Assignment 6

CMSC 473/673 — Introduction to Natural Language Processing

Due Monday December 9th, 11:59 PM

<table>
<thead>
<tr>
<th>Item</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned</td>
<td>Monday November 25th</td>
</tr>
<tr>
<td>Due</td>
<td>Monday December 9th</td>
</tr>
<tr>
<td>Topic</td>
<td>Hidden Markov Models and Syntax</td>
</tr>
<tr>
<td>Points</td>
<td>75</td>
</tr>
</tbody>
</table>

In this assignment you will gain some experience manipulating HMMs and reading about a powerful approach for syntactic processing.

As with past assignments, you are to complete this assignment on your own: that is, the code and writeup you submit must be entirely your own. However, you may discuss the assignment at a high level with other students or on the discussion board. Note at the top of your assignment who you discussed this with or what resources you used (beyond course staff, any course materials, or public Piazza discussions).

The following table gives the overall point breakdown for this assignment.

<table>
<thead>
<tr>
<th>Question</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points</td>
<td>20</td>
<td>25</td>
<td>30</td>
</tr>
</tbody>
</table>

**What To Turn In**  Turn in a writeup in PDF format that answer the questions; turn in all requested code necessary to replicate your results. Be sure to include specific instructions on how to build (compile) your code. Answers to the following questions should be long-form. Provide any necessary analyses and discussion of your results.

**How To Submit**  Submit the assignment on the submission site:

[https://www.csee.umbc.edu/courses/undergraduate/473/f19/submit](https://www.csee.umbc.edu/courses/undergraduate/473/f19/submit)

Be sure to select “Assignment 6.”
Full Questions

(Q1) (20 points) Consider an HMM with the following emission and transition parameters:

<table>
<thead>
<tr>
<th>observable type</th>
<th>start</th>
<th>to</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>X</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>END</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(a) Emission Probabilities

(b) Transition Probabilities

Note that the vocabulary item # is a symbol that will only ever appear at the end of a sequence. You may assume if a particular transition or emission combination doesn’t appear in the above table, its probability is 0.

For the observed sequence dab#, e.g., for the sequence of observed variables \( w_1 = d, w_2 = a, w_3 = b, w_4 = \# \), compute

(a) the marginal probability of the word sequence;

(b) the most likely state sequence \( z_1z_2z_3z_4 \) and corresponding joint probability;

You may do this by hand or by writing a program. If you hand compute these, show and turn in your work; if you write a program, turn in the code, tell us how to run it, and discuss any implementation issues you encountered (if any).

(Q2) (25 points) In this question, you will get familiar with EM for HMMs using the ice cream spreadsheet created by Jason Eisner, available at [http://www.cs.jhu.edu/~jason/465/hw-hmm/lect24-hmm.xls](http://www.cs.jhu.edu/~jason/465/hw-hmm/lect24-hmm.xls). You will answer questions related to different initial parameters (the red cells) and try to find out how the initial transition probabilities, emission probabilities and observations affect the end results. This will give you a chance to play around with a working implementation of EM, and to see what happens from iteration to iteration.

Turn in your responses for the following questions. In the following, some, but not all, probability values may be given. Rely on the fundamentals of probability to compute the unstated values. In answering the following, also

- write out the unstated probability values.
- consider how the various graphs, reconstructed weather (per day and pair of days), and final perplexity can change.

(a) What will happen if you strongly believe that one day’s weather should not be the same as the day after it? This corresponds to \( p(H|C) \) and \( p(C|H) \) being high.

(b) What will happen if you strongly believe that one day’s weather has nothing to do with the day after it, i.e., \( p(H|C) = p(C|H) = p(C|C) = p(H|H) \)?

(c) What will happen if you strongly believe that the colder the weather gets the more ice cream Jason wants to eat, e.g., \( p(1|C) = 0.1 \) and \( p(2|C) = 0.2 \) while \( p(1|H) = 0.7 \) and \( p(2|H) = 0.2 \)?

(d) What will happen if you strongly believe Jason just wants to eat a lot of ice cream and it has nothing to do with weather: \( p(1|C) = p(2|C) = p(1|H) = p(2|H) = 0 \)?
(e) What will happen if there are actually some additional, unmodeled factors that affect the number of ice cream Jason eats? For example, all of a sudden ice cream gets too expensive. To be specific, consider the case where

- For the first 11 days he eats 3 ice creams each day.
- Then for the next 11 days things become tight and he only buys 2 ice creams per day.
- And finally he’s starting to run out of money, so he can only afford 1 ice cream per day.

(f) Try out the new sequence of ice creams eaten: [2, 2, 1, 3, 1, 2, 2, 1, 3, 3, 1, 2, 1, 2, 3, 3, 3, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 1, 3, 2, 1, 2, 1] Try different initial probabilities to estimate the actual probabilities behind that observation and submit the weather you predicted. Explain why you think that is the distribution.

- How will initial reconstruction of the weather (the leftmost graph) change?
- How will the final graph after 10 iterations change?
- What is the \( p(1|H) \) after 10 iterations? Explain why this happened and what happened in each re-estimation step.

(Q3) (30 points) Read and write a half page summary and review of Charniak and Johnson (2005). It is available at http://aclweb.org/anthology/P/P05/P05-1022.pdf.

In addition to discussing the basic methodology and findings of this paper,

- Identify findings you found interesting, surprising, or confusing.
- How do the techniques used in this paper build on techniques we’ve previously covered?
- What are some task-specific difficulties the approach used in this paper runs into? Try to ground these difficulties in particular equations or sections of the paper.
- Give some examples of the features used.
- What is the overall takeaway (for you) from this paper?

The full Bibtex citation is

```bibtex
@InProceedings{charniak-johnson:2005:ACL,
  author = {Charniak, Eugene and Johnson, Mark},
  title = {Coarse-to-Fine n-Best Parsing and MaxEnt Discriminative Reranking},
  booktitle = {Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05)},
  month = {June},
  year = {2005},
  address = {Ann Arbor, Michigan},
  publisher = {Association for Computational Linguistics},
  pages = {173--180},
  url = {http://www.aclweb.org/anthology/P05-1022},
  doi = {10.3115/1219840.1219862}
}
```

1 Single spaced, regular font, and one column is fine.