

Machine Learning

Mathematical basis

Wang Jiewen

Today's Topics

- Probability
- Extremum
- Vector and Matrix

Today's Topics

- *Probability*
- Extremum
- Vector and Matrix

随机事件和随机试验

- 随机试验E
 - 试验可在相同条件下重复进行
 - 试验的可能结果不止一个且所有可能结果都已知
 - 每次试验哪个结果出现是未知的
- 例子
 - E1: 抛1枚匀称的硬币 一枚匀称的硬币，观察正反面出现的情况
 - E2: 投掷一颗匀称的骰子，观察出现的点数

随机事件和随机试验

- 样本空间：随机试验 E 的所有可能结果组成的集合称为试验 E 的样本空间，记为 Ω 。
 - 在 E_1 (抛硬币)中， $\Omega=\{\text{正面}, \text{反面}\}$
 - 在 E_2 (掷骰子)中， $\Omega=\{1, 2, 3, 4, 5, 6\}$

随机事件和随机试验

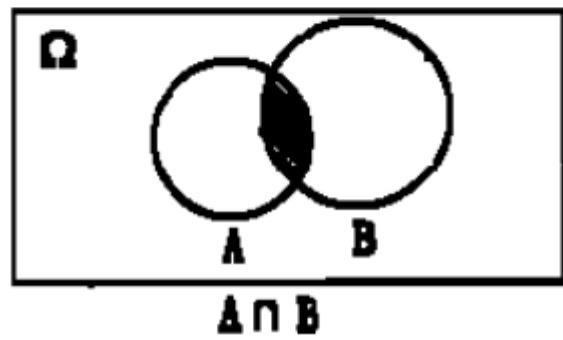
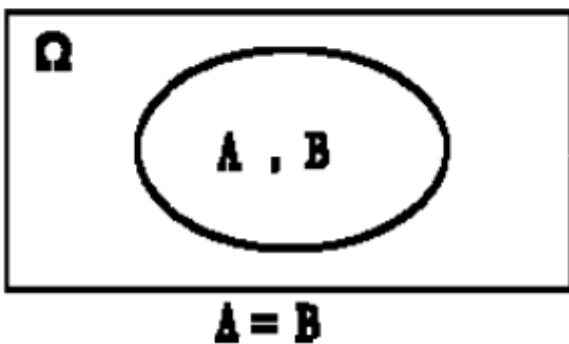
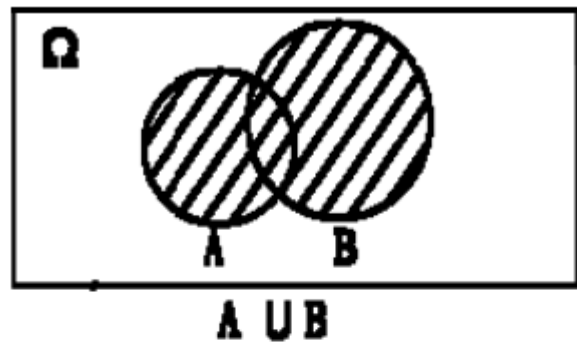
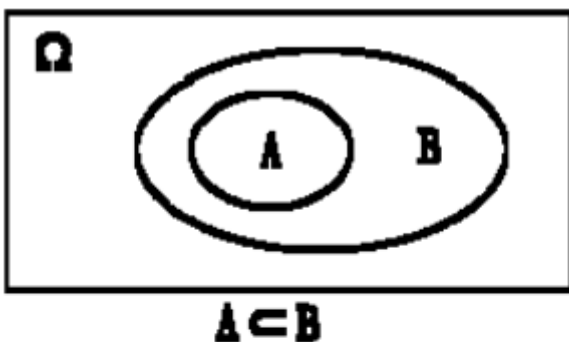
- 随机事件：在随机试验中可能出现也可能不出现的试验结果
- 随机事件常记为 A , B , C 例如：
 - 在 E_1 (抛硬币)中, A 表示“出现正面”, B 表示“出现反面”
 - 在 E_2 (掷骰子)中, A 表示“掷出2点”, B 表示“掷出偶数点”

