北风网
IBEIFENG.COM

人工智能之机器学习

数学基础、PythonApi回顾

上海育创网络科技股份有限公司

主讲人：赵翌臣

数学知识回顾

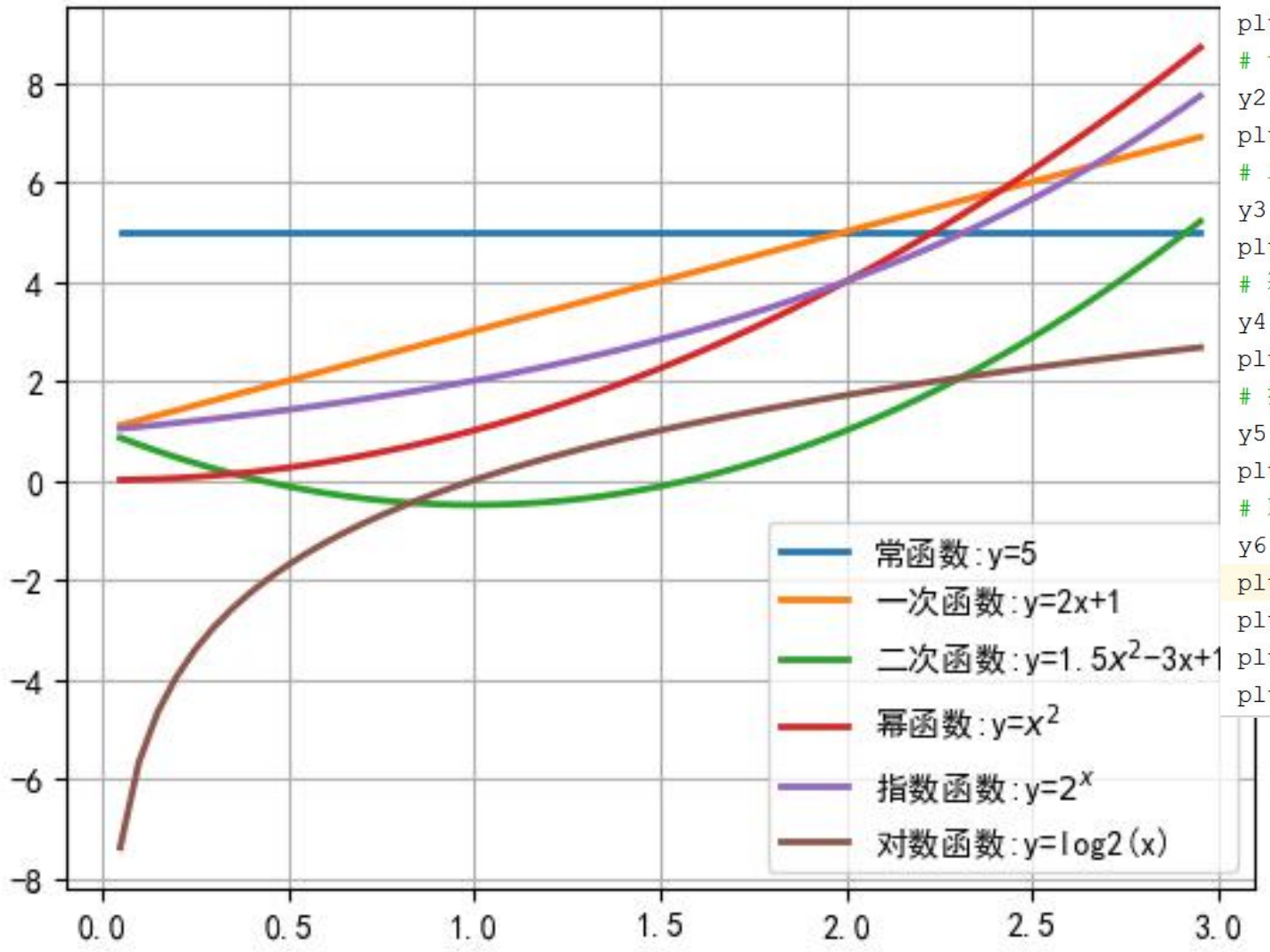
- 常见函数
- **导数、梯度《求导的方式、导数/梯度的含义/作用》**
- Taylor公式（SVM高斯核函数、XGboost）
- **联合概率、条件概率、全概率公式、贝叶斯公式**
- 期望、方差、协方差《了解这三个东西表示数据具有什么样的特性》
- 大数定律、中心极限定理
- **最大似然估计(MLE)《最大似然估计必须掌握》**
- **向量、矩阵的运算**

微积分

常见函数

- 常函数: $y = C$　　一次函数: $y = ax + b$
- 二次函数: $y = ax^2 + bx + c$　幂函数: $y = x^a$
- 指数函数: $y = a^x$, a的取值范围为: a>0&a≠1
- 对数函数: $y = \log_a(x)$, a的取值范围为: a>0&a≠1

常见函数




```
x = np.arange(0.05, 3, 0.05)
# 常函数
y1 = [5 for i in x]
plt.plot(x, y1, linewidth=2, label='常函数:y=5')
# 一次函数
y2 = [2 * i + 1 for i in x]
plt.plot(x, y2, linewidth=2, label='一次函数:y=2x+1')
# 二次函数
y3 = [1.5 * i * i - 3 * i + 1 for i in x]
plt.plot(x, y3, linewidth=2, label='二次函数:y=1.5$x^2$-3x+1')
# 幂函数
y4 = [math.pow(i, 2) for i in x]
plt.plot(x, y4, linewidth=2, label='幂函数:y=$x^2$')
# 指数函数
y5 = [math.pow(2, i) for i in x]
plt.plot(x, y5, linewidth=2, label='指数函数:y=$2^x$')
# 对数函数
y6 = [math.log(i, 1.5) for i in x]
plt.plot(x, y6, linewidth=2, label='对数函数:y=log2(x)')
plt.legend(loc='lower right')
plt.grid(True)
plt.show()
```

