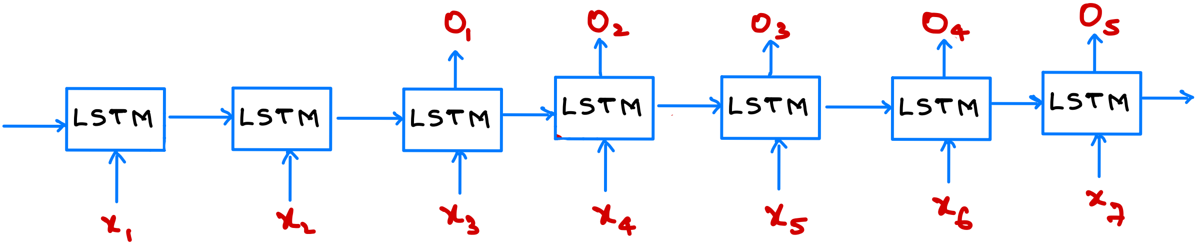
Encoders and Decoders



Why do we need Encoders and Decoders? What is the problem with LSTM (or) BiLSTM ?

In tasks like language translation, we need to understand complete sentence to translate into other language.

* LSTM only captures past data context. For example, if 3rd LSTM node should have context of 6th LSTM node. It is impossible to remember future context while using LSTM.

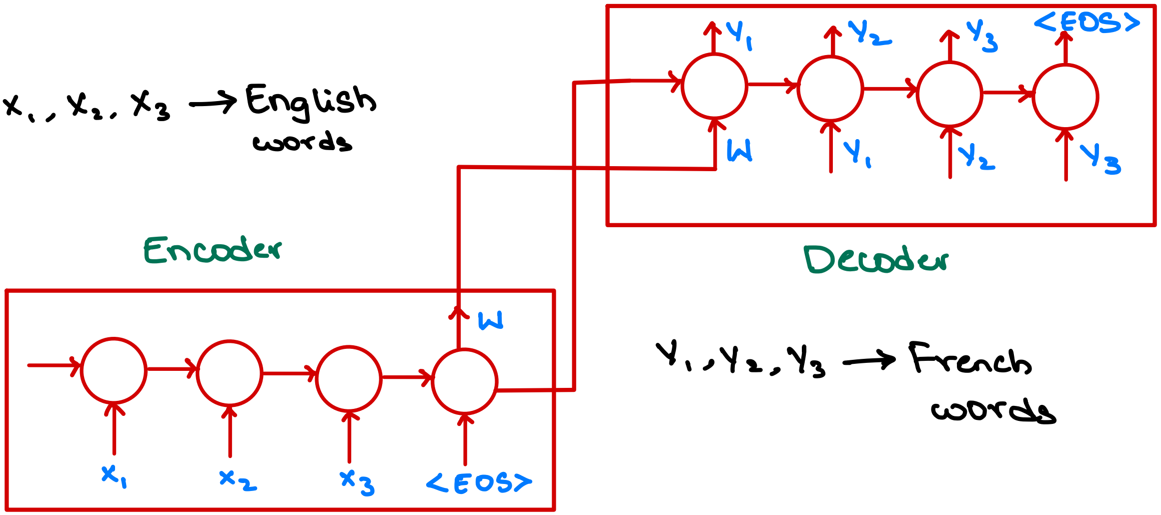
* BiLSTM resolves above issue as it captures both past and future contexts. But even BiLSTM fails if we have large sentences for translation tasks. We have 50 words in a sentence and BiLSTM can capture 10 words context. What if 10th word requires context of 45th word. It fails!

So, What's the idea?

The idea is to separate input and output layers. For this we need two sequences. Those are Encoders and Decoders.

Encoders → takes input from data source and Output input context vector.

Decoders → takes input context vector and gives output.



Note:

- * We can use any base model like LSTM, GRU, Stacked LSTM, BiLSTM etc...
- * Number of the decoder nodes need not be same as number of the encoder nodes.
- * <EOS> notifies the end of the input/output.