随机事件和随机试验

- 必然事件：每次试验必然发生的事件称为必然事件
- 不可能事件：每次试验都不可能发生的事件称为不可能事件
- $P=1$ 是事件必然发生的必要而非充分条件
- $P=0$ 是事件不可能发生的必要而非充分条件

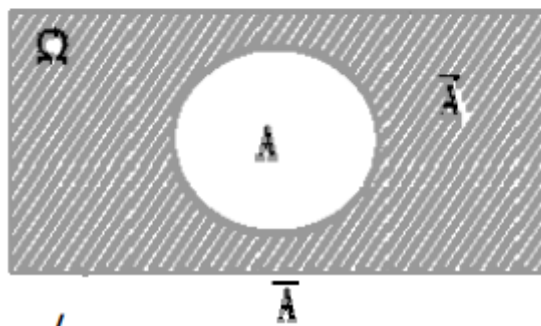
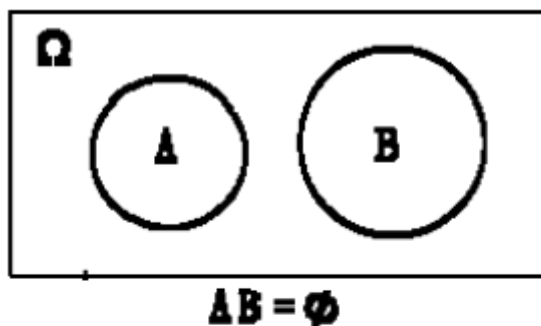
随机事件间的关系

- 包含、相等、和、积



随机事件间的关系

- 互斥、对立



概率的统计定义

- 频率

- 在 n 次重复试验中, 设事件 A 出现了 n_A 次, 则称比值 $\frac{n_A}{n}$ 为 n 次试验中事件 A 发生的频率, 记作 $f_n(A)$

试验者	抛硬币次数 n	出现正面 (A) 的次数 n_A	频率
德·摩尔根	2048	1061	0 . 5180
浦丰	4040	2148	0 . 5069
皮尔逊	12000	6019	0 . 5016
皮尔逊	24000	12012	0 . 5005
维尼	30000	14994	0 . 4998

- 概率的统计定义

- 在相同条件下, 将试验重复 n 次, 如果随着重复试验次数 n 的增大, 事件 A 的频率 $f_n(A)$ 越来越稳定地在某一常数 p 附近摆动, 则称常数 $p(0 \leq p \leq 1)$ 为事件 A 的概率, 即 $P(A) = p$ 。

概率的公理化定义

- 设某试验的样本空间为 Ω ，对其中每个事件 A 定义一个实数 $P(A)$ ，如果它满足下列三条公理，称 $P(A)$ 为 A 的概率
 - $0 \leq P(A) \leq 1$
 - $P(\Omega) = 1$
 - 若事件 A_1, A_2, \dots, A_n 互不相容，则

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

引申：距离也有类似的公理化定义。

概率的性质

- 如果 $A \subseteq B$, 则 $P(A) < P(B)$
- 如果 $A \subseteq B$, 则 $P(B - A) = P(B) - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(AB)$

条件概率

- 在已知事件 B 发生的条件下，事件 A 发生的概率称为事件 A 的条件概率，记为 $P(A|B)$

条件概率

- 例：将一枚硬币抛掷两次，观察其出现正反面的情况。
- 设事件A表示“至少有一次出现正面”
- 事件B表示“两次都出现同一面”
- 求
 - (1)事件B的概率
 - (2)已知A发生的条件下，事件B的概率

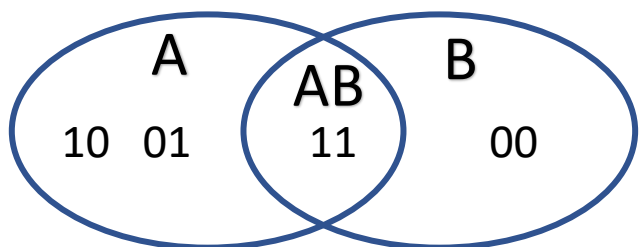
设1为正，0为反。样本空间 $S = \{11, 10, 01, 00\}$

事件 $A = \{11, 10, 01\}$ ，事件 $B = \{11, 00\}$

- (1) $P(B) = \frac{N(B)}{N(S)} = \frac{2}{4} = \frac{1}{2}$
- (2) $P(B|A) = \frac{1}{3}$ 样本空间缩小为 $\{11, 10, 01\}$ ，只有11代表事件B出现

