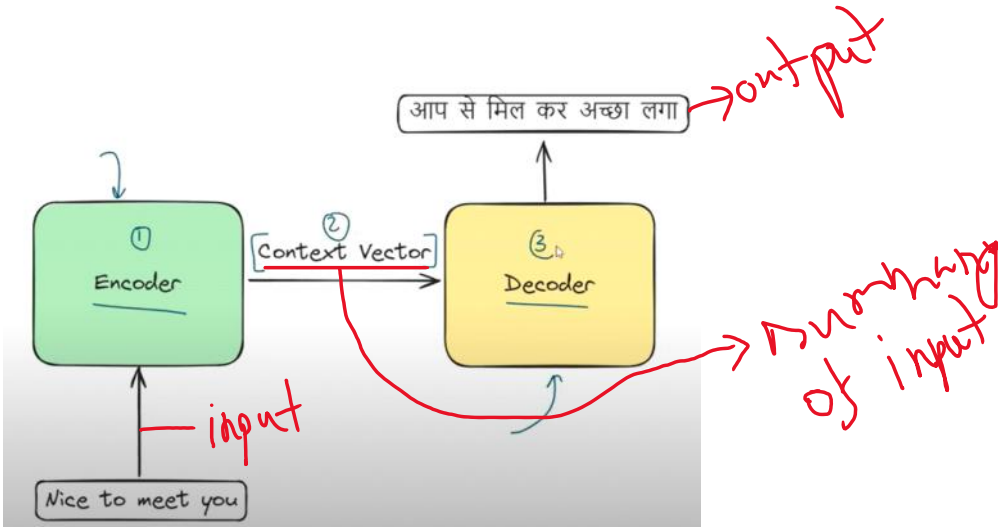
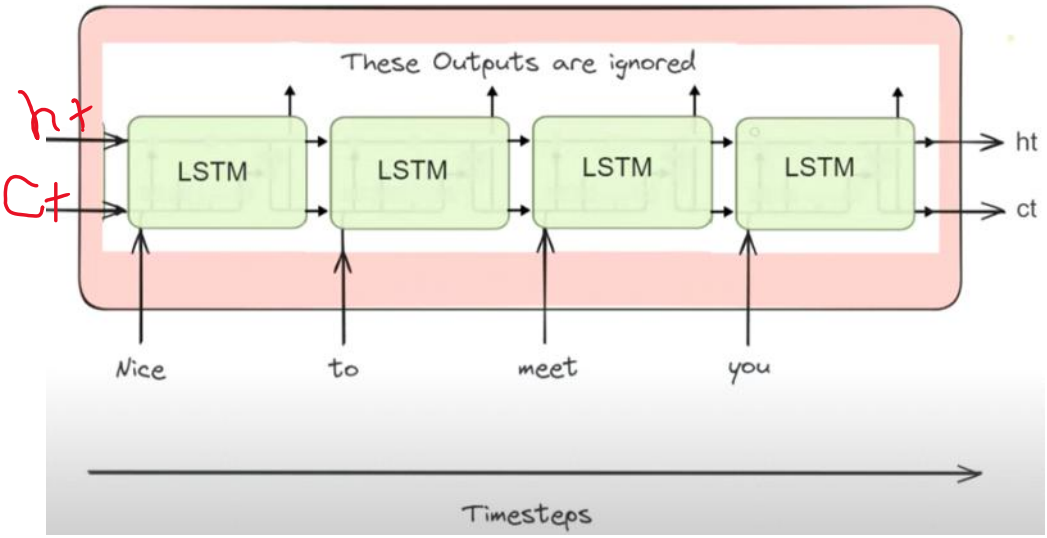
Seq to Seq

Monday, March 3, 2025 5:30 PM

► How Seq to Seq works? What are the drawbacks?

