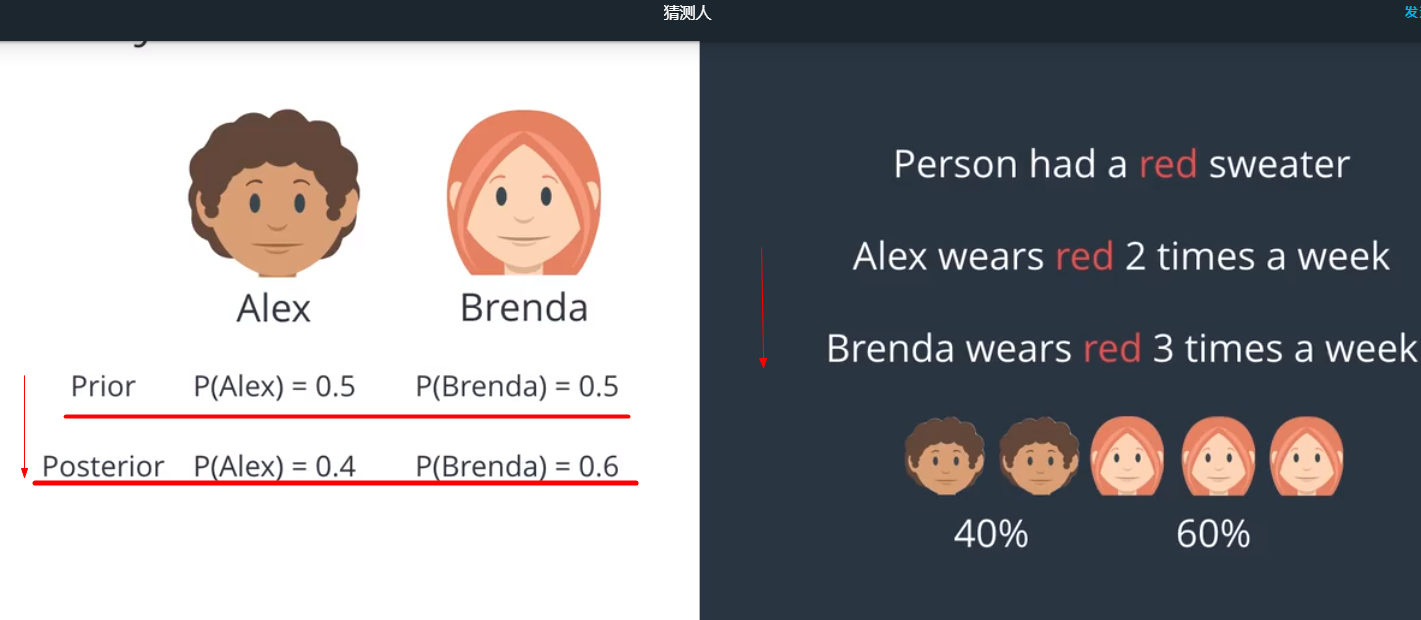
# 2猜测人



我们利用了毛衣的颜色 得到了**猜测过路人身份**更准确的概率 这称为**贝叶斯定理**

起初的猜测概率为 50% 50% , 称为**先验概率**

因为这是我们在没有新信息红毛衣之前做出的推断

而最后猜测的概率为 60% 40% 称为**后验概率**