条件概率

- 例：将一枚硬币抛掷两次，观察其出现正反面的情况。
- 设事件A表示“至少有一次出现正面”
- 事件B表示“两次都出现同一面”



设1为正，0为反。样本空间 $S = \{11, 10, 01, 00\}$
事件 $A = \{11, 10, 01\}$ ，事件 $B = \{11, 00\}$

- $P(B|A) = \frac{N(AB)}{N(A)} = \frac{1}{3}$
- 可理解为：B的样本点在A中占的比例
- $P(B|A) = \frac{N(AB)}{N(A)} = \frac{\frac{N(AB)}{N(S)}}{\frac{N(A)}{N(S)}} = \frac{P(AB)}{P(A)}$ ，由此得到了条件概率的定义

条件概率

- 定义：设A，B是两个事件，且 $P(A) > 0$ ，称 $\frac{P(AB)}{P(A)}$ 为在事件A发生的条件下事件B发生的条件概率，记作 $P(B|A)$ ，即

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

条件概率

- 乘法公式：根据条件概率定义可得

$$P(AB) = P(A)P(B|A) \quad P(A) > 0$$

- 同理，若 $P(B) > 0$ ，有

$$P(AB) = P(B)P(A|B) \quad P(B) > 0$$

- 乘法公式可以推广到多个事件

$$P(ABC) = P(AB)P(C|AB) = P(A)P(B|A)P(C|AB) \quad P(AB) > 0$$

$$P(A_1A_2A_3 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \dots P(A_n|A_1A_2A_3 \dots A_{n-1}) \\ P(A_1A_2A_3 \dots A_{n-1}) > 0$$

- 注意：A和B独立时才有 $P(AB) = P(A)P(B)$

全概率公式

- 根据设试验E的样本空间为S, (B_1, B_2, \dots, B_n) 为S的一个划分, 且 $P(B_i) > 0$ ($i = 1, 2, \dots, n$), A为E的一个事件, 则

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \dots + P(B_n)P(A|B_n) = \sum_{i=1}^n P(B_i)P(A|B_i)$$

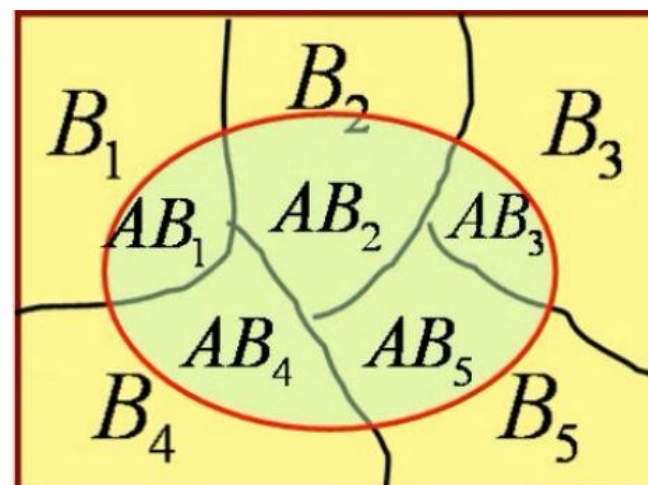
- 证明:



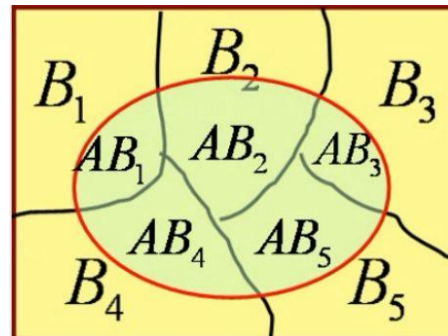
乘法公式

$$P(A) = P(AB_1) + P(AB_2) + \dots + P(AB_n)$$

- 全概率公式的意义: **由原因推结果**



全概率公式



- **例：**一电器商店出售两家工厂生产的电视机，甲厂的电视机占70%，乙厂占30%。甲厂的电视机合格率为95%，乙厂的合格率为80%，求该商店所售电视机的合格率。
- 设A=合格电视机
- 对全部电视机进行划分，得到：B=甲厂电视机，C=乙厂电视机
- 得到以下概率：

