Jeremiah O’Connor

Artificial Intelligence: Spam Filter README

For the assignment, following are the decisions that I’ve made in attempt to make the spam filtering better.

Firstly, from experimenting I found out that considering the headers brings the accuracy little better so decided to include and parse the header as well as the body and any html contents for both creating training sets and parsing unseen email.

When the training set is created, I simply use dictionary to keep track terms that are collected from that corpus of emails. The terms that are in the stop words list is exempted so that more relevant terms are taken to account.

When unseen email comes in, parsing the terms from that email is done the same ways as previously for the training sets. When calculating the probability of this unseen email being either ham or spam, I consider three cases on each term. If a given term from that email is in both ham and spam training sets, then the regular probability is calculated using both of the training set, P(t1 | spam), P(t1 | ham). However, if the word is in only one of the training set, for instance in the ham, I would give P(t1 | ham) = .99 and P(t1 | spam) = .01. If the word is not in both training sets, then I only give P(t1 | ham) = .4 because spam words tends to be all too fimiliar so it most likely to be a ham. (A Plan for Spam by Paul Graham)

On top of that, in order to deal with the underflow problem I’m converting the numbers into log domain every time we calculate the MAP and its priors.

Even though this approach works pretty well with spam, spam\_2, easy\_ham, and easy\_ham\_2 as in both training and testing set, it doesn’t work that well when hard\_ham comes into play since it looks very similar to spam email. Following is the results on every possible scenario that we can think of from the five dataset: spam, spam\_2, easy\_ham, easy\_ham\_2, and hard\_ham.

-------------------------------------------------------- test: spam\_2, Tsets: spam, easy\_ham

Result: spam-> 1333, ham-> 65 -------------------------------------------------------- test: easy\_ham\_2, Tsets: spam, easy\_ham

Result: spam-> 31, ham-> 1370 -------------------------------------------------------- test: hard\_ham, Tsets: spam, easy\_ham

Result: spam-> 191, ham-> 60 -------------------------------------------------------- test: spam\_2, Tsets: spam, easy\_ham\_2

Result: spam-> 1247, ham-> 151 -------------------------------------------------------- test: easy\_ham, Tsets: spam, easy\_ham\_2

Result: spam-> 191, ham-> 2360 -------------------------------------------------------- test: hard\_ham, Tsets: spam, easy\_ham\_2

Result: spam-> 196, ham-> 55 -------------------------------------------------------- test: spam\_2, Tsets: spam, hard\_ham

Result: spam-> 664, ham-> 734 -------------------------------------------------------- test: easy\_ham, Tsets: spam, hard\_ham

Result: spam-> 397, ham-> 2154 -------------------------------------------------------- test: easy\_ham\_2, Tsets: spam, hard\_ham

Result: spam-> 341, ham-> 1060 -------------------------------------------------------- test: spam, Tsets: spam\_2, easy\_ham

Result: spam-> 439, ham-> 62 -------------------------------------------------------- test: easy\_ham\_2, Tsets: spam\_2, easy\_ham

Result: spam-> 40, ham-> 1361 -------------------------------------------------------- test: hard\_ham, Tsets: spam\_2, easy\_ham

Result: spam-> 209, ham-> 42 -------------------------------------------------------- -------------------------------------------------------- test: spam, Tsets: spam\_2, easy\_ham\_2

Result: spam-> 474, ham-> 27 -------------------------------------------------------- test: easy\_ham, Tsets: spam\_2, easy\_ham\_2

Result: spam-> 62, ham-> 2489 -------------------------------------------------------- test: hard\_ham, Tsets: spam\_2, easy\_ham\_2

Result: spam-> 217, ham-> 34 -------------------------------------------------------- test: spam, Tsets: spam\_2, hard\_ham

Result: spam-> 332, ham-> 169 -------------------------------------------------------- test: easy\_ham, Tsets: spam\_2, hard\_ham

Result: spam-> 315, ham-> 2236 -------------------------------------------------------- test: easy\_ham\_2, Tsets: spam\_2, hard\_ham

Result: spam-> 552, ham-> 849 --------------------------------------------------------