Ch 2-3 独立性

回顾前一次课

全概率公式：事件$A_1, A_2, \cdots, A_n$为样本空间Ω的一个分割, 对任意事件B有$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$

贝叶斯公式：设$A_1, A_2, \cdots, A_n$为样本空间Ω的一个划分, 且事件B满足$P(B) > 0$. 对任意$1 \le i \le n$有

$$P(A_i|B) = \frac{P(A_iB)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

独立性: $P(AB) = P(A)P(B)$

如何判断独立性？

1、直接计算判断$P(AB) \overset{?}{=} P(A)P(B)$

2、根据实际问题判断事件的独立性

- 两人独立射击打靶且互不影响, 因此两人中靶的事件相互独立
- 从n件产品中随机抽取两件, 事件A_i表示第i件是合格品. 若有放回抽取则事件A_1与A_2相互独立; 若不放回则不独立
- 机器学习的经典假设是训练数据独立同分布采样

例：从一副扑克 (不含大王、小王) 中随机抽取一张扑克, 用事件A表示抽到10, 事件B表示抽到黑色的扑克. 事件A与B是否独立?

条件独立性

设(Ω, Σ, P)是一个概率空间, 事件$C \in \Sigma$且有$P(C) > 0$, 若事件$A, B \in \Sigma$满足

$$P(AB|C) = P(A|C)P(B|C)$$

或等价条件

$$P(A|BC) = P(A|C)$$

称事件A和B在事件C发生的情况下是**条件独立的** (conditional independent)

例题

设一个箱子中有$k+1$枚不均匀的硬币，投掷第i枚硬币时正面向上的概率为i/k $(i=0,1,2,\cdots,k)$. 现从箱子中任意取出一枚硬币，并任意重复投掷多次, 若前n次正面向上, 求第$n+1$次正面向上的概率

多事件的独立性

定义 若事件A, B, C满足

- $P(AB) = P(A)P(B), P(AC) = P(A)P(C), P(BC) = P(B)P(C)$
- $P(ABC) = P(A)P(B)P(C)$

则称 **事件$\boldsymbol{A, B, C}$相互独立**

事件A, B, C相互独立	事件A, B, C两两独立

Bernstein 反例

[Bernstein 反例] 一个均匀的正四面体, 第一面红色, 第二面白色, 第三面黑色, 第四面同时有红、白、黑三种颜色. 随意投掷一次, 用A, B, C分别表示红色、白色、黑色朝下的事件. 考虑这三事件的相互独立性与两两独立性

多事件的独立性

定义 若事件$A_1, A_2, \cdots, A_n$中任意k个事件独立, 即对任意$k \in [n]$

$$P\left(A_{i_1} \cdots A_{i_k}\right) = P\left(A_{i_1}\right) \cdots P\left(A_{i_k}\right)$$

其中$1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq \mathrm{n}$, 则称 **事件$A_1, A_2, \cdots, A_n$相互独立**

注意：n个事件的相互独立性共有$2^n - n - 1$个等式

事件$A_1, A_2, \cdots, A_n$的相互独立性与两两独立性的区别

可以类似定义多个事件的条件独立性

例题

三人独立破译一份密码, 每人单独能破译的概率分别为1/5, 1/3, 1/4, 问三人中至少有一人能破译密码的概率.

小概率原理

若n个事件$A_1,\cdots,A_n$相互独立, 其发生的概率分别为$p_1,\cdots,p_n$

事件$A_1,A_2,\cdots,A_n$中至少有一事件发生的概率为

$$P(A_1\cup\cdots\cup A_n)=1-P\left(\bar{A}_1\cdots\bar{A}_n\right)=1-(1-p_1)\cdots(1-p_n)$$

事件$A_1,A_2,\cdots,A_n$中至少有一事件不发生的概率为

$$P\left(\bar{A}_1\cup\bar{A}_2\cdots\cup\bar{A}_n\right)=1-P(A_1A_2\cdots A_n)\ =\ 1-p_1p_2\cdots p_n$$

小概率原理

若每个事件的概率p_i都非常小, 但n非常大, 则n个相互独立的事件中 至少有一事件发生 或 至少有一事件不发生 的概率都很大

若事件A在一次试验中发生的概率非常小, 但经过多次独立地重复试验, 事件A的发生是必然的, 称之为 小概率原理

例题

冷战时期美国的导弹精度90%, 苏联的导弹精度70%, 但苏联的导弹数量特别多, 导弹的数量能否弥补精度的不足?

