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

**Part-I: Theoretical Understanding of RNN, LSTM, and Encoder-Decoder**

**Task 1: Conceptual Questions :**

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

**Ans:**

| **RNN** | **LSTM** |
| --- | --- |
| Designed to process sequential data by passing hidden states from one time step to the next. | A specialized form of RNN that incorporates memory cells and gating mechanisms |
| Suffers from vanishing gradient problem | Tackle the vanishing gradient problem |
| No gates in RNN | 3 gates in LSTM:Input,Output,Forget |
| hampers learning over long sequences. | More accurate on long seq to seq |

**2)What is the vanishing gradient problem, and how does LSTM solve it?**

**Ans:** During training of deep neural networks (especially those handling long sequences), the gradient values shrink as they are backpropagated.

LSTM introduces **memory cells** that retain information over long periods.

It uses **gating mechanisms**:Input,Output,Forget Gates.

* **Forget Gate** decides what to discard.
* **Input Gate** manages what new information to store.
* **Output Gate** controls what gets passed to the next step
* These components allow LSTM to **preserve gradient flow**, thereby maintaining learning even across longer sequences.

**3)Explain the purpose of the Encoder-Decoder architecture.**

**Ans:** .It's the foundational structure that lets models learn how to **translate, summarize, or predict sequences** in a flexible way.

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

**Ans: Encoder:**

* **Function**: Processes the input sequence (e.g., a sentence) one element at a time.
* **Role**: Captures semantic and contextual information and compresses it into a context vector or internal representation.
* **Output**: This context serves as the “understanding” of the input, passed to the decoder.

### **Decoder:**

* **Function**: Generates the output sequence using the context from the encoder.
* **Role**: Starts with the context vector and produces output step-by-step (e.g., translating into another language).
* Can be enhanced with mechanisms like **teacher forcing** or **attention** to improve accuracy and fluency.

**5)How is attention different from a basic encoder-decoder model?**

**Ans**:It calculates attention weights to focus on specific parts of the input at each output time step.

* Better handling of long or complex inputs
* More context-aware output generation
* Visual interpretability via attention maps (handy for debugging!)

**Task 2: Sequence-to-Sequence Data Flow :**

Draw or describe the data flow in an encoder-decoder model using RNN/LSTM. Clearly label: ● Input sequence ● Hidden states ● Context vector ● Output sequence

Ans:

