Report: PRML Assignment 3

Keval Dodiya (CS19M023)

Karan Jivani (CS19M030)

Dataset reference:

http://www.aueb.gr/users/ion/data/enron-spam/preprocessed/enron1.tar.gz

Algorithm:

We have implemented a Naive Bayes classifier. That corresponding steps are given below.

1) **Data loading and Tokenization**
   We read all emails one by one and split it into words. Then remove special characters out of words and made a dictionary. Dictionary contains words with its frequencies.

2) **Training the model**
   We made 2 csv files one for spam and another for ham. Both are dictionaries,that contains words with its frequencies.so first we tokenize the mail and adds those words to corresponding dictionaries.

3) **Testing/Prediction**
   Suppose we have email E = {e1,e2,e3,.......} where e_i is a word. Now from bayes theorem.

$$P(spam \mid e_1 \cap e_2 \cap e_3....) = \frac{P(spam)P(e_1 \cap e_2 \cap e_3 .....|spam)}{P(e_1 \cap e_2 \cap e_3....)}$$

And in naive bayes we assumes all words are independent hence…

$$P(spam \mid e_1 \cap e_2 \cap e_3....) = \frac{P(spam)\,P(e_1\,|spam)\,P(e_2|spam)\,P(e_3|spam)...}{P(e_1 \cap e_2 \cap e_3....)}$$

So to classify upcoming email spam or ham we need to find $P(spam \mid e_1 \cap e_2 \cap e_3....)$ and $P(ham \mid e_1 \cap e_2 \cap e_3....)$.
Whichever probability is maximum, upcoming email will go into that class.

$$P(\,e_1\,) = \frac{total\ number\ of\ e_1\ in\ dataset}{total\ number\ of\ words\ in\ dataset}$$

$$P(\,e_1 \mid spam\,) = \frac{total\ number\ of\ e_1\ in\ spam\ emails}{total\ number\ of\ words\ in\ spam\ emails}\,.$$

We encountered the case where some features are not available in the dictionary so that feature become zero. That issue solved by additive smoothing that formulae given below.

$$P(\,e_k \mid spam\,) \;=\; \frac{n_k + 1}{n + |dictionary|}$$

Where, n_k = number of ek in spam messages.

n = total words in spam messages.

And last we computed $P(spam \mid e_1 \cap e_2 \cap e_3 ....)$ and $P(ham \mid e_1 \cap e_2 \cap e_3 ....)$ to classify email as spam or ham.