

Arvind Srinivasan
vs2371
Due: 9/20/2012

COMS W4705: Problem Set 1

Question 1

In order to minimize perplexity, we must minimize the quantity

$$\frac{1}{M} \sum_{i=1}^M \log p(s_i)$$

We know from the language model that

$$p(s_i) = \prod_{i=1}^n q(w_i | w_{i-2}, w_{i-1})$$

Substituting then, we get

$$\frac{1}{M} \sum_{i=1}^M \log \prod_{i=1}^n q(w_i | w_{i-2}, w_{i-1})$$

which can be rewritten as

$$\frac{1}{M} \sum_{i=1}^M \sum_{i=1}^n \log q(w_i | w_{i-2}, w_{i-1})$$

where M counts example sentences, and n counts the trigrams in the sentence. Since M will not count sentences not in the data, and the sentences are stop terminated, we can see that this is equivalent to

$$\frac{1}{M} \sum_{w_1, w_2, w_3} c'(w_1, w_2, w_3) \log q(w_3 | w_1, w_2) = \frac{1}{M} L(\lambda_1, \lambda_2, \lambda_3)$$

Since $\frac{1}{M}$ is constant, we can conclude that maximizing L will minimize perplexity.

Question 2

The first problem with this is that there are no restrictions on this being a well formed probability distribution, since $\lambda_1 + \dots + \lambda_n$ might not sum to 1. Consider the following example: $V = \{The, dog\}$.

$S1 = ThedogSTOP$

$S2 = ThedogtheSTOP$

$S3 = ThedogthedogthetheSTOP$

In this case, the smoothing function will weight the unigram estimates to greater than 1. This is because unigram and bigram dominance will skew the distribution such that the

trigrams have minimal weight on the estimate. This is because the way the distribution is set up, it is dependent on the trigrams. In the bigram estimate, the count of the bigrams don't vary with L (the smoothing function), which is a function of trigram counts.

Question 3 (15 points)

Input: a sentence $x_1 \dots x_n$, $q(s|u, v)$, $T(x)$

Initialization: $\pi(0, w, *, *) = 1, \pi(0, w, u, v) = 0 \forall u, v, w$ such that $u, v, w \neq *$

For $k = 1 \dots n$:

if $|T(u)| > 0, |T(v)| > 0, |T(w)| > 0$:

$$\pi(k, u, v, w) = \max_{w \in T(w)} \pi(k-1, w, u, v) \times q(v|w, v) \times e(x_k|v)$$

return $\max_{u \in T(u), v \in T(v), w \in T(w)} (\pi(n, u, v, w) \times q(STOP|u, v, w))$