7.1 What's is the Bayes Algorithm?

Bayes' theorem is simply to use prior probabilities and inverse conditional probabilities to find other conditional probabilities. We can use Law of Total Probability when we know reasons and need to get the probability of results; but we can use Bayes' Formula when we know results and need to get the probability of reasons. **In essences, the reasons and the results just make a transposition**.

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B) \cdot P(A|B)}{P(A)} \tag{1}$$

- P(B): Prior probability P(A)
- P(A) or P(B): marginal probability
- P(B|A): Conditional probability
- P(A|B): Inverse conditional probability, or Posterior probability
- P(AB): Joint probability

7.2 Deduction of Bayes Algorithm

The deduction process is simple, but we need to be familiar with several formulas of probability which are shown as follows:

7.2.1 Multiplication formula

$$P(AB) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \tag{2}$$

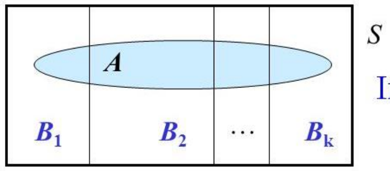
7.2.2 Law of total probability

(Reason → Result)

$$P(A) = \sum_{i=1}^{k} P(B_i) \cdot P(A|B_i), \qquad P(B_i) > 0 \tag{3}$$

Proof:

Law of Total Probability



If $P(B_i) \neq 0$, then

$$P(A \cap B_i) = P(A \mid B_i)P(B_i)$$

Hence, for the partition $\{B_1, B_2, \ldots, B_k\}$ of the sample space S,

we have $P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k)$

or equivalently,

$$\boxed{P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_k)P(B_k).}$$

7.2.3 Bayes Formula

(Result → Reason)

$$P(B_i|A) = \frac{P(B_i \cdot A)}{P(A)} = \frac{P(B_i) \cdot P(A|B_i)}{\sum_{i=1}^{k} P(B_i) \cdot P(A|B_i)} \tag{4}$$

7.3 Examples for Bayes Algorithm

7.3.1 Ensure gender based on wearing pants or skirts

Suppose there are 60% of students in school are boys and 40% are girls. We have known all of boys wear pants, but there are 50% of girls wear pants and 50% of girls wear skirts.

- **What's the probability if we pick a student who wears pants?**

This is a question about Law of Total Probability. **(Reason → Result)**

$$\begin{aligned} \text{P(pants)} &= \text{P(boy)} \cdot \text{P(pants | boy)} + \text{P(girl)} \cdot \text{P(pants | girl)} \\ &= 60\% \cdot 100\% + 40\% \cdot 50\% \\ &= 80\% \end{aligned}$$

- **What's the probability of the student we pick is a girl if he/she wears pants?**

This is a question about Bayes Formula. **(Result → Reason)**

$$\begin{aligned} P(girl|pants) &= \frac{P(girl \cdot pants)}{P(pants)} = \frac{P(girl) \cdot P(pants|girl)}{P(girl) \cdot P(pants|girl) + P(boy) \cdot P(pants|boy)} \\ &= \frac{40\% \cdot 50\%}{40\% \cdot 50\% + 60\% \cdot 100\%} \\ &= 25\% \end{aligned}$$

7.3.2 Guess words based on spelling

Assume a user typed in a word ‘tlp’ (Input data: D), we can guess he wanted to type in ‘top’ (Guess h1) or ‘tip’ (Guess h2), but what on earth does he want to input? In this case, we need to compute P(h1|D) and P(h2|D) to see which one bigger is. Due to hard computation, we can give it a transition based on Bayes Formula:

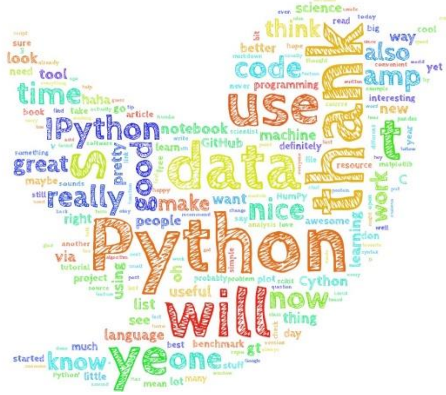
$$P(h|D) = \frac{P(h \cdot D)}{P(D)} = \frac{P(h) \cdot P(D|h)}{P(D)} \tag{5}$$

From the Eq. (5) we can figure out the probability of input data D, P(D), is always same for every guess, h1, h2. h3, …… Therefore, it would be fine to consider the numerator only.

$$\begin{gathered} P(h|D) \propto P(h) \cdot P(D|h) \\ P(top|tlp) \propto P(top) \cdot P(tlp|top) \\ P(tip|tlp) \propto P(tip) \cdot P(tlp|tip) \end{gathered} \tag{6}$$

Now, we can see that $P(h|D)$ has something to do with two parts: $P(h)$ and $P(D|h)$.

