Homework 1

Haukur Páll Jónsson
NLP 2017

November 12, 2017

Question 1. *Their* vs *There*

Answer a): Unigram model

The probability of a sentence, $w_1, w_2, ..., w_n$ is given by:

$$p(w_1, w_2, ..., w_n) = \prod_{i=1}^{n} p(w_i | w_1^{i-1})$$

In the unigram model we assume $p(w_i | w_1^{i-1}) = p(w_i)$, i.e. we assume that words are independent of each other. Therefore,

$$p(w_1, w_2, ..., w_n) = \prod_{i=1}^{n} p(w_i)$$

The approach to attempt to solve the *their* vs *there* in terms of a unigram model is a bad idea for the following reason. When evaluating a probability of a sentence; $p(w_1, w_2, ..., w_n)$ which contains either *their* or *there* and we want to see which sentence is more likely, we compare their probabilities and choose the one with higher probability. Namely:

$$argmax_{their,there}\{p(w_1, w_2, ..., w_{n-1}) * p(their), p(w_1, w_2, ..., w_{n-1}) * p(there)\} = argmax_{their,there}\{p(their), p(t$$

That is, we always pick the word (out of there/their) which has higher probability. That is, the more occurring word will always be considered the "correct" word.

Answer b): Bigram model

When using a bigram we assume; $p(w_i | w_1^{i-1}) = p(w_i | w_{i-1})$, that is, the probability of a word depends on the preceding word. Thus, the formula is:

$$p(w_1, w_2, ..., w_n) = \prod_{i=1}^{n} p(w_i | w_{i-1})$$

The bigram model would do a lot better as it can account for the fact that a noun is more frequently preceded by *their*, rather than *there*. Similarly, a verb is more frequently preceded by *there*, rather than *their*. The language model does not know what a "noun" or a "verb" is, but it knows the probability over all words preceding *their* and *there*, which will have this structure.

Question 2. Independence assumption

Answer a): The strange man

Answer b):

Answer c):

Question 3. Hidden Markov Model and Named Entity Recognition

Answer a): Transition matrix

The matrix is represented s.t. we go from row to column, therefore if we sum up the probabilities of a row we get 1.

Table 1: Transition matrix

	Per	Org	end
start	$\frac{3}{4}$	$\frac{1}{4}$	0
Per	$\frac{3}{6}$	$\frac{3}{6}$	0
Org	0	$\frac{11}{15}$	$\frac{4}{15}$

Answer b):

Answer c):

Answer d):