**Assignment\_5**

1. What are Sequence-to-sequence models?

**Ans: Sequence to Sequence (often abbreviated to seq2seq) models is a special class of Recurrent Neural Network architectures that we typically use (but not restricted) to solve complex Language problems like Machine Translation, Question Answering, creating Chatbots, Text Summarization, etc**

1. What are the Problem with Vanilla RNNs?

**Ans: RNNs suffer from the problem of vanishing gradients, which hampers learning of long data sequences. The gradients carry information used in the RNN parameter update and when the gradient becomes smaller and smaller, the parameter updates become insignificant which means no real learning is done.**

1. What is Gradient clipping?

**Ans: Gradient clipping involves forcing the gradient values (element-wise) to a specific minimum or maximum value if the gradient exceeded an expected range. Together, these methods are often simply referred to as “gradient clipping.”**

1. Explain Attention mechanism

**Ans: The attention mechanism is a part of a neural architecture that enables to dynamically highlight relevant features of the input data, which, in NLP, is typically a sequence of textual elements. It can be applied directly to the raw input or to its higher level representation.Explain Conditional random fields (CRFs)**

1. Explain self-attention

**Ans: The attention mechanism allows output to focus attention on input while producing output while the self-attention model allows inputs to interact with each other (i.e calculate attention of all other inputs wrt one input**

1. What is Bahdanau Attention?

Ans:  **An attention mechanism that learns to align and translate jointly**. It is also known as Additive attention as it performs a linear combination of encoder states and the decoder states.

1. What is a Language Model?

**Ans: A language model in NLP is a probabilistic statistical model that determines the probability of a given sequence of words occurring in a sentence based on the previous words. It helps to predict which word is more likely to appear next in the sentence.**

1. What is Multi-Head Attention?

**Ans: Multi-head Attention is a module for attention mechanisms which runs through an attention mechanism several times in parallel. The independent attention outputs are then concatenated and linearly transformed into the expected dimension**.

1. What is Bilingual Evaluation Understudy (BLEU)

**Ans: BLEU, or the Bilingual Evaluation Understudy, is a score for comparing a candidate translation of text to one or more reference translations. Although developed for translation, it can be used to evaluate text generated for a suite of natural language processing tasks**