CS134 (Fall 2012): Programming Assignment 3
Maximum Entropy Model

Chuan

October 26, 2012

In this implementation, I did not do some feature selection. Just record every word in the document as feature and train the ME model. By comparison, the Naive Bayes classification has utilized feature selection by ranking the all the words by its frequencies and select part of them. The classification results of 7-fold cross-validation tests are shown below.

Maximum Entropy classification results compared with Naive Bayes classification

Figure 1: Naive Bayes classification results

```
row = predicted, column = truth
         negative positive
negative 406.0    20.0
positive 594.0    980.0

negative        precision 0.953052      recall 0.406000  F1 0.569425
positive        precision 0.622618      recall 0.980000  F1 0.761461
accuracy rate = 0.693000
```

Figure 2: Maximum Entropy classification results

```
row = predicted, column = truth
    neg   pos
neg 857.0 280.0
pos 143.0 720.0

neg     precision 0.753738      recall 0.857000  F1 0.802059
pos     precision 0.834299      recall 0.720000  F1 0.772947
accuracy rate = 0.788500
```

we can see that the result of ME classification is apparently better than the Naive Bayes classification result both on F1 and overall accuracy. And Naive Bayes either suffers from high precision, low recall or high recall, low precision problem. Also, we have applied no feature selection and doesn't need to do smoothing by using the ME model. Although under some tasks the Naive Bayes model can get quite well results, it cannot model the dependency between features (because its strong independence assumption). So the Maximum Entropy model has apparent advantages about such classification task.