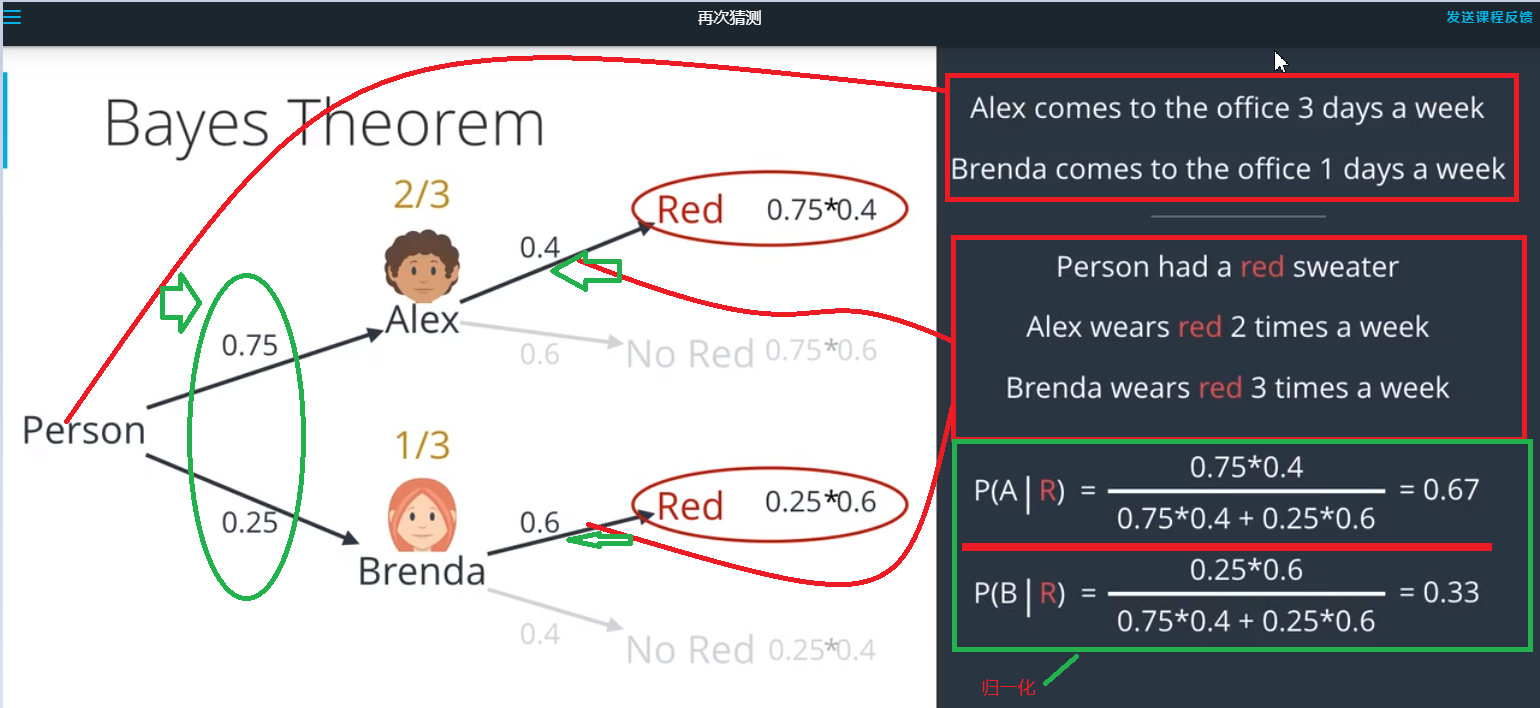
因为这是我们在得到红毛衣这条新消息之后做出的推断

3.已知推断

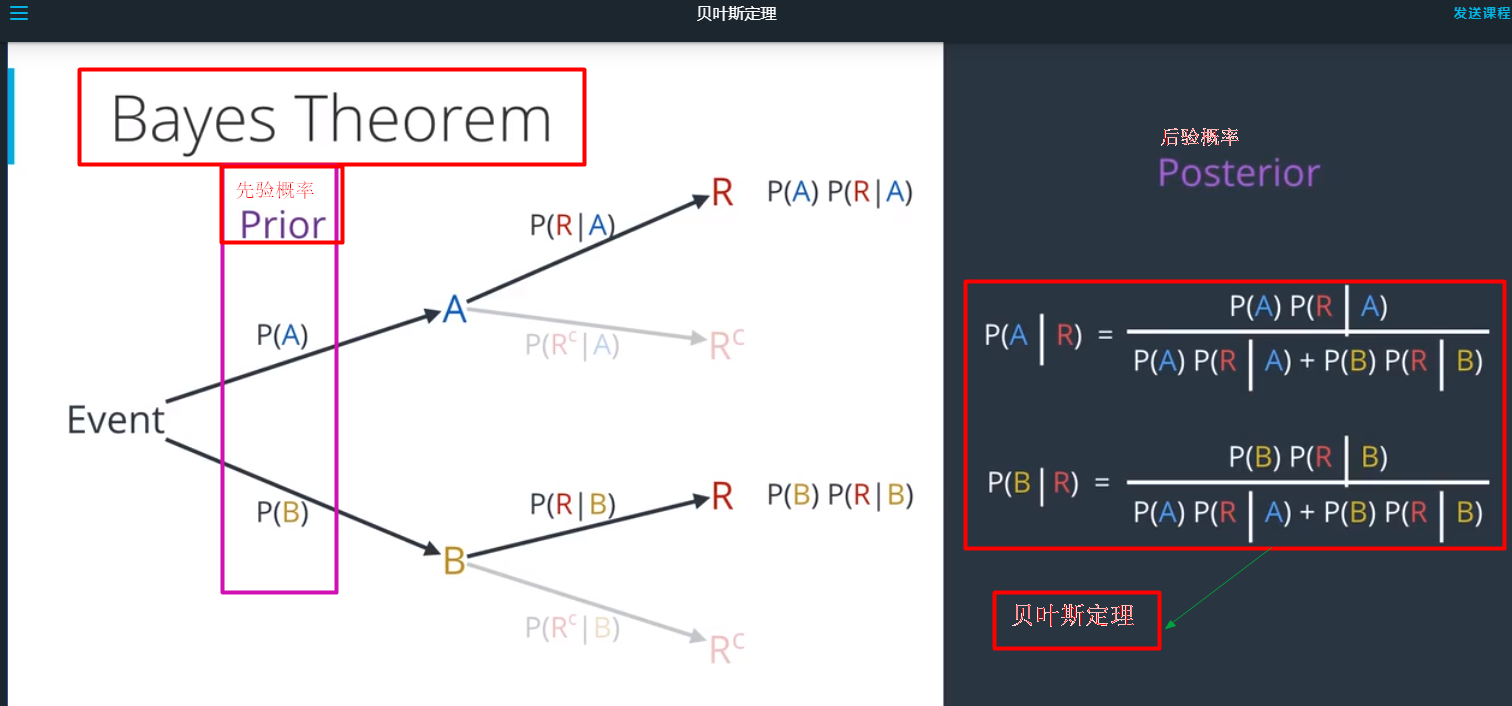
贝叶斯定理 是一个非常强大的定理,它的工作是将我们知道的 转换为 我们推测的 。

