Seq2Seq Summarization with LSTM

Objective

The objective of this assignment is to implement and evaluate a basic sequence-to-sequence (Seq2Seq) model using LSTM layers for abstractive text summarization. The goal is to build a model capable of generating concise summaries from longer text inputs.

Tasks Completed

1. Encoder-Decoder Model Architecture

A Seq2Seq architecture was developed using LSTM layers:

- **Encoder**: LSTM layer(s) that encode input sequences (articles) into a fixed-length context vector.
- **Decoder**: LSTM layer(s) that decode the context vector into output sequences (summaries).
- **Embedding Layer**: Used for both input and output text to convert tokens into dense vector representations.
- **Dense Layer with Softmax**: Predicts the next word in the summary sequence.

2. Dataset

A small dataset consisting of approximately 100–200 news articles of ‘Times of India’ was used. Preprocessing included:

- Text cleaning (lowercasing, punctuation removal)
- Tokenization and padding
- Vocabulary creation
- Word-to-index and index-to-word mappings

Example:
Input text: "The stock market experienced a volatile session on Wednesday"
Target summary: "Market shows volatility"

3. Training and Evaluation

- **Loss function**: Sparse Categorical Cross Entropy
- **Optimizer**: Adam
- **Evaluation**: Based on validation loss and qualitative review of generated summaries

Observations:

- The model generated short, relevant summaries for small articles.
- Performance degraded on longer or complex texts due to the limited dataset.

Sample Inference

Input:
"Singh says the government is not providing enough..."

Generated Summary:
"Singh says funds are being cut by the government"