Assignment 1

A Comparative Study of Part-of-Speech (POS) Tagging Models

Deadline: August 30, 2025

1. Problem Statement

The objective of this assignment is to perform Part-of-Speech (POS) tagging, which is the task of assigning a grammatical POS tag (like noun, verb, adjective, etc.) to each word in a given sentence.

- **Input:** A sequence of words (a sentence).
- **Output:** A corresponding sequence of POS tags for the input word sequence.

2. Core Tasks

Your primary goal is to implement, evaluate, and compare three distinct models given below for POS tagging. For each model, you must be able to provide a detailed explanation covering its architecture, code, and key functions.

1. **Modeling:**
    - **Hidden Markov Model (HMM)**
    - **RNN-based Encoder-Decoder Model (vanilla RNN, GRU, LSTM, or any other variant)**
    - **Large Language Model (LLM)**
2. **Evaluation:**
    - You must evaluate each model on the test set using standard metrics: **accuracy**, **precision**, **recall**, and **F1-score**.
3. **Analysis:**
    - Plot a confusion matrix for the POS tags.
    - Analyze the results from your comparison. Discuss why one model performed better than another.
    - Identify the strengths and weaknesses of each approach (e. g. tag-specific performance of different models)

3. Dataset

You can use the POS-tagged data named **Brown** with the '**universal**' tagset, which can be accessed through the 'NLTK' (https://www.nltk.org/) library in Python. You are responsible for preprocessing and splitting this data into appropriate training, validation, and testing sets to ensure a fair comparison across all models.

4. Other Instructions

1. To ensure a fair comparison, you must evaluate all models on the exact same test set.
2. If you train or fine-tune a model, do not use the test data for tuning. Use a separate validation set for this purpose. Achieving the highest possible score is not the main goal. The emphasis is on your understanding of the problem, the workings of the three different models, and the quality of your comparative analysis.
3. Though the assignment specifies an Encoder-Decoder model, you are free to ***additionally*** explore encoder-only models and discuss how they compare with the encoder-decoder models.
4. You are free to use pre-written code or code generated by an LLM. However, the core requirement is to be able to explain it thoroughly. The emphasis must be on your understanding of the implementation.

Note: Submission instructions will be conveyed later.