

## HW02: Language Modeling and POS Tagging

Hand in at: <http://www.cs.utah.edu/~hal/handin.pl?course=cmsc723>. Remember that only PDF submissions are accepted. We encourage using L<sup>A</sup>T<sub>E</sub>X to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running “`pdflatex hw00.tex`”.

### 1 Language Modeling

1. The standard  $n$ -gram model processes a sentence “left to right.” Of course, this need not be the case. Derive, showing all your steps and saying at each point what rule of probability you’re using, or if you’re making a modeling assumption, a tri-gram right-to-left language model.
2. \*(J+M 4.1). Write out the equation for trigram probability estimation (modifying Eq. 4.14). [Eq 4.14 says “ $P(w_n | w_{n-1}) = \frac{C(w_{n-1}w_n)}{C(w_{n-1})}$ ”.]

### 2 Part of Speech Tagging

The following are some erroneously tagged sentences (these are actually from the “most frequent tagger” part of project 1!). You can see the tag set used by looking at the P1 documentation. For each sentence, mark the words that are mis-tagged, and write down what the correct tags *should* be. There is at least one error per sentence.

1. the<sub>Det</sub> year-ago<sub>Noun</sub> per-share<sub>Noun</sub> earnings<sub>Noun</sub> are<sub>Verb</sub> adjusted<sub>Noun</sub> to<sub>Aux</sub> reflect<sub>Noun</sub> a<sub>Det</sub> 2-for-1<sub>Noun</sub> stock<sub>Noun</sub> split<sub>Noun</sub> last<sub>Adj</sub> may<sub>Aux</sub>
2. the<sub>Det</sub> article<sub>Noun</sub> is<sub>Verb</sub> unfortunately<sub>Noun</sub> replete<sub>Noun</sub> with<sub>Prep</sub> outrageous<sub>Noun</sub> distortions<sub>Noun</sub>
3. the<sub>Det</sub> combined<sub>Noun</sub> buying<sub>Ger</sub> rallied<sub>Noun</sub> the<sub>Det</sub> down<sub>Noun</sub> into<sub>Prep</sub> a<sub>Det</sub> small<sub>Adj</sub> gain<sub>Verb</sub> before<sub>Prep</sub> closing<sub>Noun</sub> at<sub>Prep</sub> a<sub>Det</sub> slight<sub>Noun</sub> loss<sub>Noun</sub>
4. she<sub>Noun</sub> was<sub>Verb</sub> in<sub>Prep</sub> her<sub>Noun</sub> most<sub>Adj</sub> radiant<sub>Noun</sub> expressive<sub>Noun</sub> voice<sub>Noun</sub>

For each case, say what clued you in to the error and how you knew what the correct answer was.