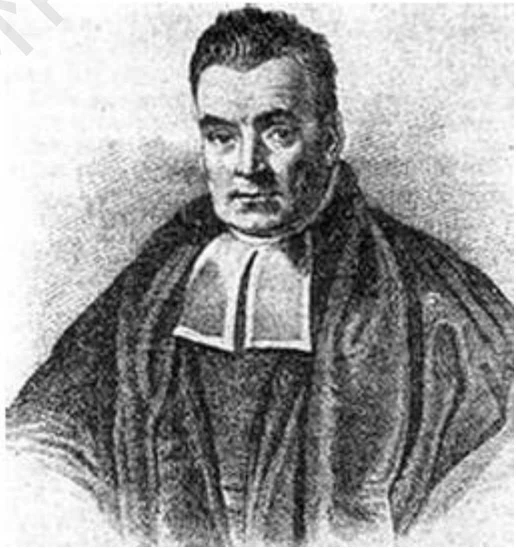
贝叶斯

贝叶斯简介：

- 贝叶斯(约1701-1761) Thomas Bayes，英国数学家
- 贝叶斯方法源于他生前为解决一个“逆概”问题写的一篇文章



贝叶斯

- 贝叶斯要解决的问题：

- **正向概率**：假设袋子里面有N个白球，M个黑球，你伸手进去摸一把，摸出黑球的概率是多大

- **逆向概率**：如果我们事先并不知道袋子里面黑白球的比例，而是闭着眼睛摸出一个（或好几个）球，观察这些取出来的球的颜色之后，那么我们可以就此对袋子里面的黑白球的比例作出什么样的推测

贝叶斯

- Why贝叶斯？

  - 现实世界本身就是不确定的，人类的观察能力是有局限性的

  - 我们日常所观察到的只是事物表面上的结果，因此我们需要提供一个猜测

贝叶斯

男生：60%
女生：40%

- 男生总是穿长裤，女生则一半穿长裤一半穿裙子

- **正向概率**：随机选取一个学生，他（她）穿长裤的概率和穿裙子的概率是多大

- **逆向概率**：迎面走来一个穿长裤的学生，你只看得见他（她）穿的是否长裤，而无法确定他（她）的性别，你能够推断出他（她）是女生的概率是多大吗？

贝叶斯

- 假设学校里面人的总数是 U 个
- 穿长裤的（男生）： U * P(Boy) * P(Pants|Boy)
  - P(Boy) 是男生的概率 = 60%
  - P(Pants|Boy) 是条件概率，即在 Boy 这个条件下穿长裤的概率是多大，这里是 100% ，因为所有男生都穿长裤
- 穿长裤的（女生）： U * P(Girl) * P(Pants|Girl)

贝叶斯

- **求解**：穿长裤的人里面有多少女生

- 穿长裤总数：U * P(Boy) * P(Pants|Boy) + U * P(Girl) * P(Pants|Girl)

- P(Girl|Pants) = U * P(Girl) * P(Pants|Girl)/穿长裤总数

  U * P(Girl) * P(Pants|Girl) / [U * P(Boy) * P(Pants|Boy) + U * P(Girl) * P(Pants|Girl)]

贝叶斯

与总人数有关吗？

U * P(Girl) * P(Pants|Girl) / [U * P(Boy) * P(Pants|Boy) + U * P(Girl) * P(Pants|Girl)

容易发现这里校园内人的总数是无关的，可以消去

P(Girl|Pants) = P(Girl) * P(Pants|Girl) / [P(Boy) * P(Pants|Boy) + P(Girl) * P(Pants|Girl)]

贝叶斯

- 化简：

  - P(Girl|Pants) = P(Girl) * P(Pants|Girl) / [P(Boy) * P(Pants|Boy) + P(Girl) * P(Pants|Girl)]

  - 分母其实就是 P(Pants)

  - 分子其实就是 P(Pants, Girl)

贝叶斯

✓ 贝叶斯公式

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

贝叶斯

- 拼写纠正实例：

  - 问题是我们看到用户输入了一个不在字典中的单词，我们需要去猜测："这个家伙到底真正想输入的单词是什么呢？

  - P(我们猜测他想输入的单词 | 他实际输入的单词)