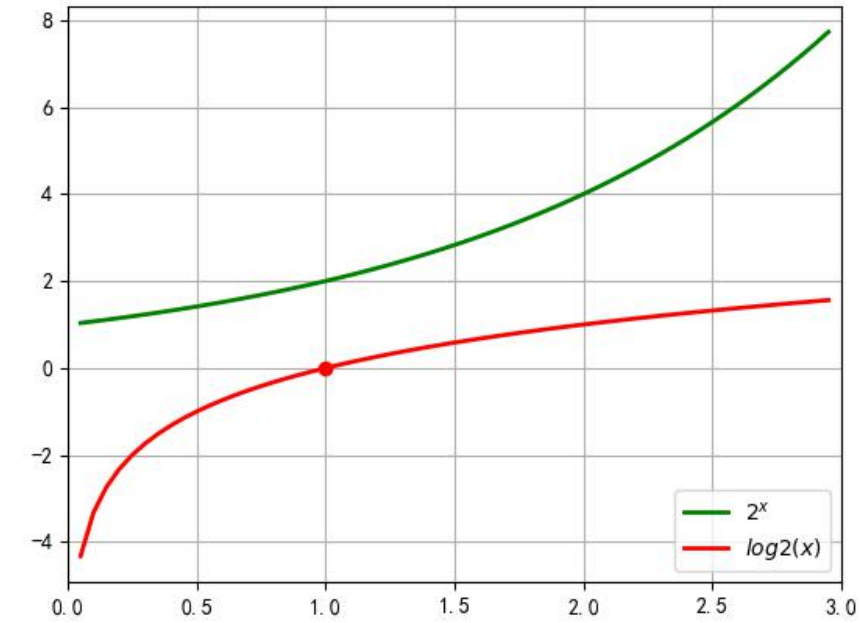
对数函数、指数函数

- 对数函数： $y=\log_a(x), a>0且a\neq 1$
- 指数函数： $y=a^x, a>0且a\neq 1$

$$\log_a x+\log_a y=\log_a(x*y)$$

$$\log_a x-\log_a y=\log_a \frac{x}{y}$$

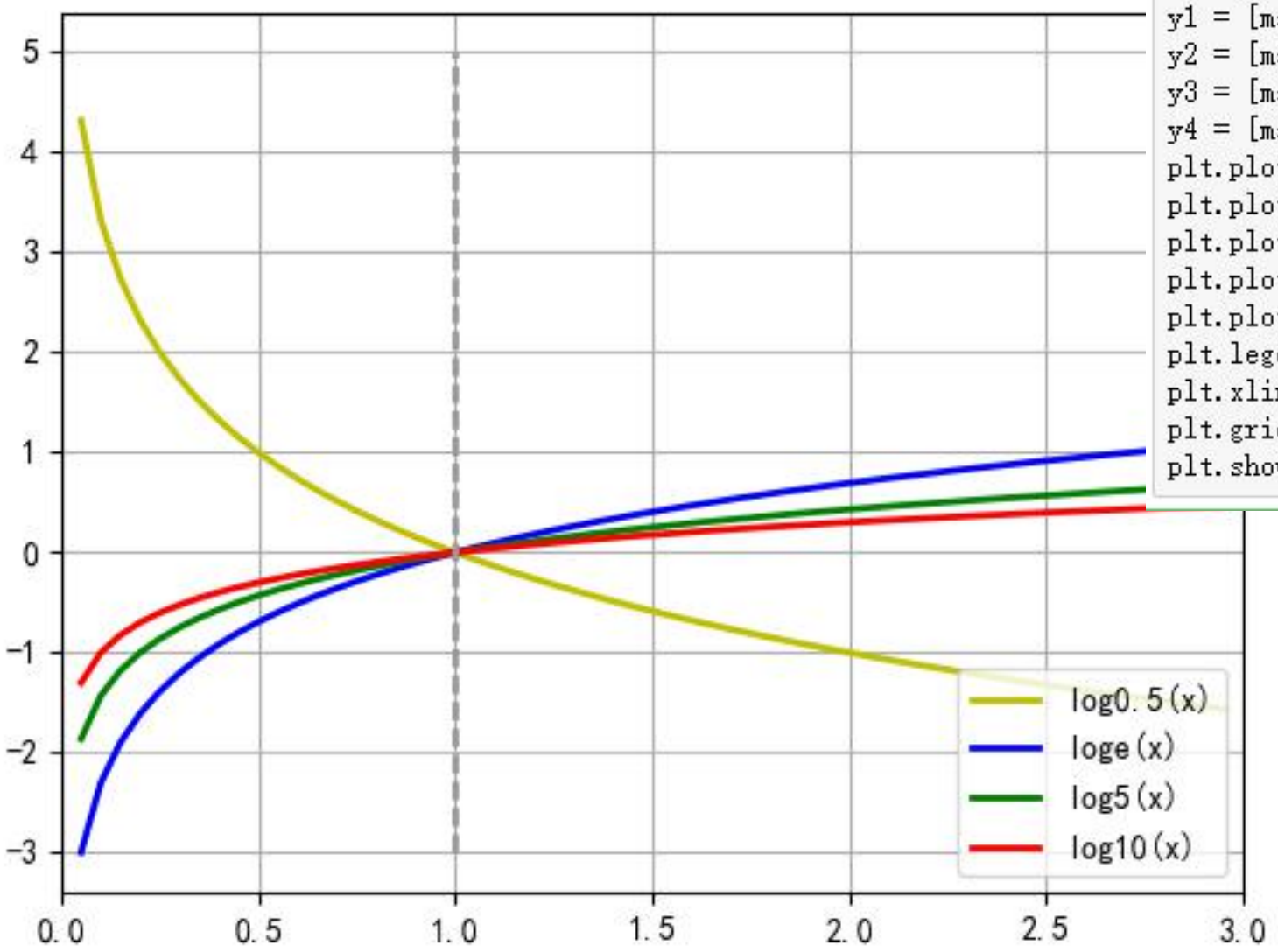
$$\frac{a^x}{a^y}=a^{x-y} \qquad a^x\cdot a^y=a^{x+y}$$