- Prior probability: $P(h)$
  It has been computed in advance. For example, there is a huge corpus including h1, h2, h3…. Let’s say there are 10,000 words in it, and h1 appears 1,000 times, h2 appears 500 times. In this way, P(h1|D) = 1,000/10,000 = 1/10, P(h2|D) = 500/10,000 = 1/20.



- Inverse conditional probability: $P(D|h)$
  The possibility of h being D.

7.3.3 Classification of spam and normal emails

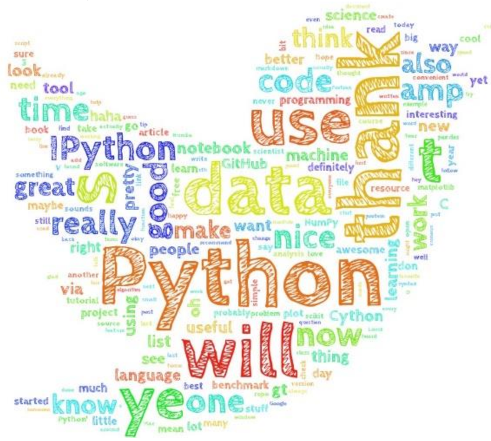
Assume D is an email with many words, h+ means normal email and h- means spam email. We can compute P(h+|D) and P(h-|D) as below:

$$\begin{gathered} P(h_+|D) = \frac{P(h_+ \cdot D)}{P(D)} = \frac{P(h_+) \cdot P(D|h_+)}{P(D)} \\ P(h_-|D) = \frac{P(h_- \cdot D)}{P(D)} = \frac{P(h_-) \cdot P(D|h_-)}{P(D)} \end{gathered} \tag{7}$$

Similarly,

$$\begin{gathered} P(h_+|D) = P(h_+) \cdot P(D|h_+) \\ P(h_-|D) = P(h_-) \cdot P(D|h_-) \end{gathered} \tag{8}$$

- Prior probability: $P(h_+)$ and $P(h_-)$
  It has been computed in advance. For example, there is a huge email library. Let's say there are 1,000 emails in it, and 900 emails are good, 100 emails are spam emails. In this way, P(h+|D) = 900/1,000 = 90%, P(h-|D) = 100/1,000 = 10%.



- Inverse conditional probability: $P(D|h)$
  $P(D|h_+)$ means the probability of email is made up of D on the condition of it is spam email. However, D here consists of many words, like d1, d2, d3, ……

$$P(D|h_+) = P(d_1, d_2, \dots, d_n|h_+) \tag{9}$$

Naïve Bayes Algorithm hypothesize the features $d_1, d_2, \dots, d_n$ are independent with each other. Therefore, we can get:

$$P(D|h_+) = P(d_1, d_2, \dots, d_n|h_+) = P(d_1|h_+) \cdot P(d_2|h_+) \cdot \dots \cdot P(d_n|h_+) \tag{10}$$

Note: $P(d_i|h_+)$ means the frequency of d_i appearing in the spam email.

7.3.4 Predict the probability of cancer based on positive testing

A kind of cancer, people with all of the cancer have 90% chance of being tested positive, without the this cancer were detected negative 90% (hint: the probability without cancer but positive is 10%), the risk of this cancer is 1% in the crowd. **Question: what is the probability that a person who is tested positive get cancer?**

$$P(cancer|positive) = \frac{P(cancer) \cdot P(positive|cancer)}{P(positve)} \tag{11}$$

- P(cancer) = 0.01
- P(positive | cancer) = 0.9
- P(positive) = P(cancer)·P(positive | cancer) + P(not cancer)·P(positive | not cancer)
  = 0.01 × 0.9 + 0.99 × 0.1 = 0.108

$$P(cancer|positive) = \frac{P(cancer) \cdot P(positive|cancer)}{P(positve)} = \frac{0.01 \times 0.9}{0.108} = 0.083$$