Date / /

$$P(\text{doesn't like 'T' \& like 'RJ'} \mid \text{likes 'RJ'}) = \frac{P(\text{doesn't like 'T' \& likes 'RJ'} \mid \text{doesn't like 'T'}) \times P(\text{likes 'RJ'})}{P(\text{doesn't like 'T'})}$$

So, we derived Baye's theorem.

Naive Baye's :

It is a classification technique based on Baye's theorem with an assumption of independence among predictors. In simple terms a Naive Bayes classifier assumes that the presence of a particular feature of a class is unrelated to the prese-nce of any other feature.

For example a fruit may be considered to be an apple if it is red, round and about 3 inches in diameter. Even if these features depend upon each other, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'naive'.

$$P(y|x) = \frac{P(x|y)\,P(y)}{P(x)}$$

likelihood; class Prior Probability

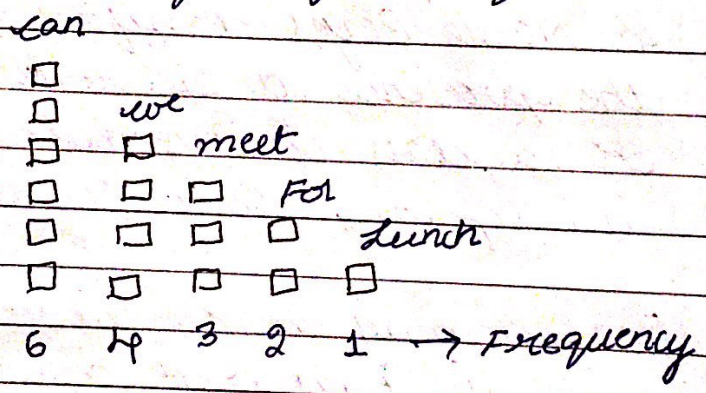
$P(y|x)$ → Posterior Probability

$P(x)$ → Predictor Probability

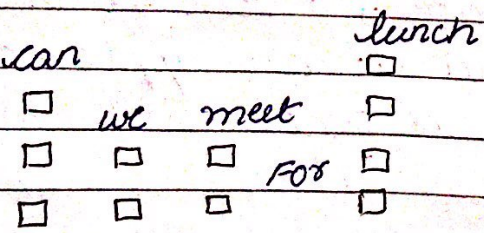
Let's take an example, imagine we received normal email from friends and family and also recieved spam (unwanted messages that are usually (solicites) scams) scams). Now we want to filter out the spam messages.

To solve this lets make a histogram of all the words that occur in normal emails from friends and family.

Message from family and friends : 10 ma



Spam Mails : 5 mails



Date / /

Contigency / likelihood table: These are constructed by listing all the levels of one variable as rows in a table and the levels of other variables as columns, then finding the joint of cell frequency of each cell.

	Normal Emails (N)	Spam (S)
can	6 [0.375]	3 (0.27)
we	4 [0.25]	2 (0.18)
meet	3 [0.19]	2 (0.18)
For	2 [0.13]	0 (0)
Lunch	1 [0.06]	4 (0.36)

$P(can|N) = \frac{6}{16} = 0.375$ $P(can|S) = \frac{3}{11} = 0.27$

$P(we|N) = \frac{4}{16} = 0.25$ $P(we|S) = \frac{2}{11} = 0.18$

$P(meet|N) = \frac{3}{16} = 0.1875$ $P(meet|S) = \frac{2}{11} = 0.18$

$P(For|N) = \frac{2}{16} = 0.13$ $P(For|S) = 0$

$P(Lunch|N) = \frac{1}{16} = 0.06$ $P(Lunch|S) = \frac{4}{11}$

Now Let's say you got an email 'meet for lunch' we want to decide if it is a normal mail or spam.

The guess is estimated from the training data.

Total messages = 10 Normal + 5 Spam = 15

$$P(N) = \frac{10}{15} = 0.66$$

Now we multiply the probability with initial guess.

$$P(N) \times P(can|N) \times P(we|N) \times P(meet|N) \times P(For|N) \times P(Lunch|N)$$

$$= 0.000092$$

Likewise do for spam and we get the value as 0

Since $0.000092 > 0$ it goes into the normal mail.

Let's take another example, 'For Lunch Lunch Lunch'. Now calculate like above we get

Spam mail = 0. but the mail should be classified as Spam mail. The work around is add 1 for each word. So, if we recompute [P(For|S) = 1/11+14 = 1/45)

Normal mail = $0.66 \times 0.14 \times (0.09)^3$
= 0.024

Spam mail = $0.50 \times 0.06 \times (0.33)^3$
= 0.029

So, with this work around we are able to classify the mails correctly.

Naive Bayes treats language like it just a bag full of words. It ignores all the rules because keeping track of every single reasonable phrase in a language would be impossible. A Naive Bayes ignores relationship it has high bias but it works well in practice so, it has low variance.

Different variants in Naive Bayes:

(i) Bernoulli Naive Bayes:

It is useful if your feature vectors are binary. One application would be text classification when 1 & 0 are 'word occurs' and 'word doesn't occur' in document respectively

Bernoulli distribution

$P(\text{True}) = p$

$P(\text{False}) = q$

Let's say x is a random variable

$x = 1$ [True]

$x = 0$ [False]

$$P(X = x) = p^x (1-p)^{1-x}$$

$$P(x) \begin{cases} p & \text{if } x = 1 \\ q & \text{if } x = 0 \end{cases}$$

(ii) Multinomial dis Naive Baye's: It can be used to find the occurence of word i.e., the frequency in a document i.e., discrete count.

Multinomial distribution:

$$P(X_1 = x_1, \cdots, X_n = x_n)$$

$$= \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_n^{x_n}$$

(iii) Gaussian NB: If the data is continious in the features then it is useful.

PDF:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$-\infty < x < \infty$