$$\log_a(1)=0$$

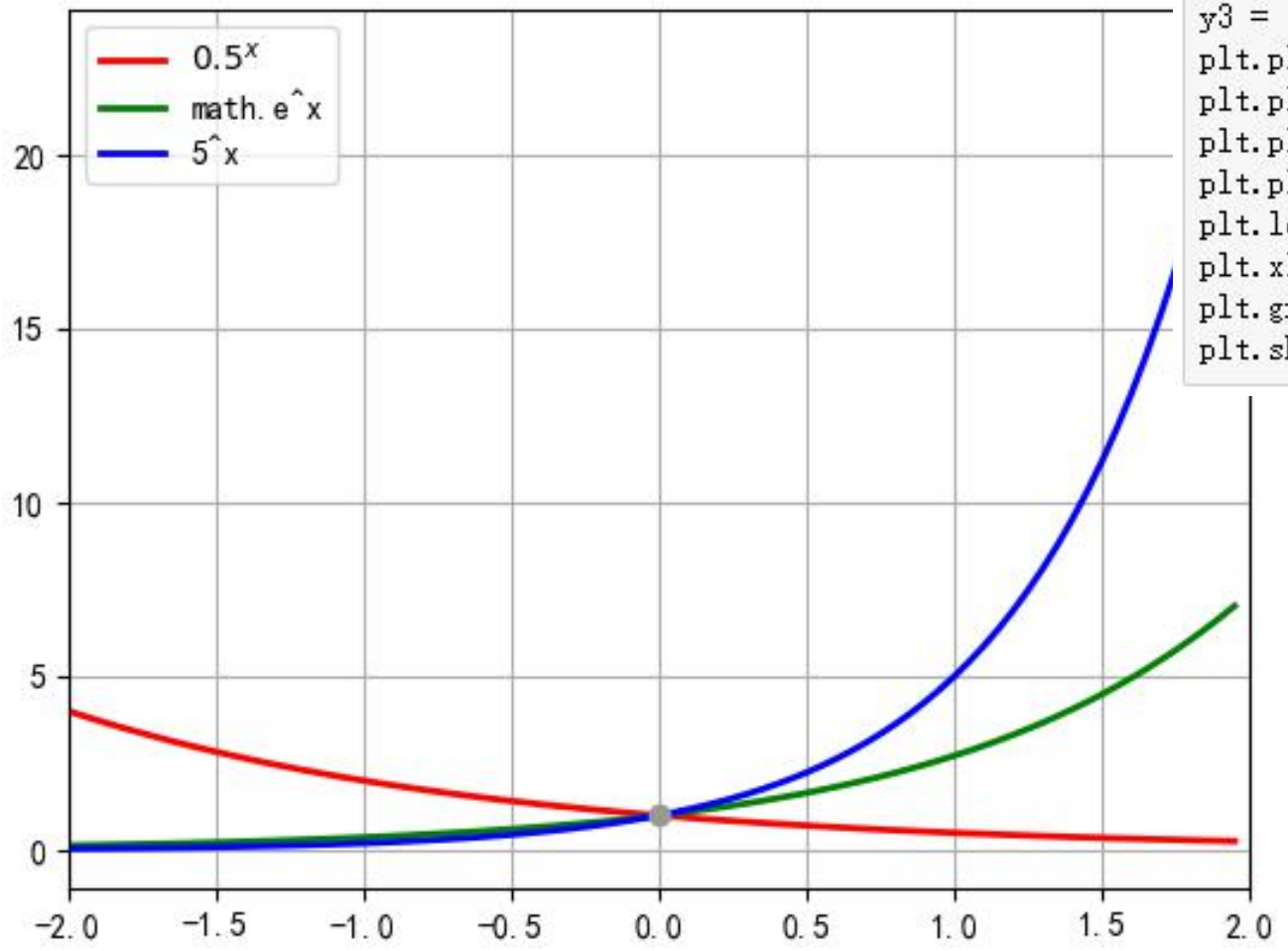
$$a^0=1$$

对数函数、指数函数




```
x = np.arange(0.05, 3, 0.05)
y1 = [math.log(i, 0.5) for i in x]
y2 = [math.log(i, math.e) for i in x]
y3 = [math.log(i, 5) for i in x]
y4 = [math.log(i, 10) for i in x]
plt.plot(x, y1, linewidth=2, color='y', label='log0.5(x)')
plt.plot(x, y2, linewidth=2, color='b', label='loge(x)')
plt.plot(x, y3, linewidth=2, color='g', label='log5(x)')
plt.plot(x, y4, linewidth=2, color='r', label='log10(x)')
plt.plot([1,1], [-3, 5], '—', color='#999999', linewidth=2)
plt.legend(loc='lower right')
plt.xlim(0, 3)
plt.grid(True)
plt.show()
```

对数函数、指数函数

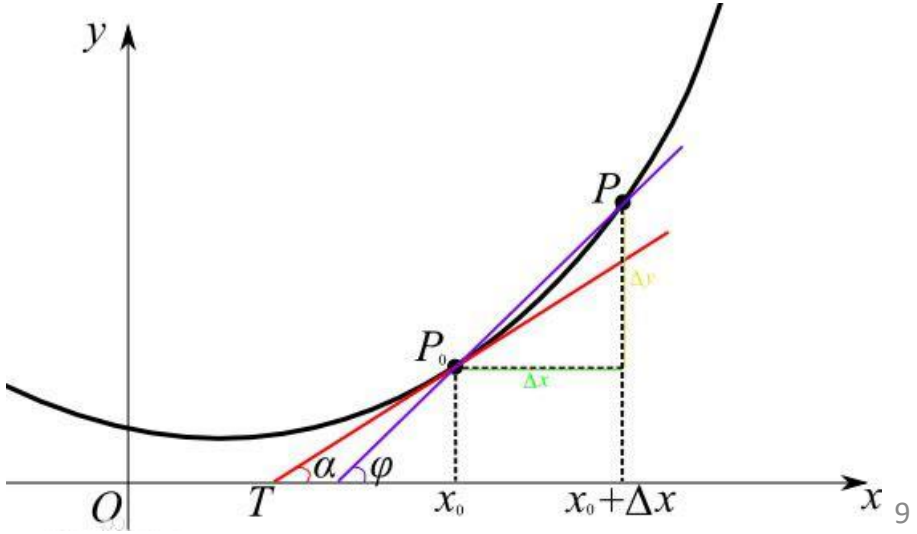



```
x = np.arange(-2, 2, 0.05)
y1 = [math.pow(0.5, i) for i in x]
y2 = [math.pow(math.e, i) for i in x]
y3 = [math.pow(5, i) for i in x]
plt.plot(x, y1, linewidth=2, color='r', label='$0.5^x$')
plt.plot(x, y2, linewidth=2, color='g', label='math.e^x')
plt.plot(x, y3, linewidth=2, color='b', label='5^x')
plt.plot([0], [1], 'o', color='#999999', linewidth=2)
plt.legend(loc='upper left')
plt.xlim(-2, 2)
plt.grid(True)
plt.show()
```

导数

- 一个函数在某一点的导数描述了这个函数在这一点附近的**变化率**，也可以认为是函数在某一点的导数就是该函数所代表的曲线在这一点的**切线斜率**。导数值越大，表示函数在该点处的变化越大。
- 定义：当函数y=f(x)在自变量$x=x_0$上产生一个增量Δx时，函数输出值的增量Δy和自变量增量Δx之间的比值在Δx趋近与0的时候存在极限值a，那么a即为函数在x_0处的导数值。

$$\lim_{\Delta x \to 0} \frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$



