**Assignment 1 Report**

**Shravan Gadbail**

**2022102025**

**Analysis of Perplexity Values for Different Smoothing Techniques**

**1. Laplace Smoothing**

Overview: Laplace smoothing, also known as Add-1 smoothing, ensures that no N-gram has a zero probability by adding a count of 1 to every possible N-gram.

Performance:

* Strengths:
  + Works reasonably well on smaller corpora where unseen N-grams are common.
  + Ensures a simple and straightforward handling of the data sparsity issue.
* Weaknesses:
  + Perplexity values are typically higher compared to other techniques because probabilities for frequent N-grams are penalized too much to account for unseen N-grams.
* Reduces the model’s ability to prioritize common N-grams, leading to less fluent predictions.
* Expected Perplexity Trend:
  + High perplexity for larger N values (e.g., trigrams and beyond), as the smoothing excessively flattens the distribution.

**2. Good-Turing Smoothing**

Overview: Good-Turing smoothing adjusts counts by redistributing probability mass based on the frequency of frequencies. Rare N-grams are adjusted downward, and unseen N-grams are assigned a small probability.

Performance:

* Strengths:
  + Handles unseen N-grams more effectively than Laplace smoothing by redistributing probabilities in a data-driven manner.
  + Performs well in balancing probabilities of frequent and infrequent N-grams.
* Weaknesses:
  + Computationally expensive due to the need for regression and frequency-of-frequencies computation.
  + Sensitive to the quality of the corpus; may struggle with extremely sparse data.
* Expected Perplexity Trend:
  + Produces lower perplexity than Laplace smoothing, especially with higher N values, as it better handles rare and unseen N-grams.

**3. Linear Interpolation**

Overview: Linear interpolation combines unigram, bigram, and trigram probabilities using weighted averages (lambdas). It ensures that the model always has some probability mass for lower-order N-grams.

* Performance:
  + Strengths:
    - Provides the most balanced approach by considering multiple levels of N-gram contexts simultaneously.
    - Results in fluent predictions by leveraging both local context (higher-order N-grams) and global context (lower-order N-grams).
    - Tends to outperform Laplace and Good-Turing smoothing on larger corpora.
* Weaknesses:
  + Requires tuning of lambda weights for optimal performance.
  + Can be computationally intensive if lambdas are not pre-optimized.
  + Expected Perplexity Trend:
  + Produces the lowest perplexity among the three methods, especially with optimized lambdas and larger datasets.