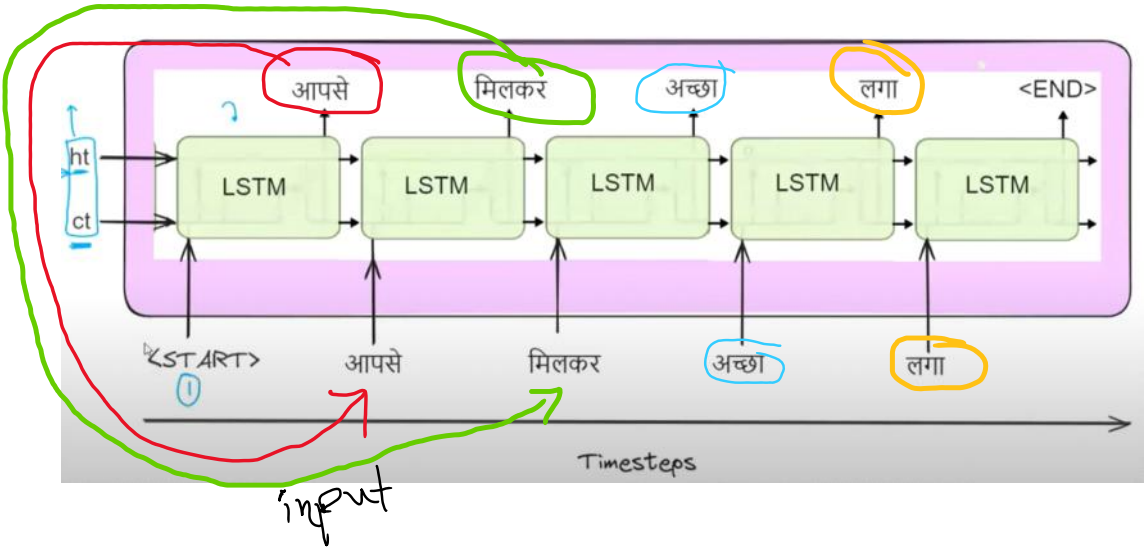


► How encoder works?



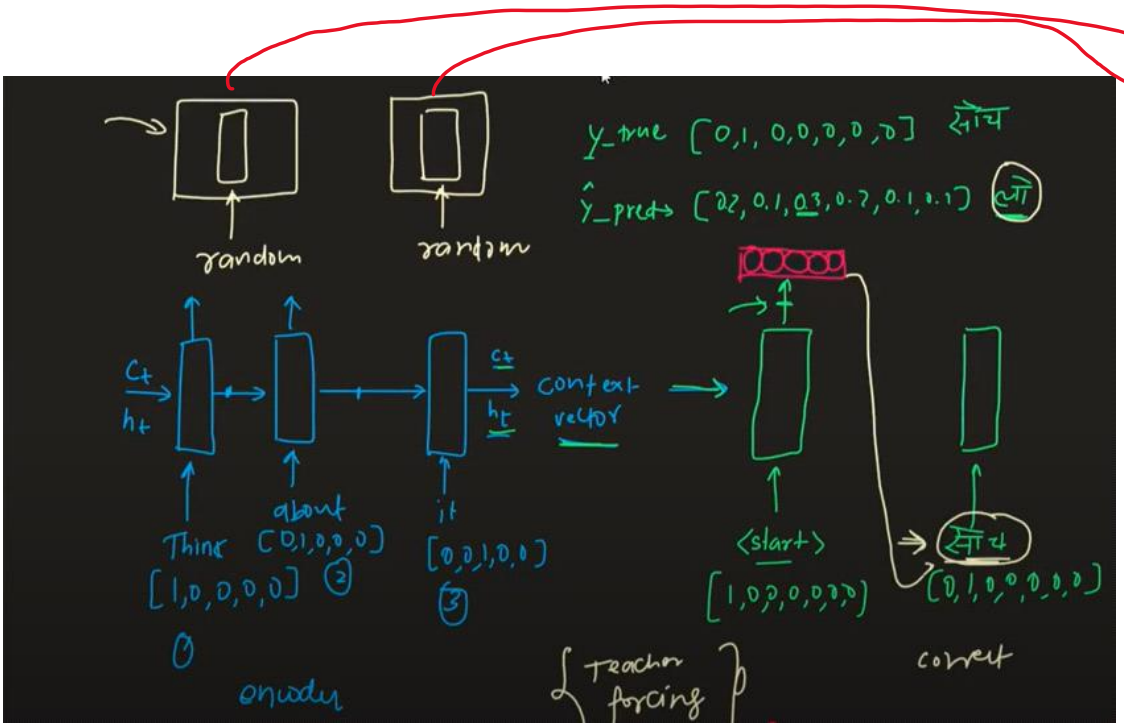
- Have one LSTM and unfold this LSTM for every input

► How decoder works?