常见导函数

$$C'=0, C\text{为常数} \qquad \left(x^n\right)'=n\cdot x^{n-1}$$

$$\left(a^x\right)'=a^x\cdot\ln a \qquad \left(e^x\right)'=e^x$$

$$\left(\log_a x\right)'=\frac{1}{x\ln a} \qquad \left(\ln x\right)'=\frac{1}{x}$$

$$\left(u\pm v\right)'=u'\pm v' \qquad \left(uv\right)'=u'v+uv' \qquad \left(\frac{u}{v}\right)'=\frac{u'v-uv'}{v^2}$$

$$y=f\left(g(x)\right)\Rightarrow y'=\frac{df}{dg}\cdot\frac{dg}{dx}$$

偏导数

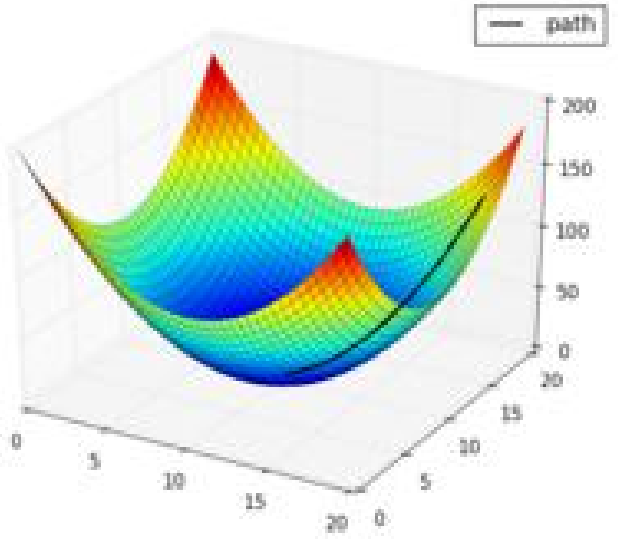
- 在一个多变量的函数中，偏导数就是关于其中一个变量的导数而保持其它变量恒定不变。假定二元函数z=f(x,y)，点(x_0,y_0)是其定义域内的一个点，将y固定在y_0上，而x在x_0上增量Δx，相应的函数z有增量$\Delta z=f(x_0+\Delta x, y_0) - f(x_0,y_0)$；Δz和Δx的比值当Δx的值趋近于0的时候，如果极限存在，那么此极限值称为函数z=f(x,y)在处对x的偏导数(partial derivative)

$$\text{对x的偏导数}:\left.\frac{\partial f}{\partial x}\right|_{\substack{x=x_0\\y=y_0}} \qquad \text{对y的偏导数}:\left.\frac{\partial f}{\partial y}\right|_{\substack{x=x_0\\y=y_0}}$$

- $z=x^2+xy^2$ 在（2,1）处的对x的偏导数=？

梯度

- 梯度：梯度是一个向量，表示某一函数在该点处的**方向导数**沿着该方向取的最大值，即函数在该点处沿着该方向变化最快，变化率最大(即该梯度向量的模)；当函数为一维函数的时候，梯度其实就是导数。



$$\nabla f(x_1, x_2) = \left(\frac{\partial f(x_1, x_2)}{x_1}, \frac{\partial f(x_1, x_2)}{x_2} \right)$$

Taylor公式

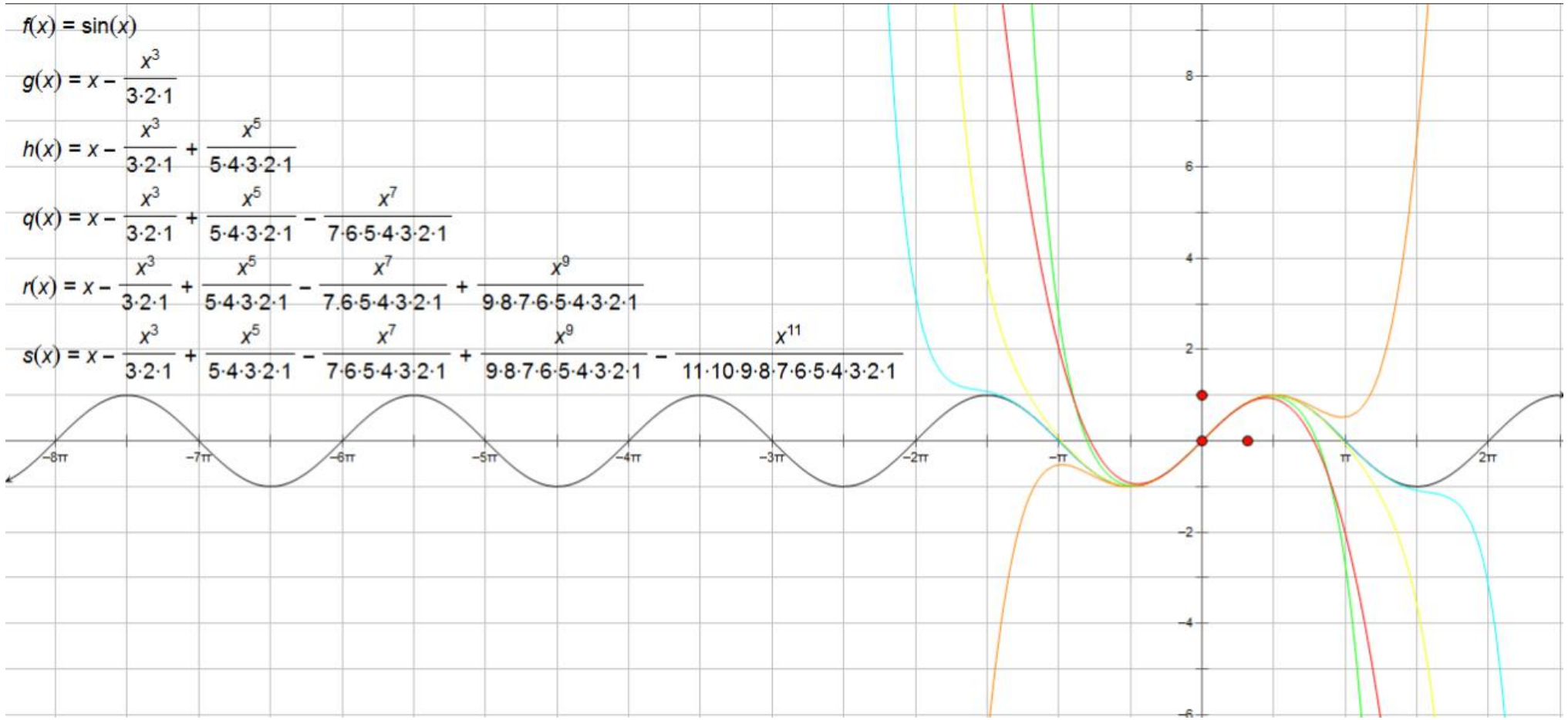
- Taylor(泰勒)公式是用一个函数在某点的信息描述其附近取值的公式。如果函数足够平滑，在已知函数在某一点的各阶导数值的情况下，Taylor公式可以利用这些导数值来做系数构建一个多项式近似函数在这一点的邻域中的值。
- 若函数f(x)在包含x_0的某个闭区间[a,b]上具有n阶函数，且在开区间(a,b)上具有n+1阶函数，则对闭区间[a,b]上任意一点x，有Taylor公式如下：<$f^{(n)}(x)$表示f(x)的n阶导数，$R_n(x)$是Taylor公式的余项，是$(x-x_0)^n$的高阶无穷小>

$$f(x)=\frac{f(x_0)}{0!}+\frac{f'(x_0)}{1!}(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\ldots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n(x)$$

$$f(x_0+\Delta x)=f(x_0)+f'(x_0)\Delta x+\frac{1}{2}f''(x_0)\Delta x^2$$

Taylor公式应用*

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + (-1)^{m-1}\frac{x^{2m-1}}{(2m-1)!} + R_{2m-1}(x)$$



