John Nickle

Naïve Bayes Classifier

Overall the Naïve Bayes did a decent job dealing with sentiment labelled sentences. With a success rate of around 75% total for all three text sources, I can definitely see the merits of using a quick classifier like this to filter out unwanted results. Results did show, however, that it seems to only be helpful when dealing with clear delimiters between categories. With the Farm Ads data, it performed terribly at about 50% overall classification success; so a 50/50 chance it could classify between two categories. This was unexpected as I was sure the classifier would detect some differences, as there is certainly a reason why some farm ads are positive and some are negative.

The results of the Naïve Bayes classifier for the sentiment labelled sentences are as follows:

Amazon.txt:
Naive Bayes: 396/500 (79.20%)
POSITIVE: 201/250 (80.40%)
NEGATIVE: 195/250 (78.00%)

Imdb.txt:
Naive Bayes: 382/500 (76.40%)
POSITIVE: 181/250 (72.40%)
NEGATIVE: 201/250 (80.40%)

Yelp.txt:
Naive Bayes: 375/500 (75.00%)
POSITIVE: 185/250 (74.00%)
NEGATIVE: 190/250 (76.00%)

As you can see the results are pretty consistent across three different text domains. For the purpose of positive vs negative reviews of products and services, the Naïve Bayes classifier seems to do pretty well.