<https://zh.wikipedia.org/wiki/%E8%B4%9D%E5%8F%B6%E6%96%AF%E5%AE%9A%E7%90%86>

4.再次猜测



# 5.贝叶斯定理



**P(A) 和 P(B) 被称为先验概率**

是我们在不知道 R 事件发生的情况下 计算得到的事件 A 和 B 的概率

P(A|R) 和 P(B|R) 被称作后验概率

是我们在已知 R 事件发生的情况下 推断出的 A 和 B 事件的概率

这就是贝叶斯定理

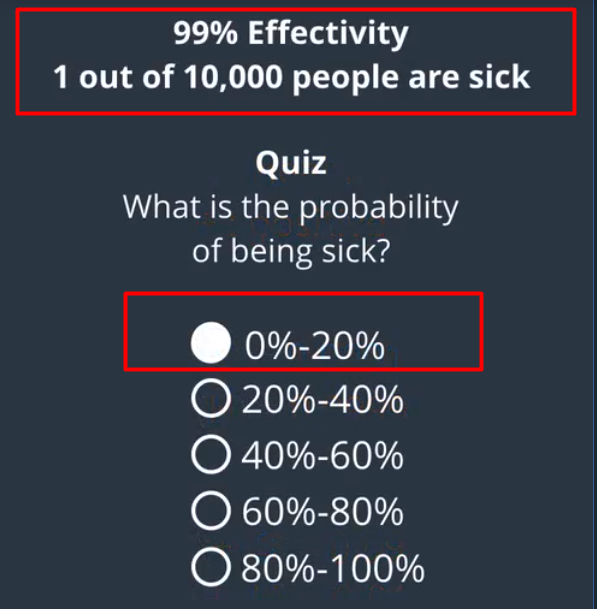