$$P(B) = 0.7, P(A|B) = 0.95$$

$$P(C) = 0.3, P(A|C) = 0.8$$

- 则根据全概率公式：

$$P(A) = P(B)P(A|B) + P(C)P(A|C) = 0.7 \times 0.95 + 0.3 \times 0.8 = 0.905$$

贝叶斯公式

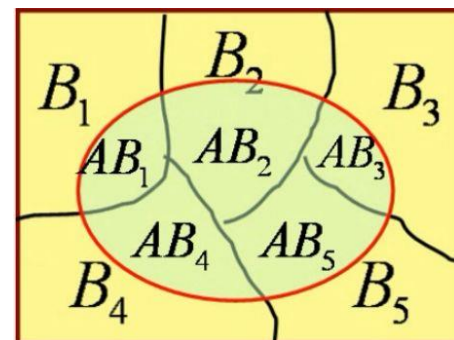
- 乘法公式: $P(AB) = P(A)P(B|A) \quad P(A) > 0$
 $P(AB) = P(B)P(A|B) \quad P(B) > 0$



- 贝叶斯公式: $P(B|A) = \frac{P(B)P(A|B)}{P(A)}$
- 进一步: (B_1, B_2, \dots, B_n) 为样本空间的一个划分, 则有

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}$$

$(i = 1, 2, \dots, n)$



- 贝叶斯公式的意义: **已知结果, 寻找原因**

贝叶斯公式

- **例：**对以往数据分析表明，机器调整得良好时，产品的合格率为95%，而当机器发生某种故障时，其合格率为50%。设机器调整良好的概率为90%，已知某日生产的第一件产品是合格品，求机器调整良好的概率。
- 用A表示产品合格，B表示机器调整良好。则：

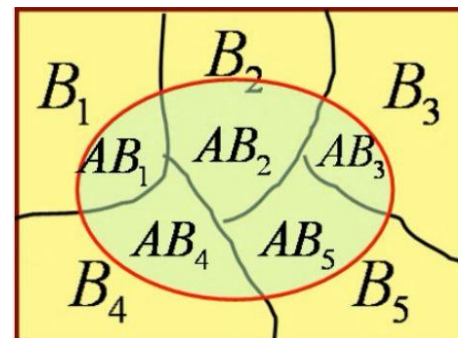
$$P(B) = 0.9, P(\bar{B}) = 1 - P(B) = 0.1$$

$$P(A|B) = 0.95, P(A|\bar{B}) = 0.5$$

- 由贝叶斯公式：

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

$$\begin{aligned} &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} \\ &= 0.945 \end{aligned}$$



先验概率与后验概率

- 对以往数据分析表明，机器调整得良好时，产品的合格率为95%，而当机器发生某种故障时，其合格率为50%。设机器调整良好的概率为90%，已知某日生产的第一件产品是合格品，求机器调整良好的概率。
- 用A表示产品合格，B表示机器调整良好。则：

$$P(B) = 0.9 \quad P(B|A) = 0.945$$

- 机器调整良好的概率 $P(B) = 0.9$ 是根据以往的数据分析所得，称为先验概率
- 条件概率是在得到产品合格的信息之后再重新加以修正的概率，称为后验概率

事件的独立性

- 对任意两个事件A、B，如果 $P(AB) = P(A)P(B)$ ，则称事件A与事件B相互独立。
- 当 $P(B) > 0$ 时，若A、B相互独立，则 $P(A|B) = P(A)$
- 若事件A与事件B相互独立，则A与 \bar{B} ，B与 \bar{A} ， \bar{A} 与 \bar{B} 也相互独立

