# Safa HW1

Gil Caplan & Ido Beigelman

May 5, 2025

## 1 Q1a

rewriting the expression where we do it as bigram probabilities:

$$PP(S) = (P(w_1, \ldots, w_n))^{-\frac{1}{N}} = \left( P(w_1) \cdot \prod_{i=1}^{n-1} P(w_{i+1} \mid w_i) \right)^{-\frac{1}{N}}$$

## 2 Q1b

$$PP(S) = \left( \frac{1}{M} \right)^{N \cdot \left(-\frac{1}{N}\right)} = M$$

We can see that the perplexity in part $Q1b$ is independent of the specific words themselves and depends only on $M$ (the size of the dataset).

This happens because the training set is distributed uniformly.

## 3 Q1c

Mathematically, it is possible for a model to yield a higher test-set perplexity on $S$ compared to the uniform model described in part (b), although it is unlikely since we do the calculation with the bigrams MLE probability model which is based on the count of the words/sequence of words. It depends on the frequency of words in the training and test datasets. Due to randomness, the distribution of words might differ, which can cause the perplexity to be higher.

However, in general, it is unlikely because, for the uniform model, the perplexity is $M$, and for a trained model, the perplexity is usually expected to be less than $M$. just to show mathematically:

$$P(w_{1\ldots N}) > P(v_{1\ldots N} <=> P(w_{1\ldots N})^{\frac{-1}{N}} < P(v_{1\ldots N})^{\frac{-1}{N}} <=> PP(W) < PP(V)$$

# 4 Q2

we will work are way from PPL given and work our way to the expression we are looking for.

$$PPL(w_1...N) = \left( \prod_{i=1}^{N} P_M(w_i \mid w_1, \ldots, w_{i-1}) \right)^{-\frac{1}{N}}$$

Taking the logarithm and exponentiating i.e. e to the power of log which is an equivalant expression:

$$= e^{\log\left( \left( \prod_{i=1}^{N} P_M(w_i|w_1,\ldots,w_{i-1}) \right)^{-\frac{1}{N}} \right)}$$

Using properties of logarithms, changing the multiplication and putting the log inside i.e. the multiplication becomes a sum as per the log properly:

$$= e^{\left( \frac{1}{N} \sum_{i=1}^{N} \log\left( \frac{1}{P_M(w_i|w_1,\ldots,w_{i-1})} \right) \right)}$$

We get the expression as required, we also took out the

$$\frac{1}{N}$$

per log rules and the minus flips the fraction in the log as well