注意：

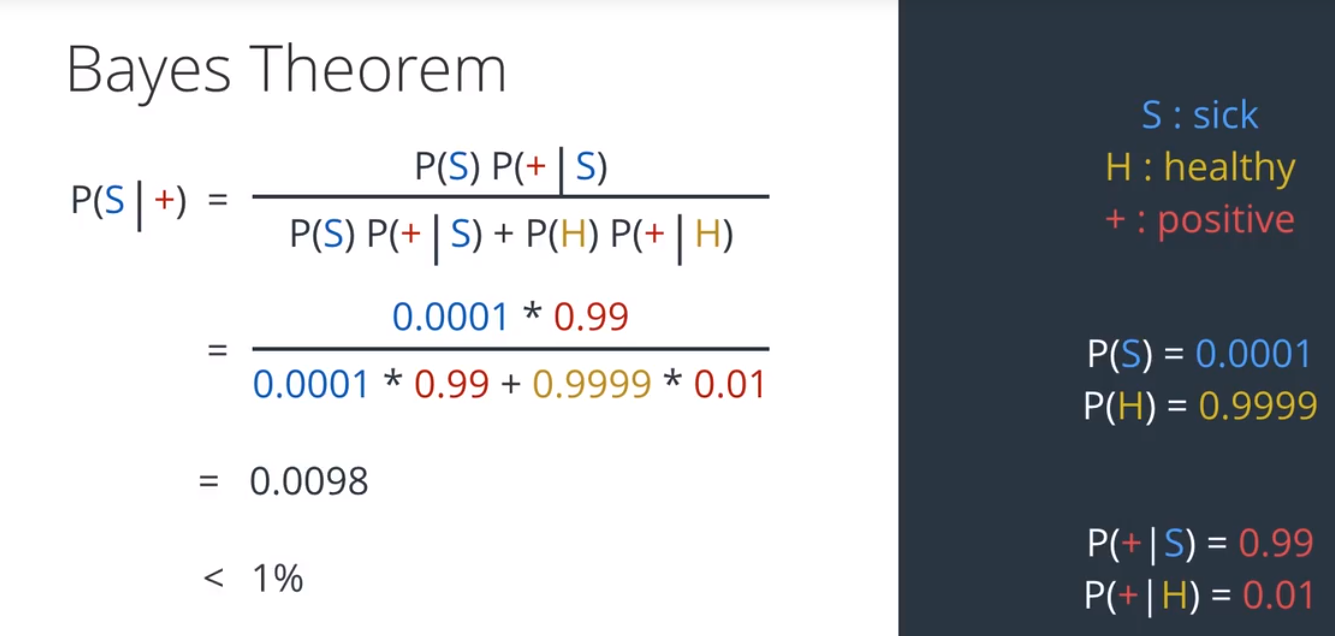
**已知 P(R|A)、P(R|B ) P(A), P(B)**

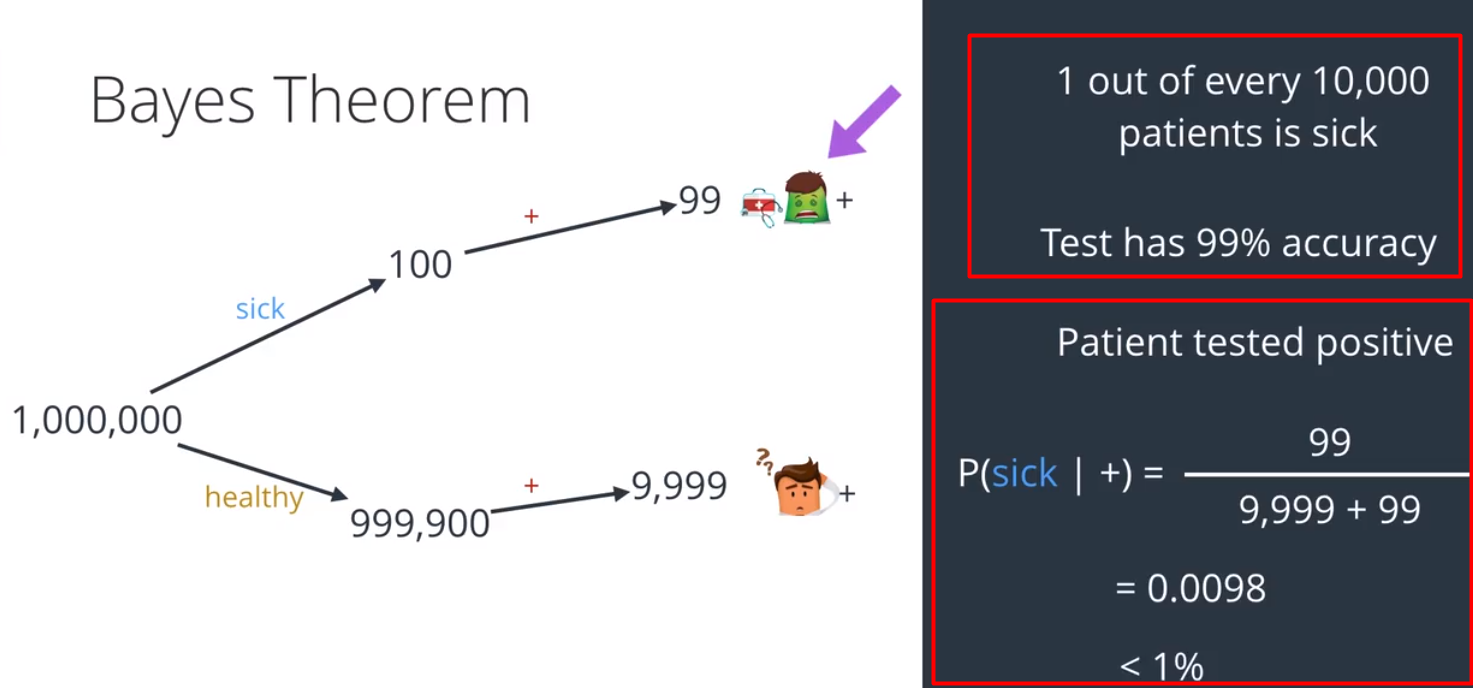
**求P(A|R)、P(B|R) )**

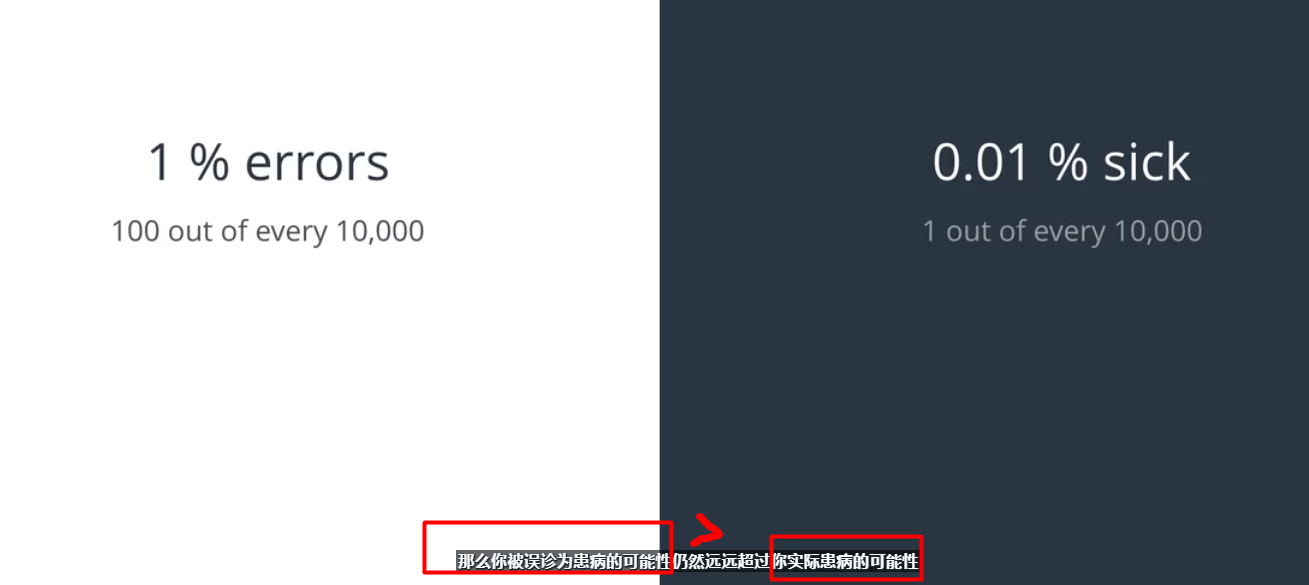
# 6.练习

检查准确率99%，与 万分之一的患病率，你认为生病的可能性是多少？

