Homework 2: N-gram Language Models

This project implements and evaluates N-gram language models (Unigram, Bigram, Trigram, 4-gram) using Maximum Likelihood Estimation (MLE) and smoothing techniques (Add-1, Linear Interpolation).

Files

- `main.py` : The main Python script to run the project.
- `model.py` : Contains the `NgramLanguageModel` class and all core logic.
- `utils.py` : Contains the `load_data` helper function and constants.
- `ptb.train.txt` : The training dataset.
- `ptb.valid.txt` : The validation (development) dataset, used for tuning.
- `ptb.test.txt` : The test dataset, used for final evaluation.
- `Homework_2_Report.pdf` : The final report, containing analysis and results.

Dependencies

This code requires Python 3 and the `numpy` library.

You can install `numpy` using pip:

```
pip install numpy
```