CS 180: Introduction to Artificial Intelligence

ME2 - Naive Bayes Classifier

Introduction

This machine exercise is an application the Naive Bayes Classifier for spam filtering.

Dataset

The dataset you will be using is in the directory *dataset*. You can find it in our Dropbox link for this class. The *dataset* folder contains two subdirectories - training and test. The *training* set contains individual emails, ham and spam separated, which you need to read to obtain the Naive Bayes model. Upon construction of the model, test it with the emails in the *test* folder.

The individual emails had been stripped of common English words, like *the*, *a*, and *of*, that are not meaningful in deciding ham/spam status. Moreover, non-words like numbers and punctuations were also removed. Lastly, words that mean the same but formed differently had been replaced by a representative word. For example, *include*, *includes*, and *including*, were represented by *include*.

Steps

Given the dataset, the following steps are to be taken to construct the Naive Bayes Classifier.

- Learn the initial distribution of ham and spam messages (*p(ham)* and *p(spam)*).
- Extract from the training set the dictionary of words.
- For each word in the dictionary, learn its distribution on both the ham and spam subsets ($p(w_i|ham)$, $p(w_i|spam)$)
- For testing, represent each test instance by a vector representing the absence(0) or presence(1) of each dictionary word in the email of consideration. For example, given dictionary D = {*king*, *world*, *love*, *CS180*}, and the test instance *I am the king of the world*, the vector W should be <1,1,0,0>.
- Having constructed the test vector, compute for $p(ham|W) = \frac{p(ham)\prod p(w_i|ham)}{p(ham)\prod p(w_i|ham)+p(spam)\prod p(w_i|spam)}$

Output

Your program must output the correct classification for the emails in the *test* subdirectory.

```
1.txt - ham
2.txt - spam
.
.
.
n.txt - ham
Total number of ham messages: 100
Toal number of spam messages: 10
```

Submission

The deliverables for this ME are the following: the source code and a short documentation explaining your work. Email the deliverables to *kedelaspenas@up.edu.ph* with the subject *CS 180: ME2 <surname>*. Deadline of submission is on 15 February 2014 11:59PM.

Notes

- Work individually.
- Rules against cheating and plagiarism will be strictly observed.

References

[1] *Exercise 6: Naive Bayes*. Open Classroom. Stanford University. Available Online: *http://openclassroom.stanford.edu/MainFolder/DocumentPage.php?course=MachineLearning&doc=exercises/ex6/ex6.html*