Naive Bayes Classifiers for Spam Filtering

Sk. Amir Hamza
2017-1-60-091
Project Documentation

Introduction

Email is very important in our every day life. Every day we found our inbox in huge number of mail. But all mail are not important. Some mail are harmful and danger for our daily lift. In this problem I try to filter those harmful mail using Naive Bayes Classifiers.

Procedure

I follow this paper https://courses.cs.washington.edu/courses/cse312/18sp/lectures/naive-bayes/naivebayesnotes.pdf for applying Naive Bayes Classifiers in our given problem.

Code Explanation

I use python 3 for solve this problem. Firstly I store all unique word, count every word occurrence in ham and spam, total spam train mail,total ham train mail. Then I find the probability of ham and spam mail and probability of all word to all unique word list with respect to ham and spam. I store all calculation in multiple array.
Now I open test mail list. I separate all word in every mail and sum total probability of spam and ham to follow the above paper. Then I check which is probability is grater. If spam probability is grater then our mail predicted spam other wise our mail predicted ham.
Then I calculate precision, recall, accuracy, f1 measure and write result file in program output.

Output Analysis

Precision: 95.6896551724138
Recall: 91.13300492610837
F-Measure: 93.35576114381833
Accuracy: 92.10000000000001