例题

假设市场上有m种不同类型的邮票，一位集邮爱好者收集第i种邮票的概率为p_i，且$p_1 + p_2 + \cdots + p_m = 1$. 假设每次集邮都是独立同分布的，若现已收集到$n$张邮票，用$A_i$表示至少收集到第$i$种类型邮票的事件，求$P(A_i)$, $P(A_i \cup A_j)$ 以及$P(A_i | A_j)$ $\quad (i \neq j)$

Ch 2-4 案例分析

案例分析：两多项式相等

给定两个较复杂的多项式

$F(x) = (x+2)^7(x+3)^5 + (x+1)^{100} + (x+2)(x+3) + x^{20}$

$G(x) = (x+3)^{100} - (x+1)^{25}(x+2)^{30} + (x-2)(x-3)\cdots(x-100)$

如何快速验证$F(x) \equiv G(x)$?

案例分析二：矩阵乘法相等

给定矩阵$A, B, C \in \{0,1\}^{n\times n}$ $(n \geq 10000000)$, 验证$AB = C$?

独立随机产生一个向量$r \in \{0,1\}^n$, 判断

$$A(Br) = Cr?$$

计算$A(Br)$ 和Cr的复杂度均为$O(n^2)$. 若$A(Br) \neq Cr$则直接有$AB \neq C$; 若$A(Br) = Cr$并不能得出$AB = C$.

将上述过程独立进行k次, 可以证明以较大的概率有$AB = C$成立, 该过程被称为Freivalds算法

Freivalds算法

输入: 矩阵 $\mathbf{A}, \mathbf{B}, \mathbf{C}$

输出: 是/否　　　　　　　　　　%% 验证 $\mathbf{AB} \stackrel{?}{=} \mathbf{C}$

- -

For $i = 1 : k$

　　随机选择向量 $\bar{\mathbf{r}}_i = (r_{i1}, r_{i2}, \cdots, r_{in})$, 其每个元素是从 $\{0, 1\}$ 独立等可能随机采样所得

　　计算向量 $\bar{\mathbf{p}}_i = \mathbf{A}(\mathbf{B})\bar{\mathbf{r}}_i - \mathbf{C}\bar{\mathbf{r}}_i$

　　If $\{\bar{\mathbf{p}}_i$ 不是零向量$\}$ then

　　　　返回“否”

　　EndIf

EndFor

返回“是”.

算法的有效性

算法返回 否, 必有$AB \neq C$, 因为找到一个向量$\bar{r}$使得$AB\bar{r} \neq C\bar{r}$;

算法返回 是, 则不一定有$AB = C$, 但可以以较大的概率保证

定理：若$AB \neq C$, 且随机向量$\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_k \in \{0,1\}^n$中每个元素是从$\{0,1\}$独立等可能随机采样所得, 则有

$$P\left[\bigcap_{i=1}^{k} AB\bar{r}_i = C\bar{r}_i\right] \leq \frac{1}{2^k}$$

隐私问题的调查

每个人都有一些隐私或秘密, 相关信息不希望被外人知晓

对于具有社会普遍性的隐私问题，需要对相关问题进行一些必要的调查. 例如当代大学生中有抑郁倾向的同学占有多大的比例, 家庭不和谐的同学所占多少比例

设计一种调查方案，使被调查者既愿意作出真实回答、又较好地保护个人隐私，最后利用全概率公式和随机事件的独立性来完成最后的信息统计

完全图着色

设平面上有n个顶点，其中任意三个顶点不在同一条直线上，用$n(n-1)/2$条边将这些顶点连接起来的图称为n个顶点的**完全图**

例如三个顶点的完全图是一个三角形.

现将图中的每条边都分别染成红色和和蓝色，讨论的问题：当$n \geq 10$时，给定一个正整数$k > n/2$，是否存在一种染色方法，使得该图上任意k个顶点, 其相应的$k(k-1)/2$条边不是同一颜色