HW5solution

LING83800

1 Products of small probabilities

Problem Use negative logarithms to compute $\frac{3}{1034} \cdot \frac{1}{8181}$, showing your work.

Hint Give the final result as a real-valued probability rather than a log probability.

Solution We compute this as follows:

$$
\begin{aligned}
-\log(\frac{3}{1034}) &\approx 5.843 \\
-\log(\frac{1}{8181}) &\approx 9.010 \\
\exp(-5.545 - 9.010) &= .000000355
\end{aligned}
$$

Or, in Python:

```
left = 3 / 1034
right = 1 / 8181
result = math.exp(math.log(left) + math.log(right))
print(result)
```

This prints `3.546453936173758e-07` when executed.

2 Conditional Probability

Problem a Suppose you roll two dice (both fair, and six-sided). We saw in class that the *expected value* is of their sum is 7. But what is the probability that the two rolls sum to 8?

Solution There are five different ways that two rolls summing to 8 can occur:

$$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$$

and $|\Omega|$ is 36 so the probability is $\frac{5}{36}$.

Problem b Let's call this combined 8-roll event 'A'. What is P(A|"first die rolled is a 3")?

Solution Let B be the event that the first die shows a 3.

Then $P(A \cup B)$ is the probability that the first die shows a 3 and the sum is 8, or $\frac{1}{36}$. (since there's only one way this can happen — namely that the second die is a 5).

Since $P(B) = \frac{1}{6}$ then $P(A|B) = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$.

Which furthermore makes intuitive sense: dice rolls are independent events here, and so the problem reduces to the probability that the second roll is exactly 5.

3 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 $P($`du hast mich gefragt`$)$. Since this is a second-order model, each word—except the first two—will be conditioned on the previous two words.
- 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.

Solution The second-order Markov estimate is given by:

$$P(\texttt{du hast mich gefragt}) = P(\texttt{du})* \\ P(\texttt{hast} \mid \texttt{du}) * P(\texttt{mich} \mid \texttt{du hast})* \\ P(\texttt{gefragt} \mid \texttt{hast mich}). \quad (1)$$

Filling in maximum likelihood probabilities, we obtain:

$$P(\texttt{du hast mich gefragt}) = \frac{20}{64} \cdot \frac{14}{20} \cdot \frac{9}{14} \cdot \frac{3}{9} \approx .046875$$

4 Bayes Rule

Suppose some far-out Brooklyn factory has machines I, II, and III that produce doohickeys. Sadly, American manufacturing is past its heyday and some of the doohickeys produced are defective:

`Machines I, II and III produce, respectively, 2%, 1%, and 3% defective doohickeys.`
`Out of the total production, machines I, II, and III produce, respectively, 35%, 25% and 40% of all doohickeys.`

When I'm not at the Grad Center, I stroll over to the doohickey folks to do some quality control and select a doohickey at random from their warehouse.

Problem a What is probability that the doohickey that I selected is defective?

Hint Use the Law of total probability here.

Solution

$$P(D) = P(I) * P(D|I) + P(II) * P(D|II) + P(III) * P(D|III)$$

$$P(D) = 0.35 * 0.02 + 0.25 * 0.01 + 0.4 * 0.03$$

$$P(D) = \frac{70}{10000} + \frac{25}{10000} + \frac{120}{10000}$$

$$P(D) = \frac{215}{10000} = 0.0215$$

Problem b Given that I found a defective doohickey, what is the conditional probability that it was produced by machine III?

Hint Use Bayes rule here

Solution

$$P(III|D) = \frac{P(III)P()D|III}{P(D)} = \frac{0.4 * 0.03}{0.0215} = \frac{120}{215} \approx 0.558$$

5 Stretch goal: the fundamental theorem of HMM tagging

Problem A *hidden Markov model* (HMM) tagger is conceptually quite similar to a speech recognizer as described in lecture. Let $\mathbf{W} = w_1, \ldots, w_n$ be a sequence of words and $\mathbf{T} = t_1, \ldots, t_n$ be the corresponding sequence of (e.g., part-of-speech) tags. Our goal in tagging is to compute:

$$\hat{\mathbf{T}} = \arg\max_{\mathbf{T}} P(\mathbf{T} \mid \mathbf{W})$$

where $P(\mathbf{T} \mid \mathbf{W})$ is the probability of the tag sequence given the observed word sequence. The corresponding *generative story* is shown in Figure 1. Use Bayes' rule to rewrite the equation for $\hat{\mathbf{T}}$ in terms of $P(\mathbf{T})$ and $P(\mathbf{W} \mid \mathbf{T})$ instead, and provide a prose explanation of each of the terms in the equation.

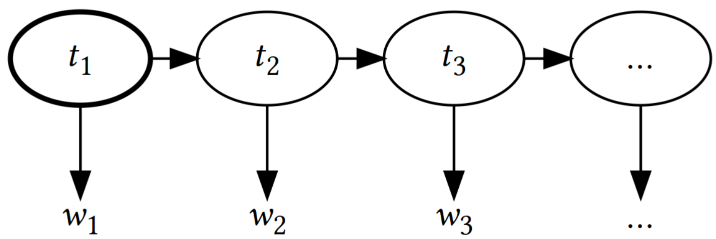

Figure 1: Illustration of a first-order hidden Markov model generative story; t: tag, w: word.

Hint As we did with $P(\mathbf{A})$ in the speech recognition case, you can omit $P(\mathbf{W})$, which is constant.

Solution Bayes' rule permits us to rewrite:

$$P(\mathbf{W} \mid \mathbf{T}) = \frac{P(\mathbf{T}) \cdot P(\mathbf{W} \mid \mathbf{T})}{P(\mathbf{W})}$$

and since $\mathbf{W}$ is constant:

$$\hat{\mathbf{T}} = \arg\max_{\mathbf{T}} P(\mathbf{T}) \cdot P(\mathbf{W} \mid \mathbf{T})$$

The $P(\mathbf{T})$ term assigns probabilities to tag sequences; in English, for example, it would assign a relatively high probability to the sequence DT NN—a determiner followed by a noun—but a smaller probability to NN DT. The term $P(\mathbf{W} \mid \mathbf{T})$ probabilistically maps tags to words; for instance, if the tag is NN, it might assign a higher probability to the word *dog* than to *anteater*.