

# NLP Homework 5

Yuwei Wan, Mou Zhang

November 2019

## 1 Question 4

**1** The baseline tagger has perplexity of 2802.383 and accuracy of 92.12%. When smooth = 1, The improved tagger has perplexity of 1690.606 and accuracy of 94.96%.

**2** The output is:

Model perplexity per tagged test word: 1690.606

Tagging accuracy (Viterbi decoding): 91.84% (known: 96.6% novel: 42.55%)

Tagging accuracy (posterior decoding): 94.96% (known: 96.6% novel: 77.99%)

## 2 Question 5

**3** Model perplexity per tagged test word: 4062.432

Tagging accuracy (Viterbi decoding): 88.49% (known: 96.42% seen: 49.21% novel: 49.20%)

Iteration 0 Model perplexity per untagged raw word: 3123.282

Model perplexity per tagged test word: 1950.309

Tagging accuracy (Viterbi decoding): 87.50% (known: 95.79% seen: 52.46% novel: 39.85%)

Iteration 1 Model perplexity per untagged raw word: 1294.014

Model perplexity per tagged test word: 1966.489

Tagging accuracy (Viterbi decoding): 86.90% (known: 95.20% seen: 51.89% novel: 39.18%)

Iteration 2 Model perplexity per untagged raw word: 1278.297

Model perplexity per tagged test word: 1991.537

Tagging accuracy (Viterbi decoding): 86.76% (known: 95.02% seen: 51.79% novel: 39.34%)

Iteration 3 Model perplexity per untagged raw word: 1271.318

Model perplexity per tagged test word: 2021.562

Tagging accuracy (Viterbi decoding): 86.71% (known: 94.96% seen: 51.79% novel: 39.28%)

Iteration 4 Model perplexity per untagged raw word: 1267.834

Model perplexity per tagged test word: 2052.855

Tagging accuracy (Viterbi decoding): 86.68% (known: 94.93% seen: 51.93% novel: 39.13%)

Iteration 5 Model perplexity per untagged raw word: 1266.309

Model perplexity per tagged test word: 2080.829

Tagging accuracy (Viterbi decoding): 86.65% (known: 94.91% seen: 51.74% novel: 39.23%)

Iteration 6 Model perplexity per untagged raw word: 1265.646

Model perplexity per tagged test word: 2100.906

Tagging accuracy (Viterbi decoding): 86.64% (known: 94.92% seen: 51.74% novel: 39.23%)  
Iteration 7 Model perplexity per untagged raw word: 1265.314

Model perplexity per tagged test word: 2111.099

Tagging accuracy (Viterbi decoding): 86.63% (known: 94.91% seen: 51.41% novel: 39.23%)

Iteration 8 Model perplexity per untagged raw word: 1265.125

Model perplexity per tagged test word: 2115.253

Tagging accuracy (Viterbi decoding): 86.62% (known: 94.91% seen: 51.41% novel: 39.23%)

Iteration 9 Model perplexity per untagged raw word: 1265.013

4 (a)

Because the start and end of the file must be ####/####.

(b)

Because that the test word has a certain type that can not be changed. But the tag of the raw type can be changed easily. Since the EM algorithm pursues the smallest expectation, each word will be given a tag that costs little. As a result, they are easier to predict and have lower cross entropy, which means they have lower perplexity. (c)

No. because we do not use the words in test to train.

(d)

It depends. When I use IC data to test the EM, the accuracy increases sharply to 100%. However, when I implement it on the EN data, the accuracy decreases with the iteration.

(e)

When the tagged data is not enough to obtain a good model, EM reestimation procedure can use untagged data to get additional value and help build a better model. It gets additional value from the order of observed variables.

(f)

How EM help depends on data. There are two situations that EM may not help: 1. The data is dispersed without obvious clustering. 2. The data has clustering but not in the shape of Gaussian distribution.

(g)

The maximum amount is 3. Because that day was very hot. No.