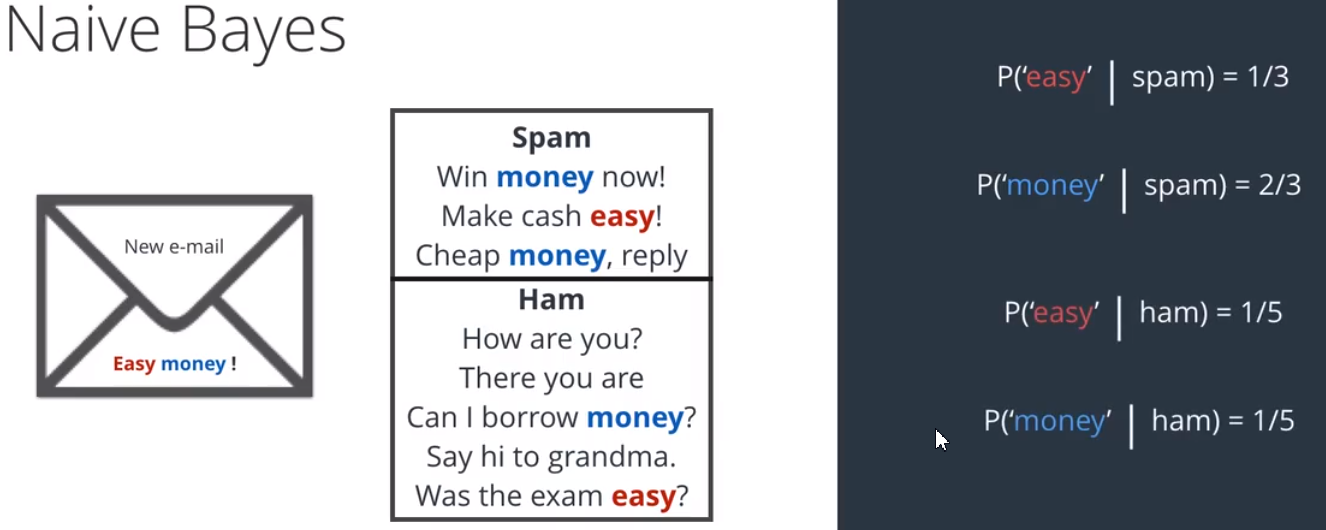


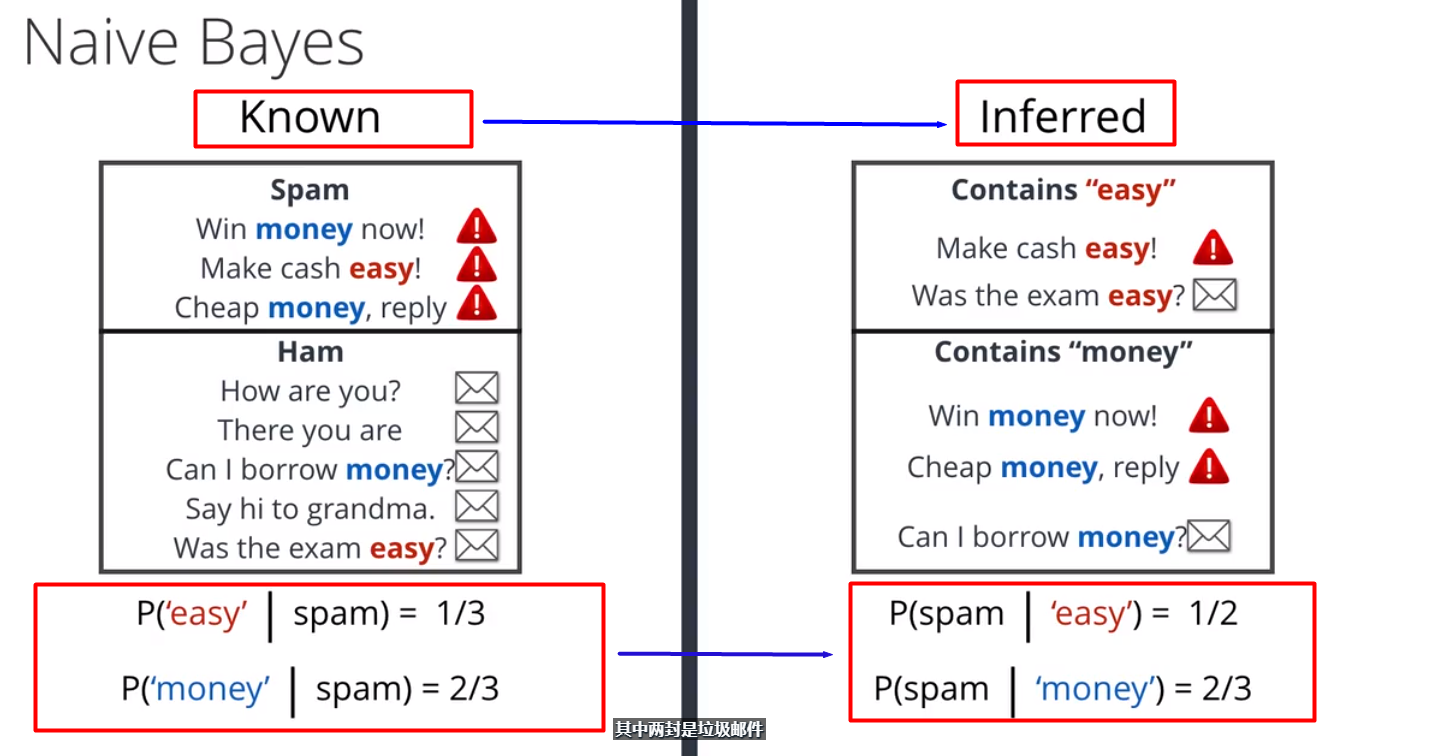




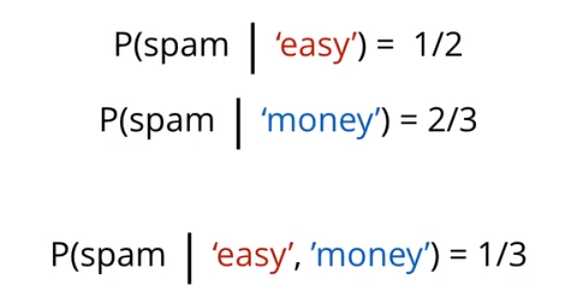


# 8.贝叶斯学习

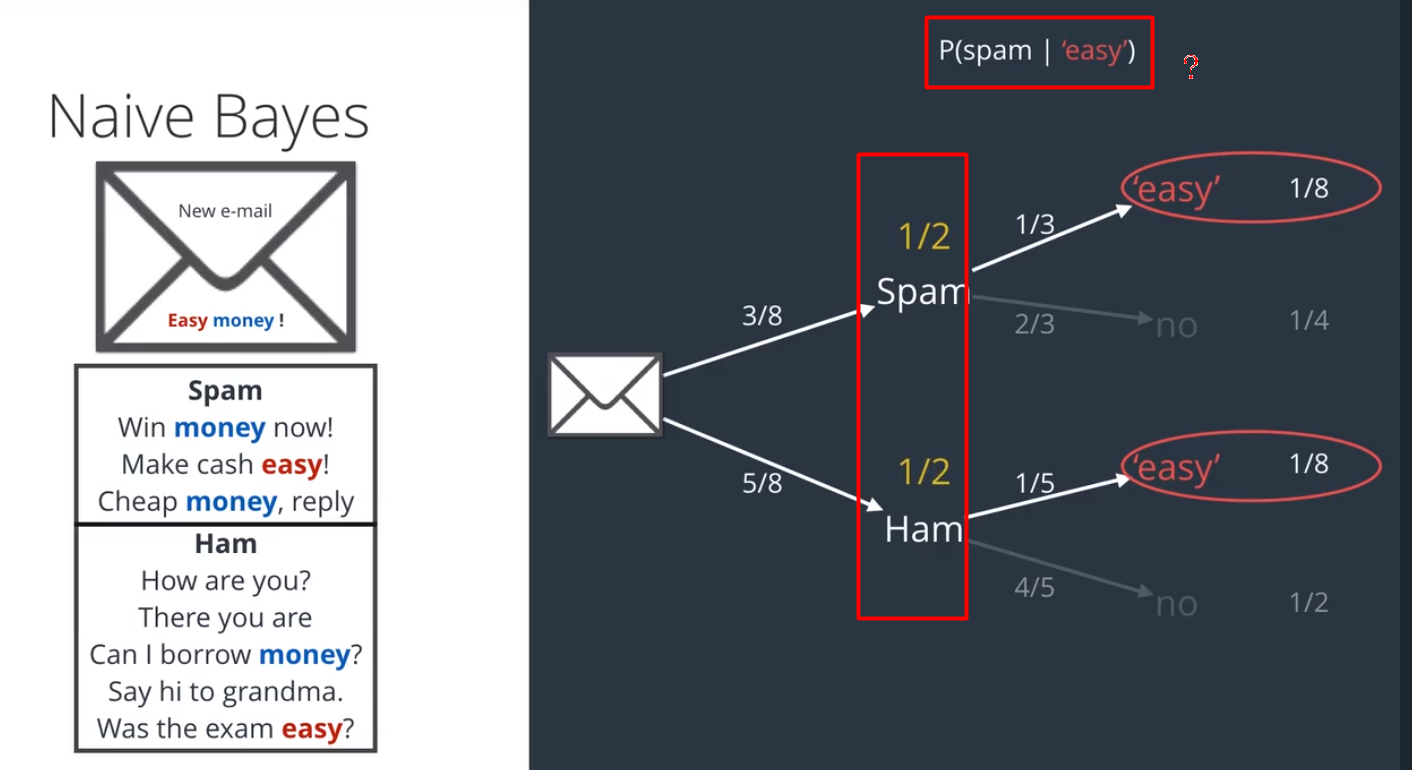




我们假设它们是彼此独立的 然后把它们相乘，得到的结果就是 如果 邮件中含有 “easy” 和 “money” 是垃圾邮件的概率。（1/2 \* 1/3 = 1/6）



