Assignment 3: Hidden Markov Models

Meg Kobashi

November 2017

1 Supervised training

The following is the results from the three experiments using different features. Using just the parts of speech had the best performance. Having more information with the words and parts of speech did not necessarily help the predictions. This might have been because the combination of the two features is too specific for the features to be useful.

Table 1: Accuracy for various features

	Precision
Words and POS	92.746
Words	91.701
POS	93.396

2 Semisupervised training

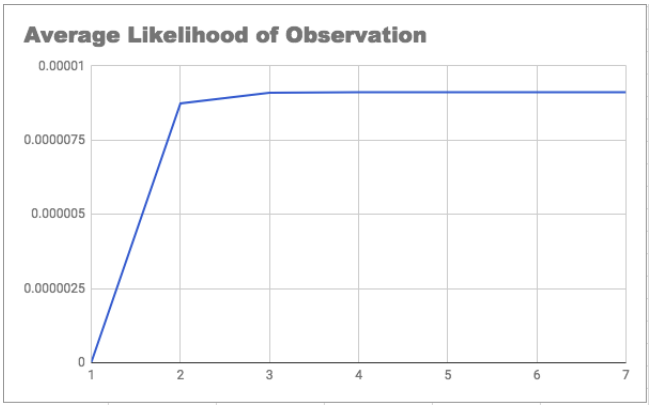

Figure 1:

The initial implementation produced NaN errors, because of floating point underflow when re-estimating the parameters in the E/M step. This was quickly resolved by using logsumexp in the intermediate calculations instead. We observe that the transition and emission probabilities slowly deviate from their initial uniform values. Furthermore, the average likelihood of the observation sequences in the unlabeled training set increases and stabilizes over a few iterations. The chart above shows this behavior.
While the training seems to be working correctly for these reasons, the classification using these optimal parameters is erroneous; the best sequence always only contains 'B'. The trellis at each time step seems to have a uniform probability across all labels so the algorithm would consecutively take the first label 'B' as the max. This seems strange since the E/M algorithm does not produce uniform probabilities for different features and labels. Had there be more time, I would once again verify the computation.