

CS 6320 Natural Language Processing

Homework 2 N-grams.

An automatic speech recognition system has provided two written sentences as possible interpretations to a speech input.

S1: The chairman made the decision to bring in a new financial planner .

S2: The profit of the company was going down last year said by the chief executive .

Using the bigram language model trained on Corpus, find out which of the two sentences is more probable. Compute the probability of each of the two sentences under the two following scenarios:

- i. Use the bigram model without smoothing.
- ii. Use the bigram model with add-one smoothing

Write a computer program to:

- A. For each of the two scenarios, construct the tables with the bigram counts for the two sentences above.
- B. For each of the two scenarios, construct the table with the bigram probabilities for the sentences.
- C. For each of the two scenarios, compute the total probabilities for each sentence S1 and S2.

What to turn in:

Your code and a Readme file for compiling the code. The Readme file should contain a command line that can be used to compile and execute your program directly.

```
python homework.py <corpus.txt> <sentence1> <sentence2>
```

The output of the program should contain:

8 tables: the bigram counts table and bigram probability table of the two sentences under two scenarios.

4 probabilities: the total probabilities of the two sentences under two scenarios.