Taylor公式应用2*

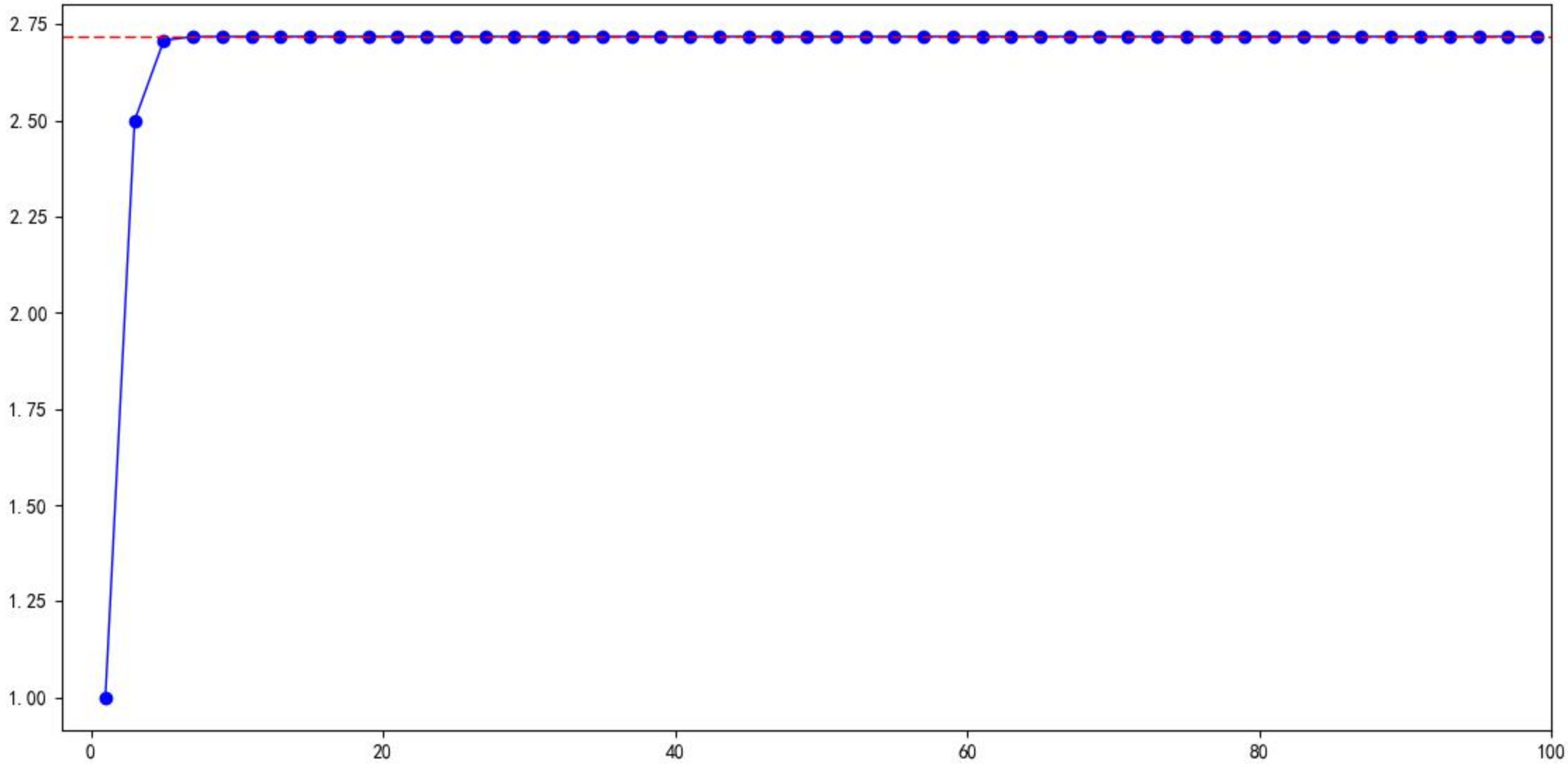
- 使用泰勒公式估算e的近似值，并估计误差值.

$$e^x \approx \sum_{k=0}^{n} \frac{e^{x_0}}{k!}(x-x_0)^k \xrightarrow{令x_0=0} e^x \approx 1+x+\frac{x^2}{2!}+\ldots+\frac{x^n}{n!}$$

$$\xrightarrow{令x=1} e \approx 1+1+\frac{1}{2!}+\frac{1}{3!}+\ldots+\frac{1}{n!} \xrightarrow{令n=10} e \approx 2.7182815$$

$$\delta = |R_{10}| = \frac{1}{11!}+\frac{1}{12!}+\ldots = \frac{1}{11!}\left(1+\frac{1}{12}+\frac{1}{12*13}+\ldots\right) < \frac{1}{11!}\left(1+\frac{1}{12}+\frac{1}{12^2}+\ldots\right) = \frac{12}{11*11!} = 2.73*10^{-8}$$

Taylor公式应用*



概率论

随机事件与样本空间

- 随机试验
  - 我们称一个试验为随机试验，如果它满足以下三个条件：
    - （1）试验可以在相同的条件下重复进行
    - （2）试验所有可能结果是明确可知道的，并且不止一个
    - （3）每次试验会出现哪一个结果，事先不能确定
  - 我们是通过随机试验来研究随机现象的，为了方便起见，将随机试验简称为**试验**，并用E表示

随机事件与样本空间

- 随机事件
  - 在一次试验中可能出现，也可能不出现的结果称为**随机事件**，简称为**事件**，并用大写字母A，B，C等表示。
  - 为讨论需要，将一次试验一定发生的事件称为**必然事件**，记为Ω。每次试验一定不会发生的事件称为不可能事件，记为Φ
- 样本空间
  - 随机试验每一个最简单、最基本的结果称为**基本事件**（或**样本点**），记为ω。基本事件（或样本点）的全体称为**基本事件空间**（或**样本空间**），记为Ω，即Ω={ω}。随机事件A总是由若干个基本事件组成，即A是Ω的子集

