ASSIGNMENT - 5

1. What are Sequence-to-sequence models?

Ans: Sequence-to-sequence models (Seq2Seq): These are neural network architectures designed to process one sequence of data (like text) and generate another sequence (like a translation or summary). They typically consist of two parts:

* Encoder: Processes the input sequence, capturing its meaning and context.
* Decoder: Generates the output sequence, conditioned on the information encoded by the encoder.

2. What are the problems with Vanilla RNNs?

Ans: Problems with Vanilla RNNs: While RNNs are powerful for sequential data, they have limitations:

* Vanishing gradients: When processing long sequences, gradients can become very small or large during backpropagation, hindering learning.
* Exploding gradients: The opposite of vanishing gradients, where gradients become very large, causing training to become unstable.
* Limited context: Vanilla RNNs struggle to capture long-range dependencies within sequences.

3. What is Gradient clipping?

Ans: A technique used to address exploding gradients in neural networks. It limits the magnitude of gradients during backpropagation, preventing them from becoming too large and destabilizing training. This helps the model converge more effectively.

4. Explain Attention mechanism

Ans: An enhancement for Seq2Seq models that allows the decoder to focus on specific parts of the input sequence when generating the output. This is particularly beneficial for longer sequences where capturing the entire context can be challenging. Attention helps the decoder prioritize relevant information from the encoder for each step of the output generation process.

5. Explain Conditional random fields (CRFs)

Ans: Probabilistic graphical models often used for sequence labeling tasks. They take into account the relationships between neighboring elements in a sequence. In NLP applications, CRFs are commonly used for tasks like part-of-speech tagging and named entity recognition.

6. Explain self-attention

Ans: A specific type of attention mechanism where the decoder attends to the input sequence itself. This allows the model to identify relationships between different parts of the input, which can be crucial for tasks like text summarization or machine translation.

7. What is Bahdanau Attention?

Ans: A specific attention mechanism introduced in 2014 that addresses limitations in the basic Seq2Seq architecture. It enables the decoder to focus on relevant parts of the input sequence dynamically at each step, improving the model's ability to handle long-range dependencies.

8. What is a Language Model?

Ans: A statistical method that predicts the next word in a sequence based on the previous words. LMs are trained on large amounts of text data and can be used for various NLP tasks like machine translation, text generation, and sentiment analysis.

9. What is Multi-Head Attention?

Ans: An extension of self-attention where the model learns multiple attention representations in parallel. This allows the model to attend to different aspects of the input sequence simultaneously, potentially capturing richer information.

10. What is Bilingual Evaluation Understudy (BLEU)

Ans: A metric used to evaluate the quality of machine translations. It compares a candidate translation to multiple reference translations and measures how well it matches them based on n-gram (sequence of n words) precision. However, BLEU has limitations and newer metrics are being explored that consider factors beyond n-gram overlap.