Classifying with probability theory: Naive Bayes Classifier

Xinyi Feng[a]

[a]*xfeng82@wisc.edu*

1 Abstract

In some circumstances when data show different characteristics, classifiers like what we have learned in this course may not always give us the correct data classification. In this case, we can introduce a new kind of algorithm called Naive Bayes Classifier (NBC).

Unlike decision tree classification and k-means classification algorithms, Naive Bayes Classifier mainly uses the knowledge of probability theory to compare the conditional probabilities of the data come from each class. We calculate them separately, and then predict the category with the largest conditional probability.

2 Math

The core of the naive Bayes classifier is the Bayes' Rule which tells us how to swap the symbols in a conditional probability statement:
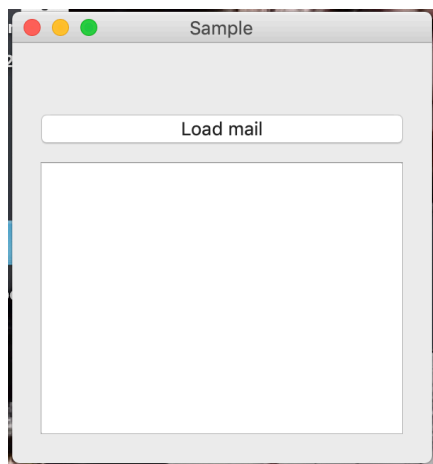
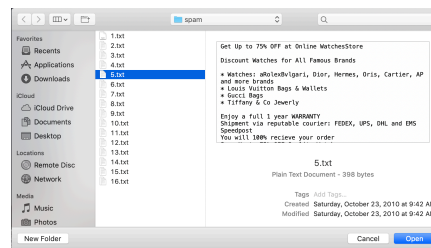$$p(c|x) = \frac{p(x|c)p(c)}{p(x)} \tag{1}$$

3 Background

To use naive Bayes on some real-life problems well need to be able to go from a body of text to a list of strings and then a word vector. In this example were going to visit the famous use of naive Bayes: email spam filtering.

We prepared a video for this part.
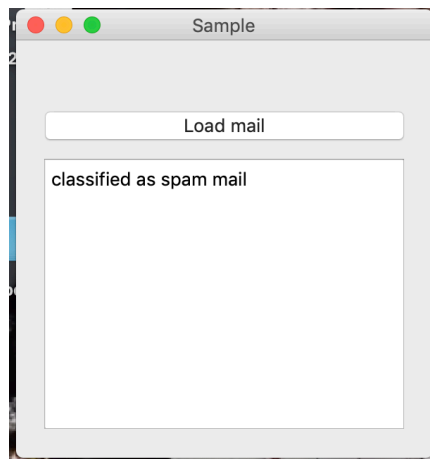
4 UI Interface

I made a simple UI Interface with Qt designer, PyUIC and Pyinstaller:

(a) Fig1

(b) Fig2

(c) Fig3

Figure 1: Preview of the software