Sentiment Analysis via MoE and representation of IMDb Movie Reviews

The project will be sent by email (to christophe.ambroise@univ-evry.fr) as a **PDF** file with the corresponding notebook (**Python** or **Rmd**). Briefly describe the problem, write the calculations you are programming. The project can be done in pairs or alone.

Introduction

Sentiment analysis is a fundamental task in Natural Language Processing (NLP), aiming to determine the sentiment expressed in a piece of text. This project explores sentiment classification of movie reviews using the IMDb dataset and implements a **Mixture of Experts (MoE)** model to improve classification performance.

Dataset

The dataset used for this project is the **IMDb Movie Reviews** dataset, available at:

- **Original Dataset**: Stanford AI - IMDb Dataset
- **Dataset with references**: Hugging Face - IMDb Dataset
- **Reference paper**: Learning Word Vectors for Sentiment Analysis (Maas et al., 2011)

This dataset consists of 50,000 movie reviews labeled as positive or negative, split evenly for training and testing.

Project Tasks

1. Data Preparation

- Download and preprocess the dataset.
- You may select a random sample of **2,000 reviews** (1,000 positive, 1,000 negative) for efficient processing.
- Preprocessing, Tokenization, stopword removal, stemming/lemmatization.

2. Feature Extraction

- Convert text data into numerical representations:
    - TF-IDF vectorization.
    - Word embeddings (using **Word2Vec**, **GloVe**)

3. Visualisation

- Use t-SNE and UMAP to represent the data

4. Model Implementation: Mixture of Experts (MoE)

- Train the MoE model with optimized hyperparameters.
- Compare performance with baseline models:
    - Logistic Regression
    - Neural Networks (MLP, CNN, LSTM)

- Evaluate using accuracy, precision, recall, and F1-score.

5. Analysis

- Investigate expert assignments for different types of reviews.
- Visualize decision boundaries and the routing mechanism.

6. Results and Discussion

- **Performance Comparison:** MoE vs. traditional models.
- **Interpretability:** How different experts contribute to classification.
- **Potential Improvements**
- ...