事件的关系与运算

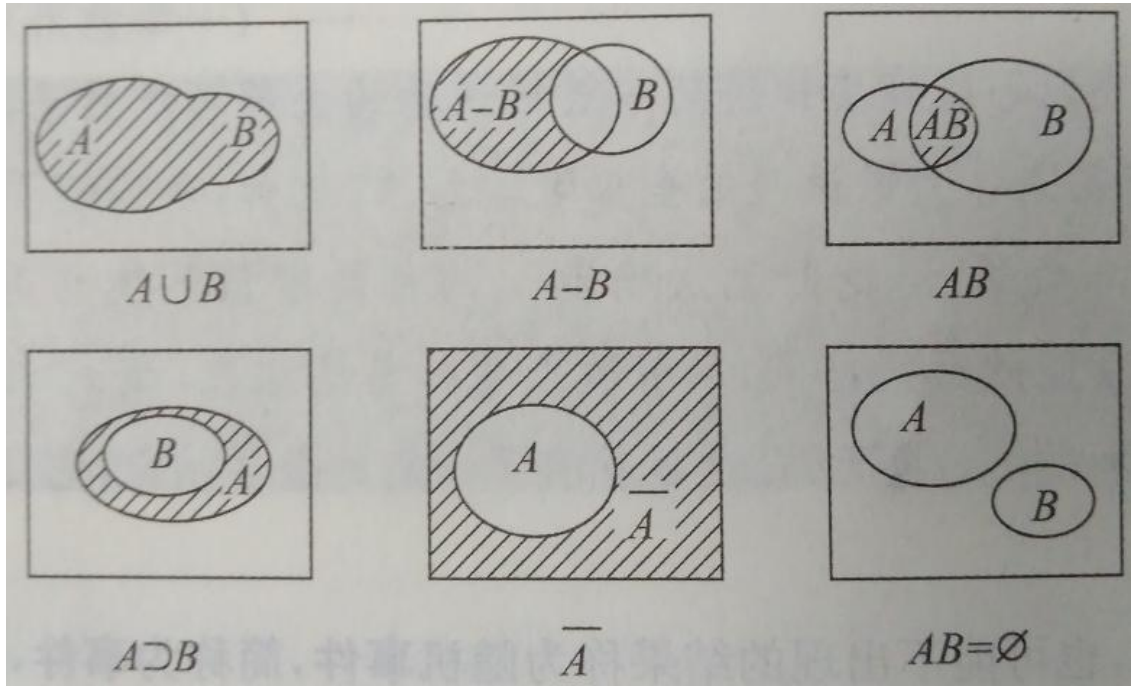
- 定义【关系：包含、相等、相容、对立；运算：和（并）、差、交（积）】
  - （1）如果事件A发生必导致事件B发生，则称事件B**包含**事件A，记为A⊂B
  - （2）如果A⊂B，B⊂A，则称事件A和B**相等**，记为A=B
  - （3）称“事件A与事件B同时发生”的事件为事件A与B的**交（积）**，记为A∩B或AB
  - （4）若AB≠Φ，则称事件“A和B**相容**”；若AB=Φ，则称事件“A和B**互不相容**”，也叫互斥。
  - （5）称“事件A与B至少有一个发生”的事件为事件A与B的**并（和）**，记为A∪B
  - （6）称“事件A发生而事件B不发生”的事件为事件A与B的**差**，记为A-B；称“事件A不发生”的事件为事件A的对立事件

事件的关系与运算

- 定义【关系：包含、相等、相容、对立；运算：和（并）、差、交（积）】
  - （7）称有限个（或可列个）事件A1， A2， …， An(…)构成一个完备事件组，如果

$$\bigcup_{i=1}^{n} A_i(\text{或}\bigcup_{i=1}^{\infty} A_i)=\Omega,\ A_iA_j=\phi(\text{对一切}i\neq j)$$

  - （8）事件的关系和运算可以用文氏图形象的表示出来，矩形表示必然事件Ω



事件的关系与运算

- 事件的关系和运算法则：
- （1）吸收律
- （2）交换律
- （3）结合律
- （4）分配律
- （5）对偶律（德摩根律）

概率的概念和基本性质

- 定义
  - 通常我们将随机事件A发生的可能性的大小的度量，称为事件A的概率，记为P(A)
- 性质
  - 性质1 $P(\Phi)=0$
  - 性质2（有限可加性）
  - 性质3（单调性）
  - 性质4（有界性）
  - 性质5（逆事件概率）
  - 性质6（加法公式）
  - 性质7（减法公式）

古典型概率和几何型概率

- 称随机试验(随机现象)的概率模型为古典概型，如果其基本事件空间(样本空间)满足:
  - （1）只有有限个基本事件(样本点);
  - （2）每个基本事件(样本点)发生的可能性都一样.
- 如果古典概型的基本事件总数为n，事件A包含k个基本事件，也叫作有利于A的基本事件为k个，则A的概率定义为

$$P(A)=\frac{k}{n}=\frac{\text{事件A所包含基本事件的个数}}{\text{基本事件总数}}$$

古典型概率和几何型概率

- 称随机试验(随机现象)的概率模型为几何概型，如果：
  - (1)样本空间(基本事件空间)是一个可度量的几何区域；
  - (2)每个样本点(基本事件)发生的可能性都一样，即样本点落入Ω的某一可度量的子区域S的可能性大小与S的几何度量成正比，而与S的位置及形状无关
- 在几何概型随机试验中，如果S_A是样本空间Ω的一个可度量的子区域，则事件A={样本点落入区域S_A}的概率定义为

$$P(A)=\frac{S_A\text{的几何度量}}{\Omega\text{的几何度量}}$$

条件概率及三个公式（乘法公式、全概率公式、贝叶斯公式）

- 条件概率
  - 设A，B为任意两个事件，若P(A)>0，我们称在已知事件A发生的条件下，事件B发生的概率为条件概率，记为P(B|A)，并定义

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

- 乘法公式
  - 如果P(A)>0，则P(AB)=P(A)P(B|A)
  - 如果$P(A_1 \ldots A_{n-1})>0$，则$P(A_1 \ldots A_n)= P(A_1)\ P(A_2|A_1)\ P(A_3|A_1A_2) \ldots P(A_n|A_1 \ldots A_n)$

条件概率及三个公式（乘法公式、全概率公式、贝叶斯公式）

- 全概率公式
  - 如果 $\bigcup_{i=1}^{n} A_i = \Omega$，$A_i A_j = \phi$(对一切$i \neq j$)，P(Ai)>0，则对任一事件B，有

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B \mid A_i)$$

  - 全概率公式是用于计算某个“结果” B发生的可能性大小。如果一个结果B的发生总是与某些前提条件Ai 相联系，那么在计算P(B)时，我们就要用Ai 对B作分解，应用全概率公式计算P(B)，我们常称这种方法为**全集分解法**。
  - 根据小偷们的资料，计算村子今晚失窃概率的问题（今晚有且仅有一个小偷作案）：P(Ai)表示小偷 i 作案的概率，P(B|Ai)表示小偷 i 作案成功的概率，那么P(B)就是村子失窃的概率

条件概率及三个公式（乘法公式、全概率公式、贝叶斯公式）

- 贝叶斯公式（又称逆概公式）
  - 如果$\bigcup_{i=1}^{n} A_i = \Omega$，$A_i A_j = \phi$(对一切$i \neq j$)，P(Ai)>0，则对任一事件B，只要P(B)>0，有

$$P(A_j \mid B) = \frac{P(A_j B)}{P(B)} = \frac{P(A_j)P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)} (i, j = 1,2,\ldots,n)$$