事件的独立性

- 若事件 A_1, A_2, \dots, A_n 满足条件:

$$P(A_i A_j) = P(A_i)P(A_j) \quad 1 \leq i < j \leq n$$

则称 n 个事件 A_1, A_2, \dots, A_n 是两两独立的。

- 若对任意整数 $k (2 \leq k \leq n)$ 和 $1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$, 恒有:

$$P(A_{i_1}, A_{i_2}, \dots, A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

则称这 n 个事件相互独立。

- 相互独立 \Rightarrow 两两独立, 两两独立 \nRightarrow 相互独立

随机变量

- 设试验的样本空间为 Ω ，在 Ω 上定义一个单值实函数
$$X = X(e), e \in \Omega$$
- 对试验的每个结果 e ， $X=X(e)$ 有确定的值与之对应，称此定义在样本空间 Ω 上的单值实函数 $X=X(e)$ 为一个随机变量
- 随机变量常用字母 X, Y, Z, X_1, X_2, \dots 来表示。

随机变量

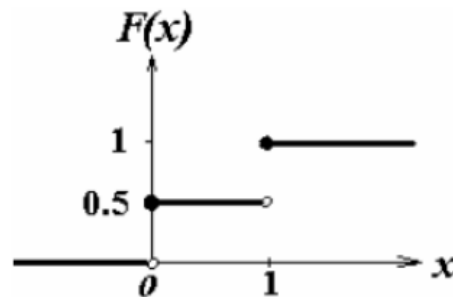
事件→实数值

- 例：在抛硬币试验中，样本空间={正面，反面}。设试验结果为随机变量 X ，出现正面用1表示，出现反面用0表示，则有：

$$X = \begin{cases} 0, & \text{结果为反面} \\ 1, & \text{结果为正面} \end{cases}$$

- 其分布函数为：

$$F(x) = P\{X \leq x\} \Rightarrow F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$



离散型随机变量

- 随机变量 X 的所有可能取值为有限个或可列无限个值。
- X 所有可能取值的概率:

$$p_k = P\{X = x_k\}, \quad k = 1, 2, 3, \dots$$

- 概率分布使用分布列描述

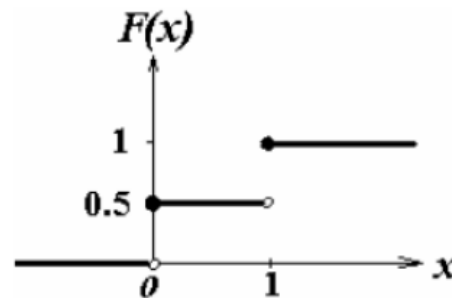
X	x_1	x_2	x_n
p	p_1	p_2	p_n

扔硬币问题

X	0	1
p	$\frac{1}{2}$	$\frac{1}{2}$

- 分布函数呈阶梯形

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$



常用离散型分布

- 0-1分布

$$P\{X = 1\} = p, P\{X = 0\} = 1 - p$$

- 二项分布 $X \sim B(n, p)$

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k q^{n-k}, k = 0, 1, 2, \dots, n$$

- 泊松分布

$$P\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, 2, \dots, \lambda > 0$$

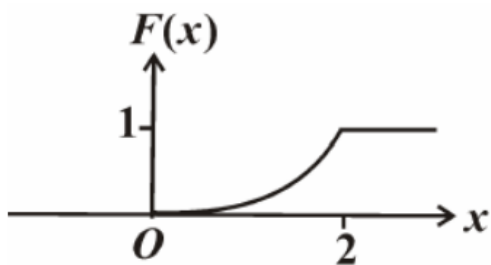
连续型随机变量

- 设随机变量 X 的分布函数为 $F(x)$ ，若存在非负函数 $f(x)$ 使得对任意实数 x ，有

$$F(x) = \int_{-\infty}^x f(t) dt$$

- 则称 X 为连续型随机变量，函数 $f(x)$ 为随机变量 X 的概率密度函数。
- 随机变量 X 落在任一区间上的概率等于它的概率密度在该区间上的积分

$$P\{X \in I\} = \int_a^b f(x) dx \quad I = (a, b]$$



常用连续型分布

- 均匀分布

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{其它} \end{cases}$$

- 指数分布

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, \text{其它} \end{cases}$$

- 高斯分布 $\mathbf{X} \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad -\infty < x < \infty$$

