Lec-2

The evolution of attention in AI

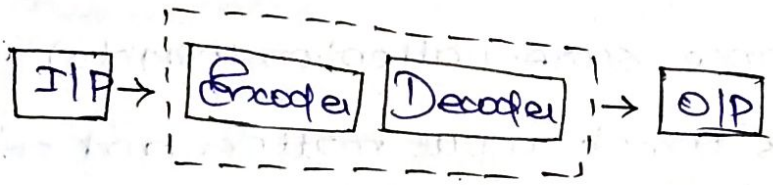
* Many real-world data types are sequences. In long-sequences, the important relationships can exist between elements far apart.

→ first and last word in a paragraph.

→ top-left and bottom right patches in an image.

* Traditional models struggled with this problem. Attention mechanism were introduced to solve long range dependency.

Seq2Seq model:



* A seq-2-seq model converts one i/p sequence to another output sequence.

* Encoder is an RNN which process the input sequentially, one at a time.

* The output of the encoder is a context vector. This single context vector represents the entire sentence meaning.

* The context vector is fed as an input to the decoder. The decoder generates one word at a time. It is also a RNN.

Problem in RNN Encoder-Decoder:

* Entire sentence meaning must fit in one vector, regardless of the sentence length.
* Decoder only sees the final context vector. It cannot see earlier hidden states / intermediate representation.
* This makes long-range dependencies extremely difficult.

Bahdanau Attention (2014):

* Instead of relying on one context vector, it allows the decoder to look at all encoder hidden assigns importance scores (attention weights)
* It decides which input matters and compute weighted contribution from all encoder states. This removes pressure from a single context vector.

* Bahdanau Attention = Attention bit two sequences.
  → Used mainly in translation.
* Self Attention = Attention to itself. (2017)
  → Every word looks at all other words.
  → Produce context aware embedding.

———x———