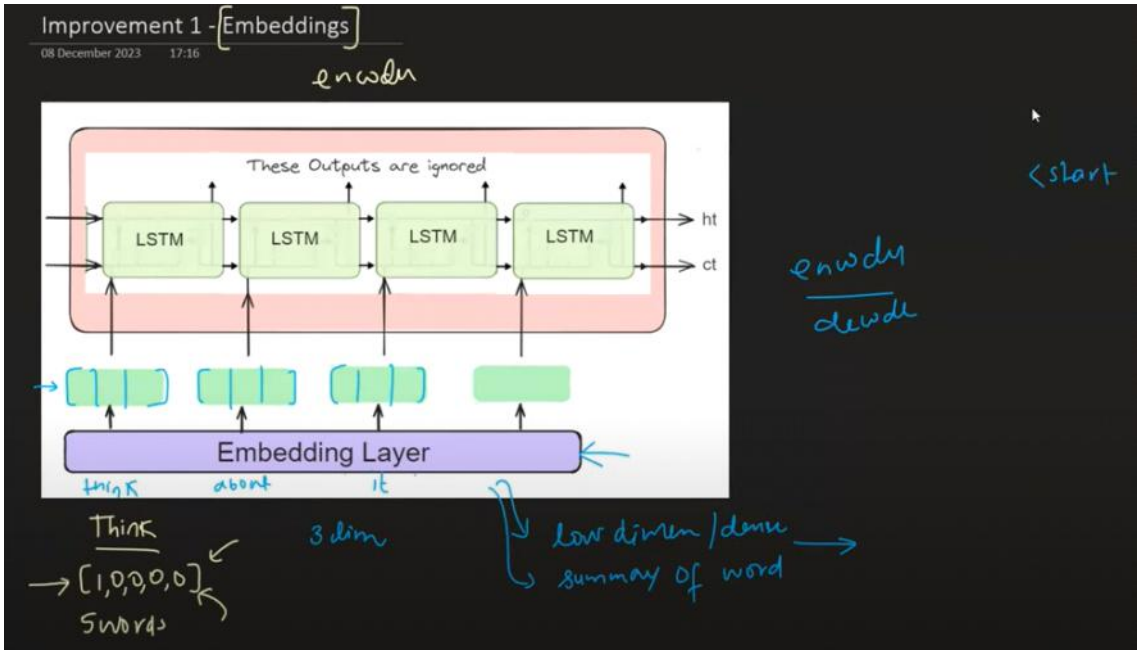
- Here LSTM generate outputs
- Every LSTM generate a single output

► How to train a model using this architecture?

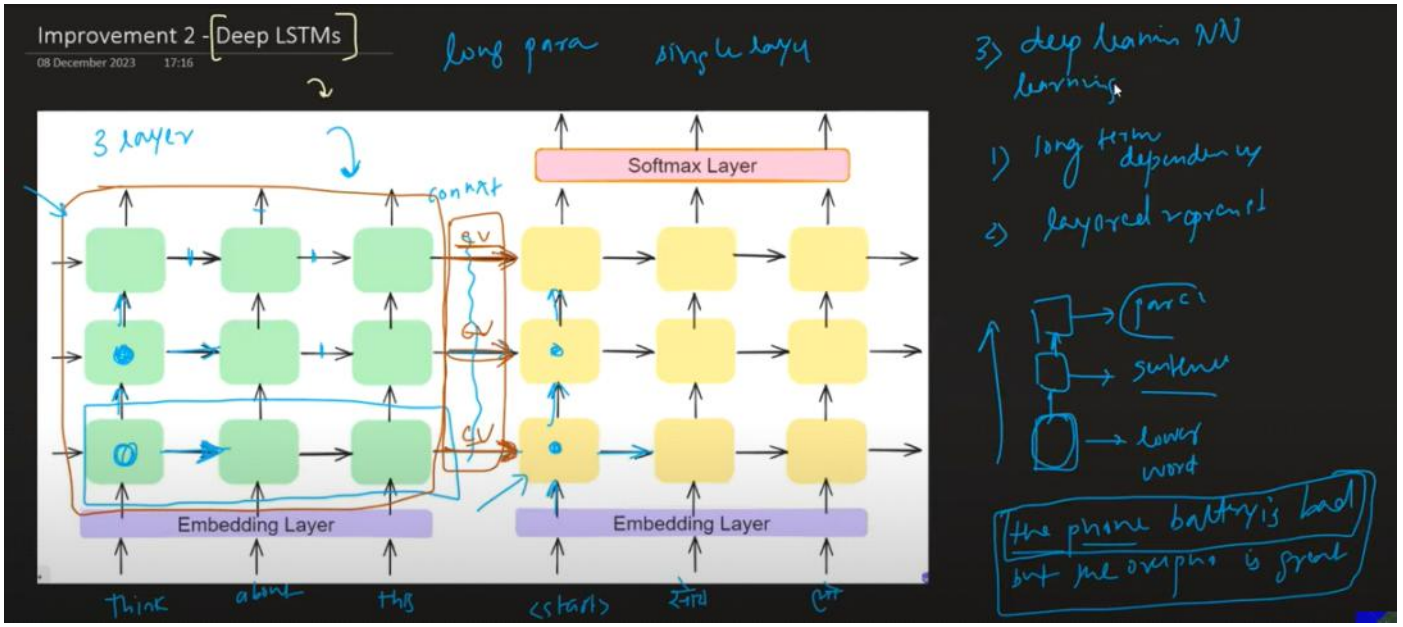


- At first the input word convert into tokens and here one hot encoding done that job
- In decoder the output from a LSTM provide as an input to next LSTM
- But what if the output in wrong from the LSTM?
- The next LSTM will still get the correct input in training time, this is called "Teacher Forcing"

► Improvements in Encoder and Decoder



- Embedding converts the input vector to a fixed length of vector which helps to take input of large vectors
- If I have a large dataset which will make my input words large cause the one hot encoding



- Increasing the number of layers can understand long sentences easily
- Number of context vector increasing with that which can increase the accuracy. First layer train the word, second layer understand sentences and third layer understand the whole paragraph.