随机变量的数字特征

- 离散型随机变量的数学期望
- 设 X 为离散型随机变量，其分布律为

$$P\{X = x_i\} = p_i, \quad i = 1, 2, \dots$$

- 如果级数 $\sum_{i=1}^{+\infty} x_i p_i$ 绝对收敛，则称它为随机变量 X 的数学期望，记作 $E(X)$ ，即

$$E(X) = \sum_{i=1}^{+\infty} x_i p_i$$

随机变量的数字特征

- 连续型随机变量的数学期望
- 设 X 为连续型随机变量，其概率密度函数为 $f(x)$ ，如果积分

$$\int_{-\infty}^{+\infty} xf(x)dx$$

绝对收敛，则称它为随机变量 X 的数学期望，记作 $E(X)$ ，即

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

随机变量的数字特征

- 随机变量函数 $Y = g(X)$ 的数学期望
- 离散型:

$$E(Y) = E[g(x)] = \sum_{i=1}^{+\infty} g(x_i) p_i$$

- 连续型:

$$E(Y) = E[g(x)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

随机变量的数字特征

- 数学期望的性质

- $E(c) = c$

- 线性组合:

$$E(c_1X_1 + c_2X_2 + \cdots + c_nX_n) = c_1E(X_1) + c_2E(X_2) + \cdots + c_nE(X_n)$$

- 随机变量间相互独立时:

$$E(X_1X_2 \cdots X_n) = E(X_1)E(X_2)E(X_n)$$

随机变量的数字特征

- 例1

X	1	2	3	4
p	1/8	1/4	1/2	1/8

- 求 $E(x + 2)^2$
- 177/8

随机变量的数字特征

- 例2
- 求 $[a,b]$ 上的均匀分布的数学期望
- $(a+b)/2$

随机变量的数字特征

- 方差
- 离散型:

$$D(X) = E[X - E(X)]^2 = \sum_i [x_i - E(X)]^2 p_i$$

- 连续型:

$$D(X) = E[X - E(X)]^2 = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) dx$$

- 性质
 - $D(c) = 0$
 - $D(cX) = c^2 D(X)$
 - $D(X) = E(x^2) - [E(x)]^2$

练习题

- 一批产品由4个工厂生产，分别生产了3000、2000、2500和2500件，其次品率分别为5%、8%、15% 和10%。
 - (1) 求这批产品的次品率。0.0935（全概率公式）
 - (2) 任取一件产品，发现是次品，求该产品出自工厂1的概率。
0.16（贝叶斯公式）

练习题

- 根据以往的记录，某种诊断肝炎的试验有如下效果：对肝炎病人的试验呈阳性的概率为 0.95；非肝炎病人的试验呈阴性的概率为 0.95. 对自然人群进行普查的结果为：有千分之五的人患有肝炎。现有某人做此试验结果为阳性，问此人确有肝炎的概率为多少？ 0.087（贝叶斯公式）

练习题

- 设随机变量 X 服从 $[0, \pi]$ 的均匀分布, 求: $\frac{2}{\pi}, \frac{\pi^2}{3}, \frac{\pi^2}{12}$

$$E(\sin X), E(X^2), E(X - E(X))^2$$

- 求 $[a, b]$ 上的均匀分布的方差 $\frac{(b-a)^2}{12}$

Today's Topics

- Probability
- *Extremum*
- Vector and Matrix

偏导数

- 求函数 $f(x, y)$ 的偏导数
 - 求对 x 的偏导数，固定 y
 - 求对 y 的偏导数，固定 x
- 例
 - 1. 求 $z = x^2 + 3xy + y^2$ 在 $(1, 2)$ 处的偏导数 8
 - 2. 设 $z = x^3y^2 - 3xy^3 - xy + 1$ ，求 $\frac{\partial^2 z}{\partial x \partial y}$ 和 $\frac{\partial^2 z}{\partial y \partial x}$
均为 $6x^2y - 9y^2 - 1$

