# Machine Learning

**Assignment 8.4**

Submitted By: Ranji Raj

December 19, 2020

Let the input to Naive Bayes classifier be **bag of words, W** such that $w_1, .., w_n \in W$ which we can process into **list**, S comprising words, W.

## a) Estimating probability of a word as spam or ham

Let, $\boldsymbol{P(\neg spam|s[W])}$=Probability that an email is not-spam given that it contains list of words. $\boldsymbol{P(spam|s[W])}$=Probability that an email is spam given that it contains list of words.
Therefore,

$$P(\neg spam|S[w]) = \frac{P(S[w]|\neg spam).P(\neg spam)}{P(S[W])}$$

and

$$P(spam|S[w]) = \frac{P(S[w]|spam).P(spam)}{P(S[W])}$$

| sr.no. | doc | category |
|---|---|---|
| | | |
| | | |
| | | |
| | | |

Table 1: Training data

| sr.no. | word | count | category |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |

Table 2: Frequency table

From Training data we can obtain,

$$P(\neg spam) = \frac{\text{Number of documents belonging to not-spam}}{\text{Total number of documents}}$$

$$P(spam) = \frac{\text{Number of documents belonging to spam}}{\text{Total number of documents}}$$

From frequency table we can obtain,

$$P(S[w]|\neg spam) = P(w_1|\neg spam), ..., P(w_n|\neg spam)$$

$$P(S[w]|spam) = P(w_1|spam), ..., P(w_n|spam)$$

$$P(w_1|spam) = \frac{\text{Count of word 1 in category spam}}{\text{Total number of words in category spam}}$$

$$P(w_1|\neg spam) = \frac{\text{Count of word 1 in category not-spam}}{\text{Total number of words in category not-spam}}$$

We can drop $P(S[W])$ because it is constant and won't affect the estimation.

## b) Calculating probability of the size of the mail wrt. spam or ham

Bayesian probability for **single keyword** $k$ is given as,

$$P(k) = \frac{s(k)}{s(k) + \neg s(k)}$$

where, $s(k)$ is the number of spam emails with keyword k and $\neg s(k)$ is the number of not-spam emails with keyword k.
Bayesian probability for **single keyword set** $ks$ is given as,

$$P(ks) = \frac{s(ks)}{s(ks) + \neg s(ks)}$$

where, $s(ks)$ is the number of spam emails with single keyword set ks and $\neg s(ks)$ is the number of not-spam emails with single keyword set ks.
Bayesian probability for **multi-keyword set** $ks$ is given as,

$$P(mk) = \frac{s(mk)}{s(mk) + \neg s(mk)}$$

where, $s(mk)$ is the number of spam emails with multi-keyword set mk and $\neg s(mk)$ is the number of not-spam emails with multi-keyword set mk.

Two keywords are assigned a weight of $MK_{WEIGHT}$ (constant value), three keywords are assigned a weight of $MK_{WEIGHT} * 3$, four keywords or more are assigned a weight of $MK_{WEIGHT} * 4$. Single keywords are not assigned any weights.

The keyword scores are totaled to get the spam score for a given mail.

## c) Problems encountered when using regular Naive Bayes

Possibility that our classifier detects a new word that is not present in training data. In that case it multiplicative probability will be equal to zero.

To mitigate this we use Laplace smoothing,

$$P(w|spam) = \frac{\text{Count of word belonging to category spam} + 1}{\text{Total count of words belonging to spam} + \text{number of distinct words in training data}}$$