HW5: probability theory

LING83800

1 Products of small probabilities

Problem Use negative logarithms to compute $\frac{1}{256} \cdot \frac{1}{8181}$, showing your work.

2 Language models

Problem Rammstein's 1997 German-language single "Du Hast" consists of the following (case-folded) lyrics, repeated twice:

```
du
du hast
du hast mich
du
du hast
du hast mich
du
du hast
du hast mich
du
du hast
du hast mich
du
du hast
du hast mich
du hast mich
du hast mich gefragt
du hast mich gefragt
du hast mich gefragt und ich hab nichts gesagt
willst du bis der tod uns scheidet
treue sein für alle tage
nein
nein
```

Compute the maximum likelihood probability of the phrase *du hast mich gefragt* 'you asked me' according to a second-order Markov model.

Hint First, write out the second-order Markovian formula for $P(\texttt{du hast mich gefragt})$. Then, use maximum likelihood to compute each probability in the formula. Do not count n-grams across a line; for example, there is no bigram `nein nein` at the end of the lyric, since those occur on separate lines.

3 Stretch goal: Bayes' rule

Problem Suppose that an employer requires their employees to take regular tests for the presence of a certain illicit substance. Let us suppose that the probability that an employee consumes said substance is %3, or equivalently, $P(\text{drug}) = .03$. There are two types of ways the test can be faulty: either the test is positive though the employee does not consume the substance (a *false positive*) or it is negative though the employee consumes the substance (a *false negative*). Let us suppose that the false positive rate is 2%, or equivalently, $P(+ \mid \neg\text{drug}) = .02$, and that the false negative rate is 10or equivalently, $P(- \mid \text{drug}) = .1$.

1. Using multiplication, addition, and the total law of probability, compute the probability that a test will come out positive.

2. Using Bayes' rule, compute $P(\neg\text{drug} \mid +)$, or the probability of obtaining a positive test when the illicit substance is absent.

Hint

1. Once you've solved part 1, you just need to plug in the probabilities to the Bayes' rule formula.

2. You may be surprised how large this probability is.