离散概率（II）

离散数学课程组

南京大学计算机科学与技术系

内容提要

- 贝叶斯定理
- 广义贝叶斯定理
- 贝叶斯Spam过滤器

引言

- 疾病检测呈阳性，真的得病？概率多大?
- 疾病检测呈阴性，得病的概率多大?
- 已知部分证据（经验数据），估计特定事件的概率

Thomas Bayes (1701-1761)

贝叶斯定理

贝叶斯定理.

设E和F是样本空间S中的事件, $p(E)\neq 0$, $p(F)\neq 0$. 则

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

贝叶斯定理（举例）

- 有2个盒子，第一个盒子含2个绿球和7个红球。第二个盒子含4个绿球和3个红球。Bob随机选择盒子，并在选择的盒子里随机选取一个球。如果Bob选中的是红球，该球来自第一个盒子的概率？
- 解. 设E是Bob选中红球的事件， F是Bob选择第一个盒子的事件。

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

$$p(F|E) = \frac{(7/9)(1/2)}{(7/9)(1/2) + (3/7)(1/2)} = \frac{7/18}{38/63} = \frac{49}{76} \approx 0.645.$$

贝叶斯定理的推导

- 根据条件概率的定义

$$p(F \mid E) = p(E \cap F) / p(E)$$

- 另外，我们有：

$$p(E \cap F) = p(E \mid F)\, p(F)$$

- 所以，我们有：

$$p(F|E) = \frac{p(E|F)p(F)}{p(E)}$$

贝叶斯定理的推导

我们有

$$p(E) = p(E \cap F) + p(E \cap \overline{F}) = p(E|F)p(F) + p(E|\overline{F})p(\overline{F})$$

因此,

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

贝叶斯定理的应用

- 假设有一种罕见的疾病，100,000人只有1人会得这种病。如果某人得了此病，检测准确率高达99%；如果某人没有得此病，检测准确率为99.5%.
  - 疾病检测呈阳性，得此病的概率多大?
  - 疾病检测呈阴性，没有得此病的概率多大?
- 解. 设D是此人得此病的事件，E是疾病检测呈阳性的事件。需要计算 $p(D|E)$, $p(\overline{D}|\overline{E})$ 。

贝叶斯定理的应用

$p(D) = 1/100,000 = 0.00001 \qquad p(\overline{D}) = 1 - 0.00001 = 0.99999$

$p(E|D) = .99 \quad p(\overline{E}|D) = .01 \quad p(E|\overline{D}) = .005 \quad p(\overline{E}|\overline{D}) = .995$

$$p(D|E) = \frac{p(E|D)p(D)}{p(E|D)p(D) + p(E|\overline{D})p(\overline{D})}$$

$$= \frac{(0.99)(0.00001)}{(0.99)(0.00001) + (0.005)(0.99999)}$$

$$\approx 0.002$$

为何结果如此小？

呈阳性，也不必太担心！

贝叶斯定理的应用

$$p(\overline{D}|\overline{E}) = \frac{p(\overline{E}|\overline{D})p(\overline{D})}{p(\overline{E}|\overline{D})p(\overline{D}) + p(\overline{E}|D)p(D)}$$

$$= \frac{(0.995)(0.99999)}{(0.995)(0.99999) + (0.01)(0.00001)}$$

$$\approx 0.9999999$$

$$\boxed{\text{呈阴性，高枕无忧！}}$$

$$p(D|\overline{E}) \approx 1 - 0.9999999 = 0.0000001.$$

广义贝叶斯定理

假设E是取自样本空间S中的事件，$F_1, F_2, \ldots, F_n$ 是互斥的事件，且 $\bigcup_{i}^{n} F_i = S.$ 又假定$p(E)\neq 0$，$p(F_i)\neq 0$ $(i=1, \ldots, n)$. 则

$$p(F_j|E) = \frac{p(E|F_j)p(F_j)}{\sum_{i=1}^{n} p(E|F_i)p(F_i)}.$$

举例

- 朋友来看我，乘坐交通工具的概率和这些工具可能晚点的概率分别是
  - 乘坐概率：自驾(0.3), 公交 (0.1), 高铁(0.4), 飞机(0.2)
  - 晚点概率：自驾(0.3), 公交 (0.15), 高铁(0.05), 飞机(0.5)
  - 朋友迟到了, 何种原因最有可能导致这种现象？
- 解：A自驾，B公交，C高铁，D飞机，E迟到。
  - $p(A)=0.3$, $p(B)=0.1$, $p(C)=0.4$, $p(D)=0.2$;
  - $p(E|A)=0.3$, $p(E|B)=0.15$, $p(E|C)=0.05$, $p(E|D)=0.5$;
  - 求$p(A|E)$, $p(B|E)$, $p(C|E)$, $p(D|E)$中的最大者。