# 10.贝叶斯学习3



**先验概率**

垃圾邮件 八分之三 非垃圾邮件 八分之五

**后验概率**

这两种情况发生的概率相同 都是八分之一

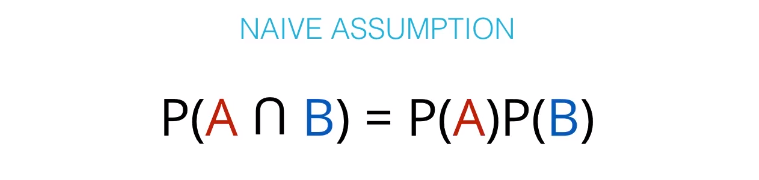
我们把这两个概率归一化

它们都是 50%

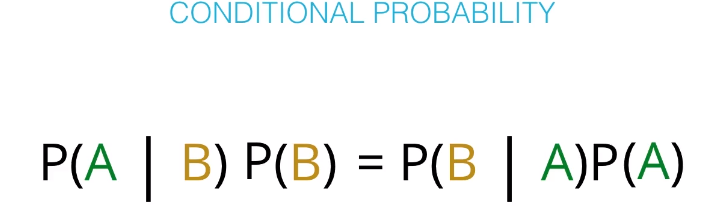
所以我们的后验概率都是 50%

# 11.朴素贝叶斯算法-重点

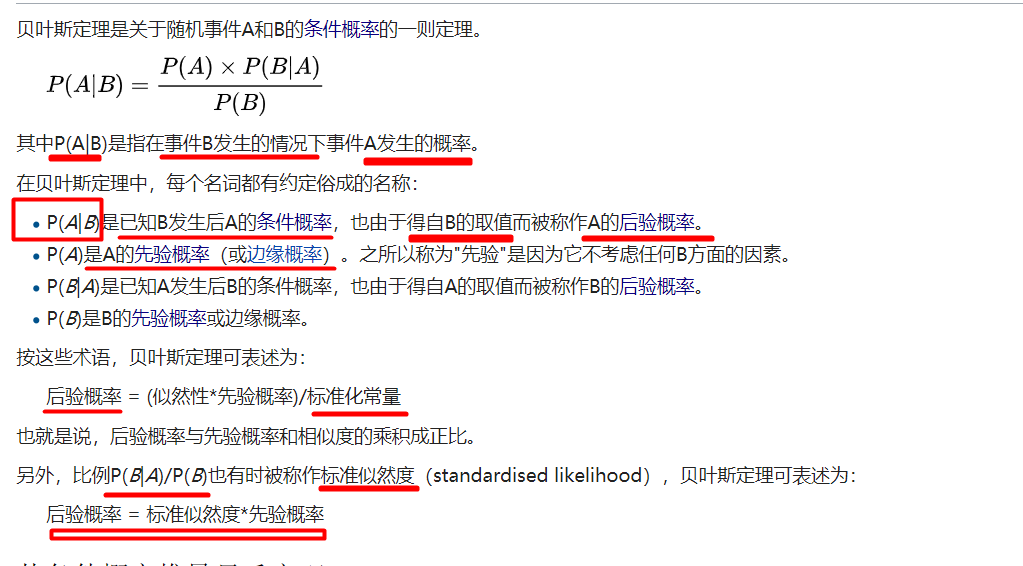
贝叶斯假设



条件概率



**以上是 A 交 B 的概率的两种写法。**



注意：

我们将做出一个非常朴素的假设,我们看看两个事件同时发生的概率

P(A & B)还可以写成 P(A ∩ B) 等于 P(A) 和 P(B) 的积

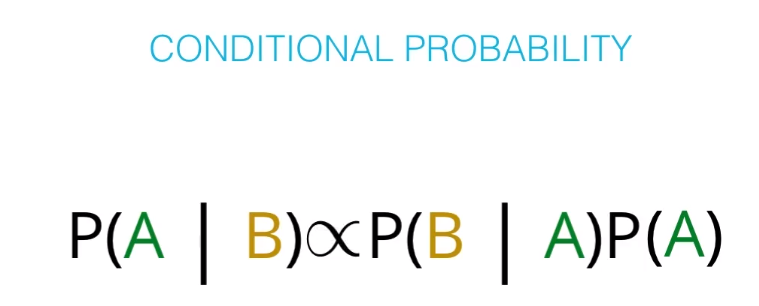
只有这两个事件相互独立时才会发生,如果不相互独立 则不成立

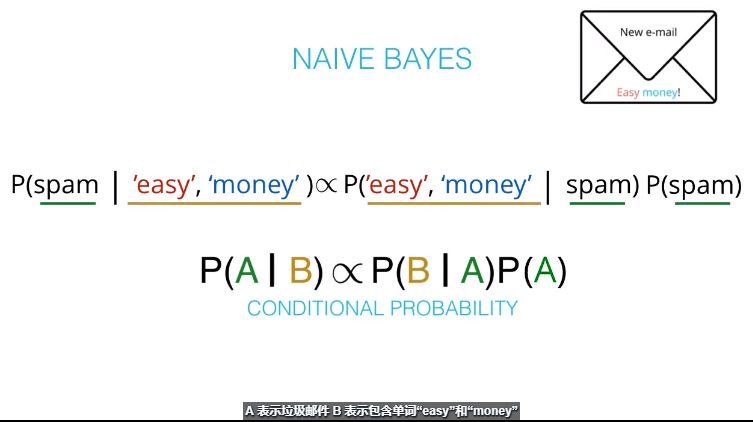
例如 如果 A 是外面很热的概率, B 是外面很冷的概率,它们的概率都为正,两个事件同时发生的概率是多少？