二元函数的极值

- 第一步：令一阶偏导数为0，求得驻点
- 第二步：计算驻点处的二阶偏导数，按照以下法则判断：

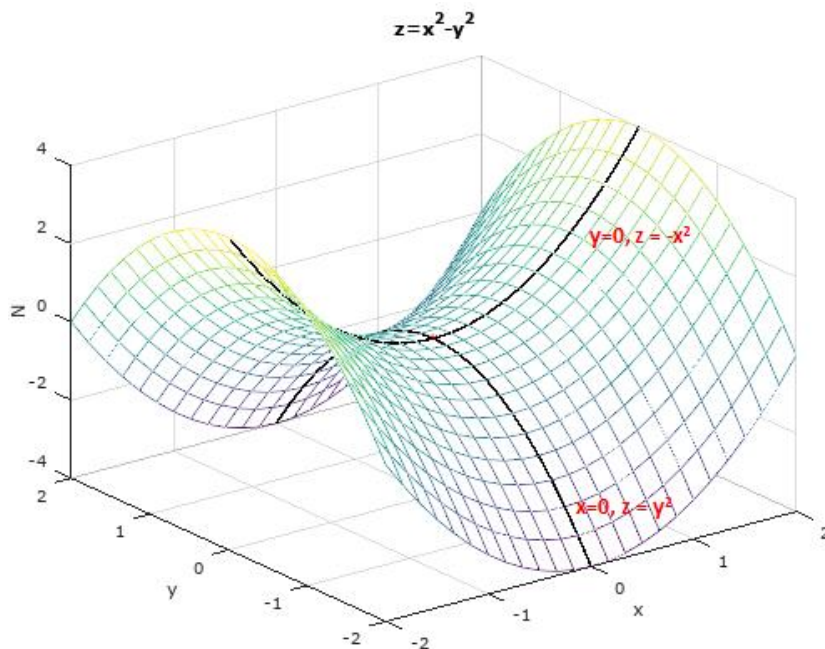
$$A = f_{xx}(x_0, y_0), \quad B = f_{xy}(x_0, y_0), \quad C = f_{yy}(x_0, y_0)$$

1) $AC - B^2 > 0$ && $A > 0 \Rightarrow$ 极小值

2) $AC - B^2 > 0$ && $A < 0 \Rightarrow$ 极大值

3) $AC - B^2 < 0 \Rightarrow$ 不是极值（鞍点）

4) $AC - B^2 = 0 \Rightarrow$ 不确定



二元函数的极值

- 例：求函数 $f(x, y) = x^3 - 3xy + y^3$ 的极值
- 驻点是 $(0, 0) / (1, 1)$ ，前者为鞍点，后者为极小值点

Hessian矩阵

- 二元情况下: $H(x_0, y_0) = \begin{bmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix}$
- 多元情况下:

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- 多元函数极值的判定条件与Hessian矩阵的正定性有关

拉格朗日乘子法求条件极值

目标: $z=f(x, y)$ 条件: $\varphi(x, y)=0$

1° $L=f(x, y)+\lambda\varphi(x, y)$

2° 令
$$\begin{cases} L_x = f_x + \lambda\varphi_x = 0 \\ L_y = f_y + \lambda\varphi_y = 0 \\ L_\lambda = \varphi(x, y) = 0 \end{cases} \Rightarrow \begin{cases} x = ? \\ y = ? \end{cases}$$

拉格朗日乘子法求条件极值

- 例：求表面积为 a^2 的体积最大长方体

[解] 1° 设长、宽、高为 x 、 y 、 z

$$2xy + 2xz + 2yz = a^2 \quad V = xyz$$

即目标： $V = xyz$ ； 条件： $2xy + 2xz + 2yz - a^2 = 0$

$$2^\circ L = xyz + \lambda(2xy + 2xz + 2yz - a^2)$$

$$\text{令} \begin{cases} L_x = yz + 2\lambda(y + z) = 0 \\ L_y = xz + 2\lambda(x + z) = 0 \\ L_z = xy + 2\lambda(x + y) = 0 \\ L_\lambda = 2xy + 2xz + 2yz - a^2 = 0 \end{cases}$$

$$\text{解得 } x = y = z = \frac{a}{\sqrt{6}} \quad \therefore V_{\max} = \frac{a^3}{6\sqrt{6}}$$

https://blog.csdn.net/q_42294351

练习题

- 在椭圆 $x^2 + 4y^2 = 4$ 上求一点，使其到直线 $2x + 3y - 6 = 0$ 的距离最短。
- 极值点为 $(8/5, 3/5)$ ，最短距离为 $\frac{\sqrt{13}}{13}$
- 注意排除了一个不符合条件的点