  - 如果在B发生的条件下探求导致这一结果的各种“原因” Ai 发生的可能性大小 P(Ai |B)，则要应用贝叶斯公式
  - 若村子今晚失窃，计算哪个小偷嫌疑最大的问题（嫌疑最大就是后验概率最大）

贝叶斯公式

- 假设小偷1和小偷2在某村庄的作案数量比为3:2，前者偷窃成功的概率为0.02，后者为0.01，现村庄失窃，求这次失窃是小偷1作案的概率。
- 【分析】A_1={小偷1作案}，A_2={小偷2作案}，B={村庄失窃}

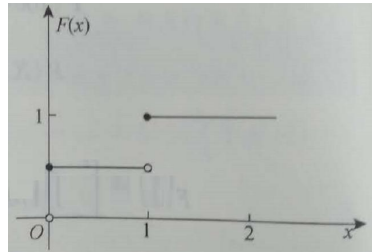
$$P(A_1)=3/5,\quad P(A_2)=2/5$$

$$P(B\mid A_1)=0.02,\quad P(B\mid A_2)=0.01$$

$$P(A_1\mid B)=\frac{P(A_1)P(B\mid A_1)}{\sum_{i=1}^{2}P(A_i)P(B\mid A_i)}=\frac{3/5\times 0.02}{3/5\times 0.02+2/5\times 0.01}=3/4$$

随机变量及其分布函数

- 随机变量
  - 投一个硬币，“正面朝上”是一个基本事件，“反面朝上”也是一个基本事件，可以分别用X=0和X=1来表示。
  - “正面朝上”和“反面朝上”都是随机事件，那么X的值也会随机而定，因此，就称X为随机变量
  - 注：随机事件是从静态的观点来研究随机现象，而随机变量则是一种动态的观点
- 分布函数
  - 设X是随机变量，x是任意实数，称F(x)=P{X≤x}为随机变量X的分布函数，或称X服从分布F(x)，记为X~F(x)

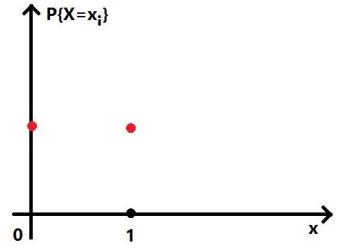


离散型和连续型随机变量

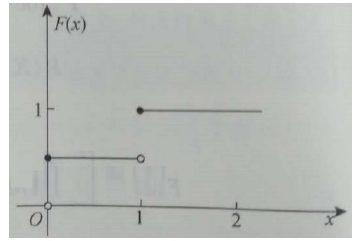
- 离散型随机变量及其概率分布
  - 如果随机变量X只可能取有限个或可列个值x_1， x_2， …，则称X为**离散型随机变量**，称 $p_i = P\{X = x_i\}, i = 1,2,...$ 为X的**概率分布律**，记为 $X \sim p_i$，概率分布常常用表格形式表示，即

X	x_1	x_2	…
P	p_1	p_2	…

X	0	1
P	0.5	0.5



  - 离散型随机变量X的分布函数为： $F(x)=P\{X \le x\}$



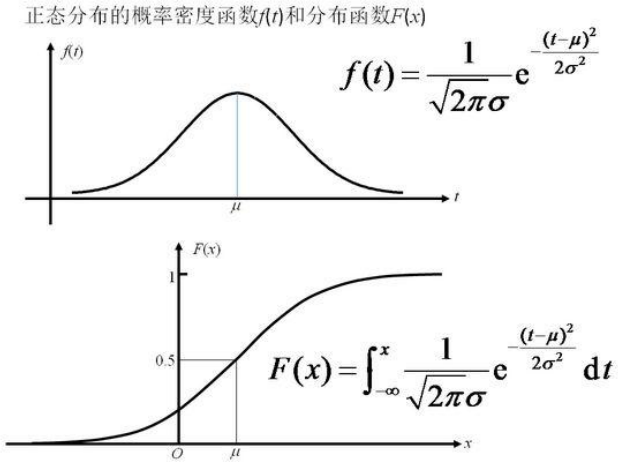
  - 注：既可以说X服从某一概率分布，也可以说X服从某一分布

离散型和连续型随机变量*

- 连续型随机变量及其概率分布
  - 如果随机变量X能在实数域R上取值，则称X为**连续型随机变量**，称 f(x) 为X的**概率密度函数**，记为X~f(x)
  - 连续型随机变量X的分布函数为：

$$F(x)=P\{X\le x\}=\int_{-\infty}^{x} f(t)dt,\quad x\in R$$

  - 注：既可以说X服从某一概率分布，也可以说X服从某一分布



期望

- 期望(mean)：也就是均值，是概率加权下的“平均值”，是每次可能结果的概率乘以其结果的总和，反映的是随机变量平均取值大小。常用符号μ表示：
- 连续性数据：

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

- 离散型数据：

$$E(X)=\sum_{i} x_i p_i$$（演示见下页）

期望

X	2	4	6	8	10
P(x)	0.2	0.2	0.2	0.2	0.2

$$E(X)=\sum_i x_i p_i$$

$$=2*0.2+4*0.2+6*0.2+8*0.2+10*0.2$$

$$=6$$

期望

假设C为一个常数，X和Y是两个随机变量，那么期望有以下性质：

$$E(C)=C \qquad E(CX)=CE(X)$$

$$E(X+Y)=E(X)+E(Y)$$

如果X和Y相互独立，那么$E(XY)=E(X)E(Y)$

方差

- 方差(variance)是对随机变量或一组数据离散程度的度量，用来度量随机变量和其数学期望之间的偏离程度。即方差是衡量数据原数据和期望/均值相差的度量值。

$$Var(X)=D(X)=\sigma^2=\frac{1}{N}\sum_{i=i}^{N}(X-\mu)^2$$

$$D(X)=\sum_{i=1}^{n}p_i(x_i-\mu)^2 \qquad D(X)=\int_a^b(x-\mu)^2 f(x)dx$$

$$D(X)=E\left((X-E(X))^2\right)$$

$$=E\left(X^2-2XE(X)+E(X)^2\right)=E\left(X^2\right)-\left(E(X)\right)^2$$

方差

X	2	4	6	8	10
P(x)	0.2	0.2	0.2	0.2	0.2

$$D(X)=\sum_{i=1}^{n}p_i(x_i-\mu)^2$$

$$=0.2*16+0.2*4+0.2*0+0.2*4+0.2*16$$

$$=8$$

$$E(X)=6 \qquad E(X^2)=44$$

$$D(X)=E(X^2)-(E(X))^2$$

$$=44-6^2$$

$$=8$$

方差

- 假设C为一个常数，X和Y是两个随机变量，那么方差有以下性质：

$$D(C)=0 \qquad D(CX)=C^2D(X) \qquad D(C+X)=D(X)$$

$$D(X\pm Y)=D(X)+D(Y)\pm 2Cov(X,Y)$$

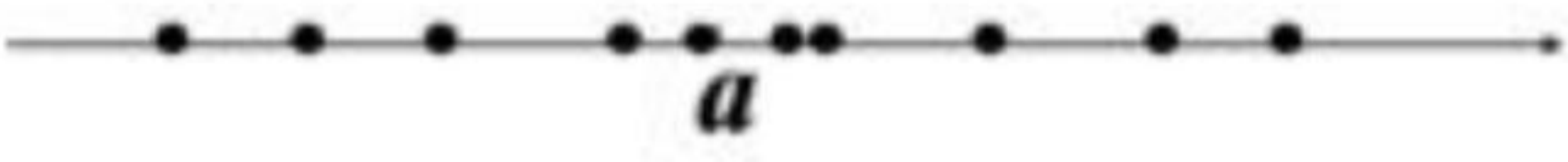
协方差$Cov(X,Y)=E\{(X-E(X))\cdot(Y-E(Y))\}$

