Language Modeling (NLP Assignment-2)

Dhananjay Sharma

1 Problem Statement

We are given a vocabulary of four words: `a`, `b`, `c`, and `d`, along with start (`<s>`) and end (`</s>`) tokens. The task is to create a dataset with eight sentences, calculate bigram counts, and compare unsmoothed (Model U) and smoothed (Model S) probabilities for a given sentence in the dataset using add-one smoothing.

2 Dataset D

The dataset D consists of the following eight sentences:

1. `<s> a b </s>`
2. `<s> b c </s>`
3. `<s> c a </s>`
4. `<s> d b </s>`
5. `<s> b d </s>`
6. `<s> c c </s>`
7. `<s> a d </s>`
8. `<s> b a </s>`

3 Bigram Counts

The bigram counts based on the dataset D are shown in Table 1.

4 Model U: Unsmoothing

For Model U, probabilities are calculated based on observed counts:

$$P(\texttt{<s> a}) = \frac{2}{8} = 0.25$$

Bigram	Count
<s> a	2
<s> b	3
<s> c	1
<s> d	1
a b	1
b c	1
c a	1
d b	1
b d	1
c c	1
a d	1
b a	1
b </s>	1
c </s>	1
a </s>	1
d </s>	1

Table 1: Bigram counts in dataset D

$$P(\texttt{a b}) = \frac{1}{2} = 0.5$$

Using these counts, we calculate the probability of the sentence <s> d b </s>:

$$P(\texttt{<s> d b </s>}) = P(\texttt{<s> d}) \times P(\texttt{d b}) \times P(\texttt{b </s>})$$
$$= 0.125 \times 1.0 \times 0.25 = 0.03125$$

5 Model S: Add-One Smoothing

For Model S, we apply add-one smoothing, adding one to each bigram count. The adjusted counts and probabilities are shown in Table 2.

Bigram	Adjusted Count	Probability
<s> a	3	$\frac{3}{32} \approx 0.09375$
<s> b	4	$\frac{4}{32} = 0.125$
<s> c	2	$\frac{2}{32} = 0.0625$
<s> d	2	$\frac{2}{32} = 0.0625$
a b	2	$\frac{2}{32} = 0.0625$
d b	2	$\frac{2}{32} = 0.0625$
b </s>	2	$\frac{2}{32} = 0.0625$

Table 2: Adjusted bigram counts and probabilities for Model S with add-one smoothing

Calculating the probability of the sentence `<s> d b </s>` under Model S:

$$P(\texttt{<s> d b </s>}) = P(\texttt{<s> d}) \times P(\texttt{d b}) \times P(\texttt{b </s>})$$
$$= 0.0625 \times 0.0625 \times 0.0625 = 0.000244$$

6 Summary of Results

The probability of the sentence `<s> d b </s>` is higher in Model U than in Model S, as shown below:

- Model U: $P(\texttt{<s> d b </s>}) \approx 0.03125$
- Model S: $P(\texttt{<s> d b </s>}) \approx 0.000244$

This example demonstrates that unsmoothed models can assign higher probabilities to sentences in the dataset due to the absence of probability mass redistribution to unobserved events, a characteristic introduced in smoothed models.