贝叶斯

- 用户实际输入的单词记为 D （D 代表 Data， 即观测数据）

- 猜测1：P(h1 | D)，猜测2：P(h2 | D)，猜测3：P(h1 | D) 。。。
  统一为：P(h | D)

- P(h | D) = P(h) * P(D | h) / P(D)

贝叶斯

- 用户实际输入的单词记为 D （ D 代表 Data ， 即观测数据）

- 对于不同的具体猜测 h1 h2 h3 .. ， P(D) 都是一样的，所以在比较 P(h1 | D) 和 P(h2 | D) 的时候我们可以忽略这个常数

- $P(h \mid D) \propto P(h) * P(D \mid h)$
  对于给定观测数据，一个猜测是好是坏，取决于“这个猜测本身独立的可能性大小（先验概率，Prior ）”和“这个猜测生成我们观测到的数据的可能性大小。

贝叶斯

- 用户实际输入的单词记为 D （ D 代表 Data ，即观测数据）

- 对于不同的具体猜测 h1 h2 h3 .. ， P(D) 都是一样的，所以在比较 P(h1 | D) 和 P(h2 | D) 的时候我们可以忽略这个常数

- $P(h \mid D) \propto P(h) * P(D \mid h)$
  对于给定观测数据，一个猜测是好是坏，取决于“这个猜测本身独立的可能性大小（先验概率，Prior ）”和“这个猜测生成我们观测到的数据的可能性大小。

贝叶斯

- 拼写纠正实例：

  - 贝叶斯方法计算： P(h) * P(D | h)，P(h) 是特定猜测的先验概率

  - 比如用户输入tlp，那到底是 top 还是 tip？这个时候，当最大似然不能作出决定性的判断时，先验概率就可以插手进来给出指示——“既然你无法决定，那么我告诉你，一般来说 top 出现的程度要高许多，所以更可能他想打的是 top”

贝叶斯

模型比较理论

- 最大似然：最符合观测数据的（即 P(D | h) 最大的）最有优势

- 奥卡姆剃刀： P(h) 较大的模型有较大的优势

- 掷一个硬币，观察到的是“正”，根据最大似然估计的精神，我们应该猜测这枚硬币掷出“正”的概率是 1，因为这个才是能最大化 P(D | h) 的那个猜测

贝叶斯

模型比较理论

如果平面上有 N 个点，近似构成一条直线，但绝不精确地位于一条直线上。这时我们既可以用直线来拟合（模型1），也可以用二阶多项式（模型2）拟合，也可以用三阶多项式（模型3），特别地，用 N-1 阶多项式便能够保证肯定能完美通过 N 个数据点。那么，这些可能的模型之中到底哪个是最靠谱的呢？

奥卡姆剃刀：越是高阶的多项式越是不常见

贝叶斯

- 垃圾邮件过滤实例：

  - 问题：给定一封邮件，判定它是否属于垃圾邮件
    D 来表示这封邮件，注意 D 由 N 个单词组成。我们用 h+ 来表示垃圾邮件，h- 表示正常邮件

  - P(h+|D) = P(h+) * P(D|h+) / P(D)
    P(h- |D) = P(h-) * P(D|h-) / P(D)

贝叶斯

- 垃圾邮件过滤实例：

- **先验概率**：P(h+) 和 P(h-) 这两个先验概率都是很容易求出来的，只需要计算一个邮件库里面垃圾邮件和正常邮件的比例就行了。

- D 里面含有 N 个单词 d1, d2, d3， P(D|h+) = P(d1,d2,..,dn|h+)
P(d1,d2,..,dn|h+) 就是说在垃圾邮件当中出现跟我们目前这封邮件一模一样的一封邮件的概率是多大！

P(d1,d2,..,dn|h+) 扩展为： P(d1|h+) * P(d2|d1, h+) * P(d3|d2,d1, h+) * ..

贝叶斯

- 垃圾邮件过滤实例：

P(d1|h+) * P(d2|d1, h+) * P(d3|d2,d1, h+) * ..
假设 di 与 di-1 是完全条件无关的（朴素贝叶斯假设特征之间是独立，互不影响）
简化为 P(d1|h+) * P(d2|h+) * P(d3|h+) * ..

对于P(d1|h+) * P(d2|h+) * P(d3|h+) * ..只要统计 di 这个单词在垃圾邮件中出现的频率即可