

CG168 Week 2 homework

Mark Johnson

18th September 2008

Homework:

1. *Train a Naive Bayes classifier from the training data that you prepared for the last homework. Test your classifier on how well it can predict the next move given features computed from the gold parses on both the training and the development data. With $\alpha = 1$, no unknown word pruning and the features from the last homework, my implementation of Naive Bayes obtains an accuracy of 84% on the training data. Report your accuracies on the training and on the test data.*
2. *Now use your Naive Bayes classifier as an oracle in a dependency parser, i.e., from the current state, predict the next move, and use that next move to update the state. This parser should take the word vector \mathbf{w} and part-of-speech vector \mathbf{p} and produce a governing heads vector \mathbf{h} as output. Evaluate the parses by calculating the proportion of words whose heads are correct. When evaluating on the training data, my Naive Bayes classifier obtained an accuracy of 68%. Report your accuracies on the training and on the test data.*

Hint: To aid debugging it may be worth creating a tiny corpus only containing a couple of small sentences. With a small data set like this it is possible to check the values of the various statistics by hand.