Assignment 2

Assignment Task: Text classification and language model task Using RNN ,LSTM and Attention with PyTorch

Dataset: For task text classification

https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews

Dataset: For task language model

https://www.kaggle.com/datasets/rajathmc/cornell-moviedialog-corpus

Steps to be Completed (Two Notebooks):

Notebook 1: RNN Implementation

1. **Reading the Dataset:**
    - Download the dataset.
2. **Preprocessing:**
3. **Word Embedding:**
    - Use pre-trained word embeddings (e.g., GloVe, Word2Vec) or train embeddings from scratch.
4. **Model Building (RNN):**
    - Define the RNN architecture using PyTorch, including input size, hidden layers, and output size.
    - Implement the forward pass and define the loss function.
5. **Training:**
    - Train the RNN model on the training dataset.
    - Implement techniques like early stopping and learning rate decay if necessary.
6. **Evaluation:**
    - Evaluate the RNN model on the test dataset using metrics

7. **Comparison Preparation:**
    - Save the results and the model for comparison with the LSTM model.

Notebook 2: LSTM Implementation

1. **Reading the Dataset:**
    - Reuse the preprocessed dataset from the first notebook.
2. **Word Embedding:**
    - Use the same word embeddings as used in the RNN notebook.
3. **Model Building (LSTM):**
    - Define the LSTM architecture using PyTorch, including input size, hidden layers, and output size.
    - Implement the forward pass and define the loss function.
4. **Training:**
    - Train the LSTM model on the training dataset.
    - Implement techniques like early stopping and learning rate decay if necessary.
5. **Evaluation:**
    - Evaluate the LSTM model on the test dataset using metrics.
6. **Comparison:**
    - Compare the performance of RNN and LSTM based on evaluation metrics.
    - Discuss the differences observed and possible reasons for these differences.
    - Determine which model performs better for the task of text summarization.

Notebook 3: LSTM using attention Implementation

Structure of Submission

1. **Notebooks**:
    - **Notebook 1**: RNN Implementation

- **Notebook 2**: LSTM Implementation
- **Notebook 3**: LSTM with Attention Implementation

2. **Report**:
   - A document detailing your process, results, and analysis (PDF or DOC format).
3. **All files should be in a zip file**.