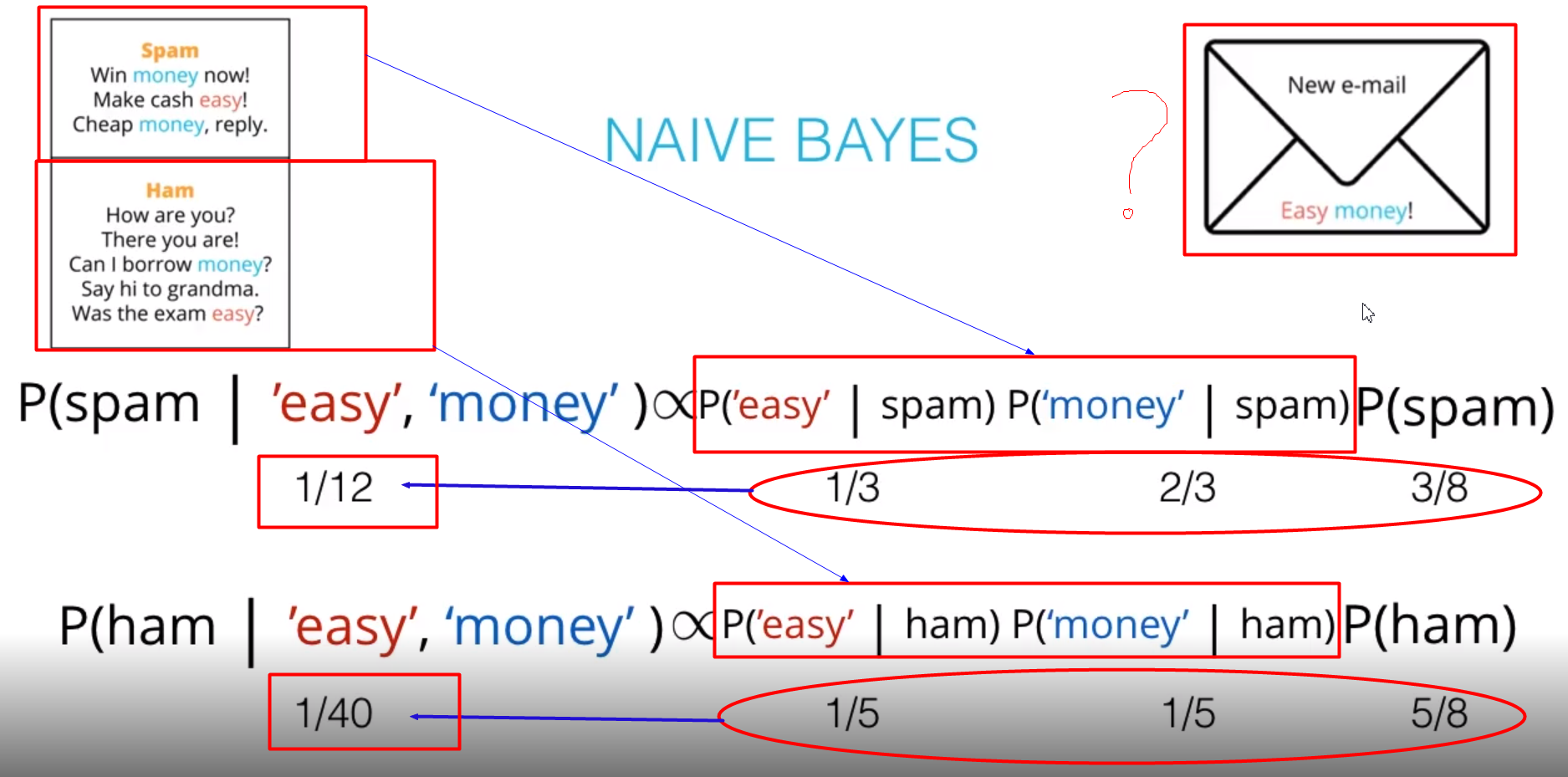
将为 0 因为不可能同时很热又很冷,这个公式不成立,因为很冷和很热是相互依赖的

**忽略P(B):**

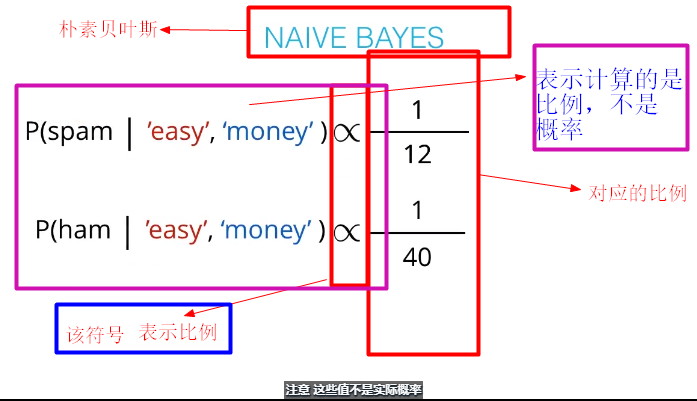
采用的技巧是忘记 P(B)，现在不是等于，而是变成 P(A | B) 与 P(B | A)P(A) 成比例



