

Email Spam Filter

By:

- Nisarg Hareshbhai Shah
- Sugandha Kher





Naive Bayes Classifier

- Bayes' theorem written as:

$$P(A|B) = P(B|A)P(A) / P(B)$$

- To calculate the Spam email, we could use the above formula as:

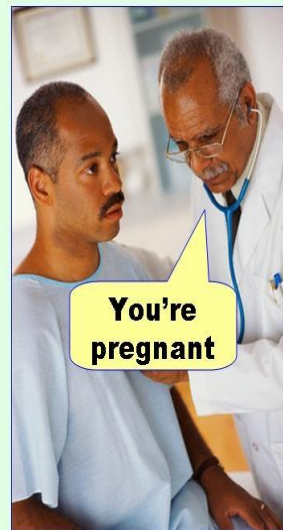
$$P(\text{Spam} | \text{EmailContent}) = P(\text{EmailContent} | \text{Spam}) P(\text{Spam}) / P(\text{EmailContent}).$$

- If $P(\text{Spam} | \text{EmailContent}) > P(\sim\text{Spam} | \text{EmailContent})$, email is classified as Spam.
- Traditional spam filter tends to use bag-of-word model to calculate the frequency of occurrence for a word in order to train the classifier.
- But, we plan to use tf-idf model which takes into consideration the importance of word in a training set.
- Dataset Used: Email Spam Dataset of CSMining group.

Evaluation

- Precision: Fraction of relevant instances among retrieved instances
- Recall: Fraction of relevant instances retrieved over a total number of relevant instances
- False Positive: Emails marked as spam when they are not spam
- False Negative: Emails marked as ham when they are spam
- True Positive: Emails marked as spam when they are spam
- True Negative: Emails marked as ham when they are ham

Type I error
(false positive)



Type II error
(false negative)