在众多线索中探究

$$p(A|E)=\frac{p(E|A)p(A)}{p(E|A)p(A)+p(E|B)p(B)+p(E|C)p(C)+p(E|D)p(D)}$$

p(A|E)=90/225=2/5, p(B|E)=15/225=1/15, p(C|E)=20/225=4/45,

p(D|E)=100/225=4/9

误事的很可能是飞机！

贝叶斯Spam过滤器

- 如何确定一个电子邮件是Spam?
  - 假设我们有一个垃圾邮件的集合B 和一个不是垃圾的邮件集合G。利用贝叶斯定理来预测一个新的电子邮件是Spam 的概率。
- 考察一个特定的单词w, 它在B 和G中出现的次数分别为$n_B(w)$和$n_G(w)$.
- 设S是邮件为Spam的事件, E是邮件内容含单词w的事件. 需要计算 $p(S|E)$,

$$\boxed{\text{需要估算} p(E \mid S) \text{和 } p(E \mid \overline{S})}$$

贝叶斯Spam过滤器

估算

$p(E \mid S) = \boldsymbol{p(w)} = \boldsymbol{n_B(w)/|B|}$

$p(E \mid \overline{S}) = \boldsymbol{q(w)} = \boldsymbol{n_G(w)/|G|}$

$$p(S|E) = \frac{p(E|S)p(S)}{p(E|S)p(S) + p(E|\overline{S})p(\overline{S})}$$

假设 $\boldsymbol{p(S)} = ½$

垃圾邮件的频率

$$p(S|E) = \frac{p(E|S)}{p(E|S) + p(E|\overline{S})}$$

$$r(w) = \frac{p(w)}{p(w) + q(w)}$$

若大于某个经验值，则被认为是Spam

贝叶斯Spam过滤器

举例: “Rolex” 在 2000 封垃圾邮件的250个当中出现，而在1000封非垃圾邮件中只有5封包含这个单词。估计一条含有“Rolex”的消息是Spam的概率. 假设收到的消息是Spam和不是Spam是等可能的。假设把一条消息作为Spam而拒绝的阈值为0.9，那么我们应该拒绝这条消息吗？

解: $p(Rolex)$ = 250/2000 =0.125, $q(Rolex)$ = 5/1000 = 0.005.

$$r(Rolex) = \frac{p(Rolex)}{p(Rolex) + q(Rolex)} = \frac{0.125}{0.125 + .005} = \frac{0.125}{0.125 + .005} \approx 0.962$$

将含有“Rolex”的消息分类为Spam，并拒绝这种消息。

使用多个单词的贝叶斯Spam过滤器

- 多个单词作为证据来改善精度.
- 设E_1和E_2分别为消息种包含单词w_1和w_2 的事件.
- 简化起见，假定事件是独立的，且$p(S) = ½$.

$$p(S|E_1 \cap E_2) = \frac{p(E_1|S)p(E_2|S)}{p(E_1|S)p(E_2|S) + p(E_1|\overline{S})p(E_2|\overline{S})}$$

$$r(w_1, w_2) = \frac{p(w_1)p(w_2)}{p(w_1)p(w_2) + q(w_1)q(w_2)}$$

使用多个单词的贝叶斯Spam过滤器

单词“stock”出现在2000条垃圾消息中的400条里，1000个非垃圾消息中的60条里，在200条垃圾消息和25条非垃圾消息中包含单词“undervalued”。对于一条既含有“stock”和“undervalued”的新消息，估计它是垃圾消息的概率。

解. $p(stock) = 400/2000 = 0.2$, $q(stock) = 60/1000=0.06$,
$p(undervalued)=200/2000 =0.1$, $q(undervalued)=25/1000=0.025$

$$r(stock, undervalued) = \frac{p(stock)p(undervalued)}{p(stock)p(undervalued) + q(stock)q(undervalued)}$$

$$= \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.06)(0.025)} \approx 0.930$$

假设阈值为0.9，那么我们应该拒绝它。

使用多个单词的贝叶斯Spam过滤器

- 考虑更多的单词，通常会得到更精准的垃圾过滤器。在独立性假设下，考虑*k*个单词:

$$p(S|\bigcap_{i=1}^{k} E_i) = \frac{\Pi_{i=1}^{k} p(E_i|S)}{\Pi_{i=1}^{k} p(E_1|S) + \Pi_{i=1}^{k} p(E_i|\overline{S})}$$

$$r(w_1, w_2, ...w_n) = \frac{\Pi_{i}^{k} p(w_i)}{\Pi_{i=1}^{k} p(w_i) + \Pi_{i=1}^{k} q(w_i)}$$

Spam战争！

作业

- 教材[6.3]
  - p. 325： 4, 6, 10, 16, 21