Assignment – 2: Encoder-Decoder Models using RNN and LSTM

Name: Mahak Gupta

Course: Generative AI

Assignment: Task 1 – Theoretical Understanding

Date: July 24, 2025

# Task 1 – Conceptual Questions

1. 1. What is the difference between RNN and LSTM?

RNNs are useful for sequence data but struggle with remembering long-term dependencies due to vanishing gradients. LSTMs solve this by using gates (input, forget, and output) that control the flow of information and preserve long-term memory.

1. 2. What is the vanishing gradient problem, and how does LSTM solve it?

The vanishing gradient problem occurs when gradients become extremely small during training, making it hard for the model to learn long-term dependencies. LSTMs solve this using a memory cell and gating mechanisms that allow gradients to pass through unchanged across time steps.

1. 3. Explain the purpose of the Encoder-Decoder architecture.

Encoder-decoder architecture is used to convert one sequence to another. The encoder compresses the input into a fixed-length vector, and the decoder generates the output sequence from it. It’s commonly used in tasks like machine translation.

1. 4. In a sequence-to-sequence model, what are the roles of the encoder and decoder?

The encoder processes the input sequence and outputs a context vector that summarizes the information. The decoder takes this vector and generates the output sequence one word at a time.

1. 5. How is attention different from a basic encoder-decoder model?

A basic encoder-decoder uses a single fixed context vector, which may lose important information in long sequences. Attention allows the decoder to focus on different parts of the input sequence at each decoding step, improving performance.