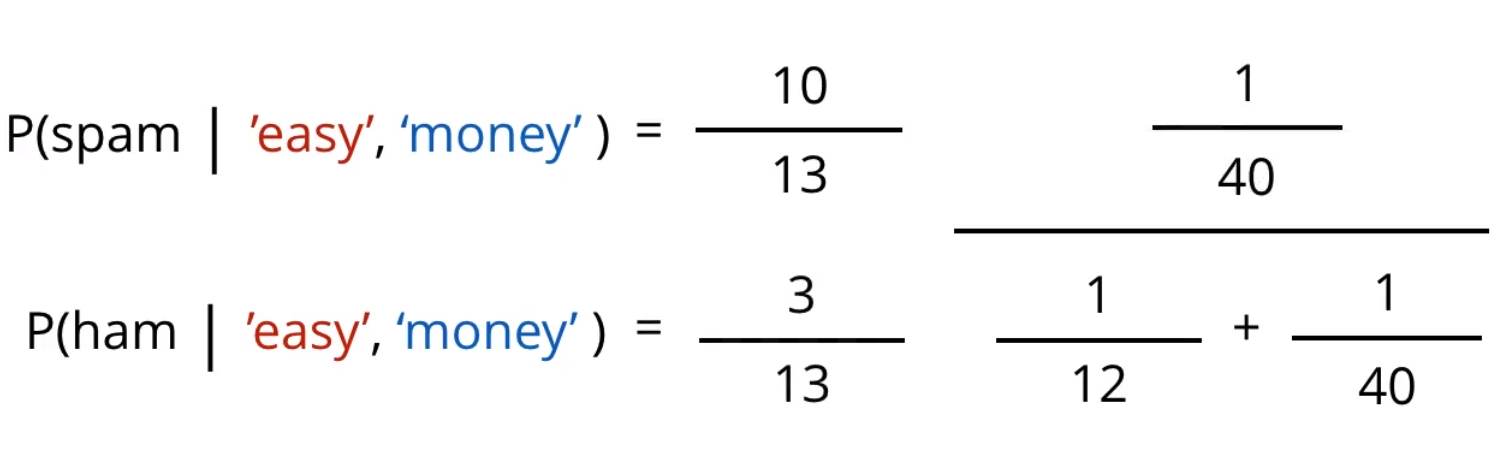




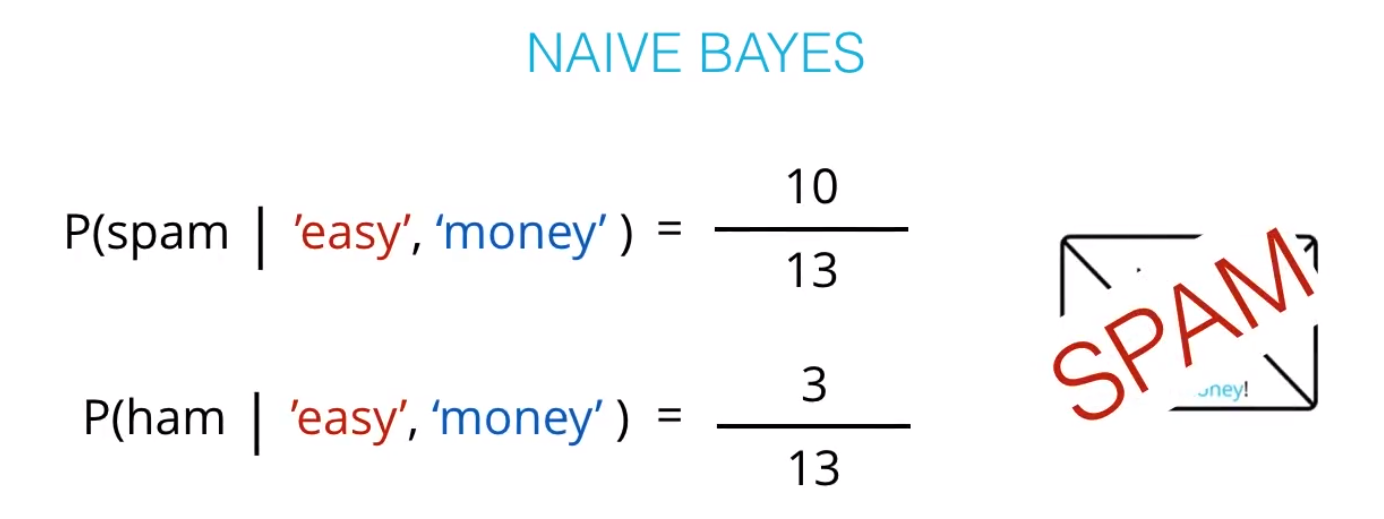
## 标准化



## 计算结果



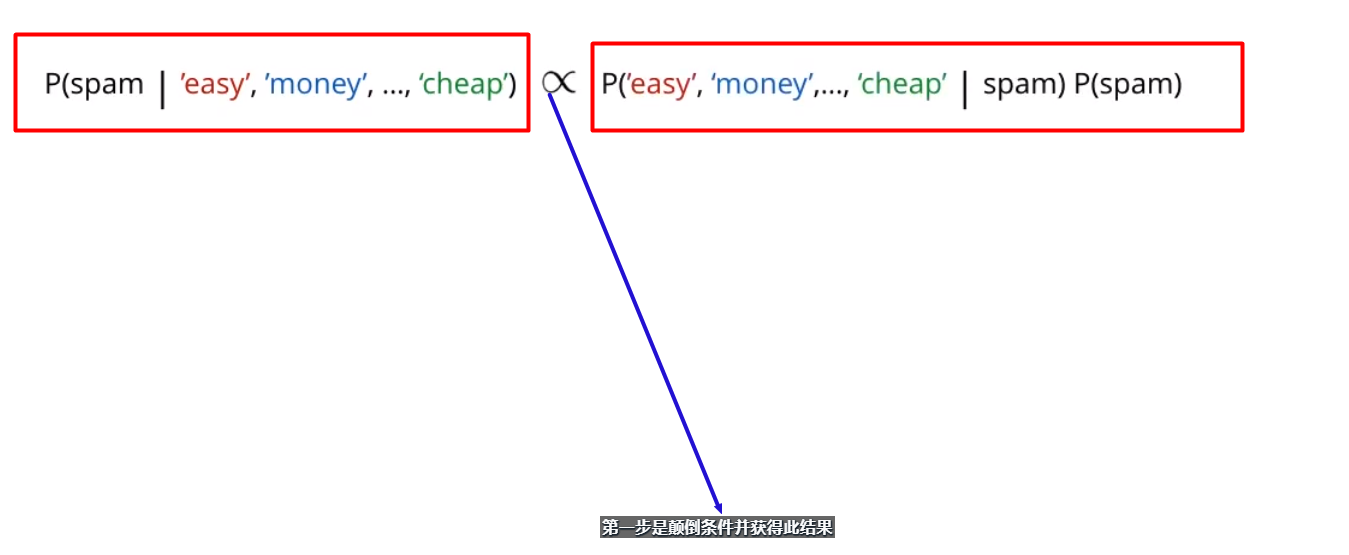
结论



结论是它很有可能是垃圾邮件

## 朴素贝叶斯原理

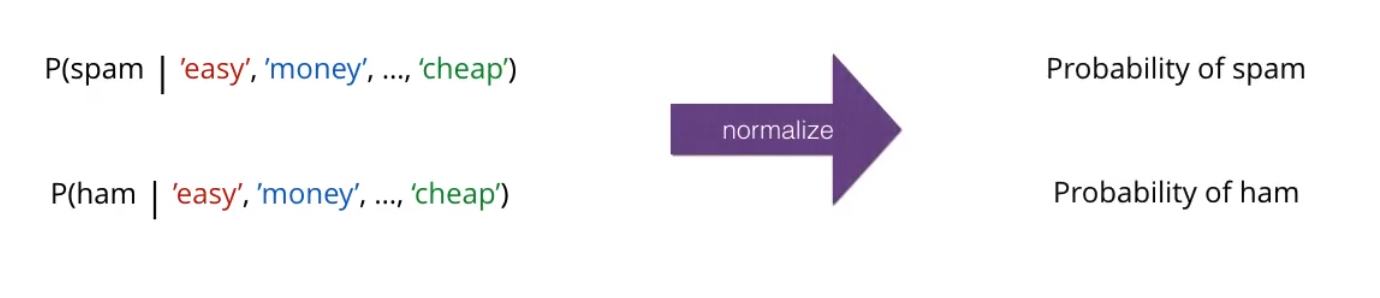
1、将概率的形式进行颠倒



2、转换为另一种方便计算的形式



3、标准化



**标准化计算方法**

**是将 每个概率 除以 二者之和， 确保它们加起来为 1**

**原理描述：**

条件：

假设有大量单词作为特征，可以告诉我们电子邮件是否为垃圾邮件，例如“easy” “money” “cheap”等。

第一步是颠倒条件并获得此结果

第二步做出朴素假设 将其拆分为多个简单因子的积

我们可以通过查看数据快速计算这些因子，对垃圾邮件和非垃圾邮件都执行此操作，获得一些和不等于 1 的值

最后 标准化数据，获得邮件是否为垃圾邮件的最终概率

这就是朴素贝叶斯算法的原理