Today's Topics

- Probability
- Extremum
- *Vector and Matrix*

向量

$$\begin{bmatrix} 1 \\ 0 \\ 5 \\ 6 \\ 2 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 & 1 \\ 0 & 7 & 2 \\ 3 & 5 & 0 \end{bmatrix}$$

Example: Hyperplane 超平面

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad \longrightarrow \quad (w_0, w_1, \dots, w_n) \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} + b = 0$$

向量范数

- **L1/L2 Regularization, and other mathematical expression**

$$x = [x_0, x_1, \dots, x_m]^T$$

$$\text{1-Norm: } \|x\|_1 = \sum_{i=1}^m |x_i|$$

$$\text{2-Norm: } \|x\|_2 = \sqrt{\sum_{i=1}^m x_i^2}$$

$$\text{p-Norm: } \|x\|_p = (\sum_{i=1}^m |x_i|^p)^{\frac{1}{p}}$$

$$\infty\text{-Norm: } \|x\|_\infty = \max_i |x_i|$$

矩阵转置

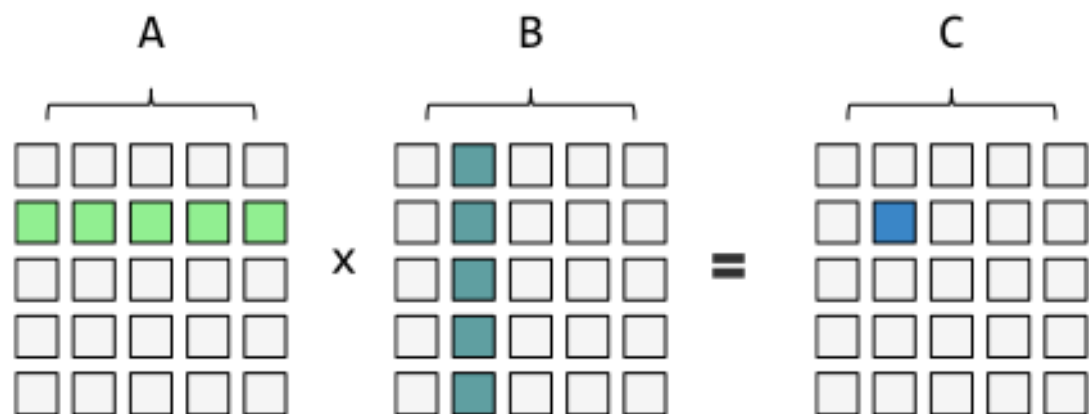
$$\mathbf{A}_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \longrightarrow \mathbf{A}^T_{n \times m} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & & & \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \longrightarrow \mathbf{A}^T = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \longrightarrow \mathbf{B}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

矩阵乘法

- Neural network, convolution ...



$$C[i][j] = \text{sum}(A[i][k] * B[k][j]) \text{ for } k = 0 \dots n$$

A矩阵的列数=B矩阵的行数

矩阵乘法

- **Example**

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 9 & 14 \end{bmatrix}$$

$$11 = 1 \times 1 + 3 \times 0 + 2 \times 5$$

$$10 = 1 \times 3 + 3 \times 1 + 2 \times 2$$

$$9 = 4 \times 1 + 0 \times 0 + 1 \times 5$$

$$14 = 4 \times 3 + 0 \times 1 + 1 \times 2$$

特征值与特征向量

$$\lambda x = Ax$$

- 求特征值: $|A - \lambda I| = 0$
- 例: 求 $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ 的特征值和特征向量
- 2, 4

矩阵的迹

- 矩阵的迹 $tr(A)$ 是矩阵的对角线元素之和
- $tr(A)$ = 特征值之和
- $tr(ABCD) = tr(BCDA) = tr(CDAB) = tr(DABC)$