Assignment 2:Theory Questions

# Task 1

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

RNN is good for sequences but it forgets long-term stuff. LSTM is like an upgraded version of RNN which can remember more by using gates like input, forget and output.

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

In RNNs, when training, the gradients get very small and stop learning long sequences. LSTM fixes this using a memory cell that keeps useful info and avoids this problem.

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

It is used for things like translation. Encoder reads the input sentence and turns it into a context. Decoder then uses that to make the output sentence step-by-step.

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

Encoder takes the input and gives a hidden state . Decoder uses that to generate the output sentence one word at a time.

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

In normal encoder-decoder, decoder uses just the last hidden state. Attention lets the decoder look at all encoder outputs so it can focus on important words.

# Task 8

Ques 1. What are the challenges in training sequence-to-sequence models?

Long sentences are hard to remember. Model can forget things or give wrong output if training data is small. It also takes time to train.

Ques 2. What does a “bad” translation look like? Why might it happen?

A bad translation can miss words or repeat them. It happens if the model didn’t learn properly or if the sentence is too different from training data.

Ques 3. How can the model be improved further?

We can add attention, use more data, train for more epochs, or use better techniques like GRU or bidirectional LSTM.