如果X和Y相互独立，那么$D(X\pm Y)=D(X)+D(Y)$

- 假设X、Y相互独立，X~N(0,σ^2)，Y~N(1,σ^2)，那么X+Y~N?，X-Y~?

方差

- 已知某零件的真实长度为a，现用甲、乙两台仪器各测量10次，将测量结果X用坐标轴上的点表示，并且甲仪器的测量如图中的点，乙仪器的测量结果全是a，此时两台仪器的测量均值都是a，但是我们可能会认为乙机器性能更好，因为乙机器的测量值在a附近。因此可见，研究随机变量与其均值的偏离程度是十分必要的。



标准差

- 标准差(Standard Deviation)是离均值平方的算术平均数的平方根，用符号σ表示，其实标准差就是方差的算术平方根。
- 标准差和方差都是测量离散趋势的最重要、最常见的指标。标准差和方差的不同点在于，标准差和变量的计算单位是相同的，比方差清楚，因此在很多分析的时候使用的是标准差。

$$\sigma = \sqrt{D(X)} = \sqrt{\frac{\sum (X-\mu)^2}{N}}$$

标准差

X1	2	4	6	8	10
P(x1)	0.2	0.2	0.2	0.2	0.2

X2	4	5	6	7	8
P(x2)	0.2	0.2	0.2	0.2	0.2

$$D(X_1)=8 \qquad \sigma_1=\sqrt{D(X_1)}=\sqrt{8}=2.8284$$

$$D(X_2)=2 \qquad \sigma_2=\sqrt{D(X_2)}=\sqrt{2}=1.4142$$

协方差

- 协方差常用于衡量两个变量的总体误差；当两个变量相同的情况下，协方差其实就是方差。
- 如果X和Y是相互独立的，那么二者之间的协方差为零。但是如果协方差为零，只能推出X和Y是不相关的。

$$
\begin{aligned}
Cov(X,Y) &= E\left[(X-E(X))\cdot(Y-E(Y))\right] \\
&= E\left[XY - XE(Y) - YE(X) + E(X)E(Y)\right] \\
&= E(XY) - E(X)E(Y)
\end{aligned}
$$

协方差

- 假设a、b为常数，X和Y是两个随机变量，那么协方差有性质如下所示：

$$Cov(X,Y)=Cov(Y,X)$$

$$Cov(aX,bY)=abCov(X,Y)$$

$$Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$$

协方差

- 协方差是两个随机变量具有相同方向变化趋势的度量：
  - 若Cov(X,Y) > 0, 则X和Y的变化趋势相同；
  - 若Cov(X,Y) < 0, 则X和Y的变化趋势相反；
  - 若Cov(X,Y) = 0，则X和Y不相关，也就是变化没有什么相关性

协方差矩阵

- 对于n个随机变量$X_1,X_2,X_3....X_n$, 任意两个随机变量X_i和X_j都可以得到一个协方差，从而形成一个n*n的矩阵，该矩阵就叫做协方差矩阵，协方差矩阵为对称矩阵。

$$c_{ij} = E\{[X_i - E(X_i)][X_j - E(X_j)]\} = Cov(X_i, X_j)$$

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

大数定律

- 大数定律的意义：随着样本容量n的增加，样本平均数将接近于总体平均数(期望μ)，所以在统计推断中，一般都会使用样本平均数估计总体平均数的值。
- 也就是我们会使用一部分样本的平均值来代替整体样本的期望/均值，出现偏差的可能是存在的，但是当n足够大的时候，偏差的可能性是非常小的，当n无限大的时候，这种可能性的概率基本为0。
- 大数定律的主要作用就是为使用**频率**来估计**概率**提供了理论支持；为使用部分数据来近似的模拟构建全部数据的特征提供了理论支持。

大数定律


```
# 解决中文显示问题
mpl.rcParams['font.sans-serif'] = [u'SimHei']
mpl.rcParams['axes.unicode_minus'] = False

# 给定随机数的种子
random.seed(28)

def generate_random_int(n):
    """产生n个1-9的随机数"""
    return [random.randint(1, 9) for i in range(n)]
```

```
if __name__ == '__main__':
    number = 8000
    x = [i for i in range(number + 1) if i != 0]
    # 产生number个[1,9]的随机数
    total_random_int = generate_random_int(number)
    # 求n个[1,9]的随机数的均值, n=1,2,3,4,5.....
    y = [np.mean(total_random_int[0:i + 1]) for i in range(number)]

    plt.plot(x, y, 'b-')
    plt.xlim(0, number)
    plt.grid(True)
    plt.show()
```

中心极限定理

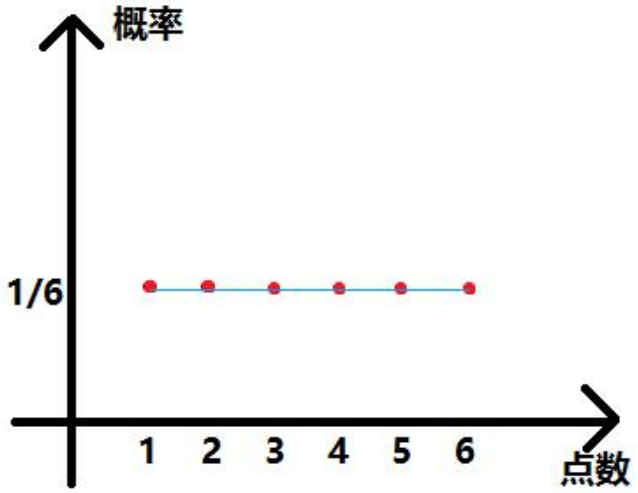
- 中心极限定理(**Central Limit Theorem**)：在**独立同分布**的情况下，当随机变量个数趋于无穷时，这些随机变量的和近似服从正态分布。
- 假设{$X_1, X_2,\ldots,X_n$}为**独立同分布**的随机变量序列，即具有相同的期望μ和方差为σ^2，当n趋近于无穷时，Y_n近似地服从正态分布，其中，Y_n为随机序列{$X_1, X_2,\ldots,X_n$}的和：

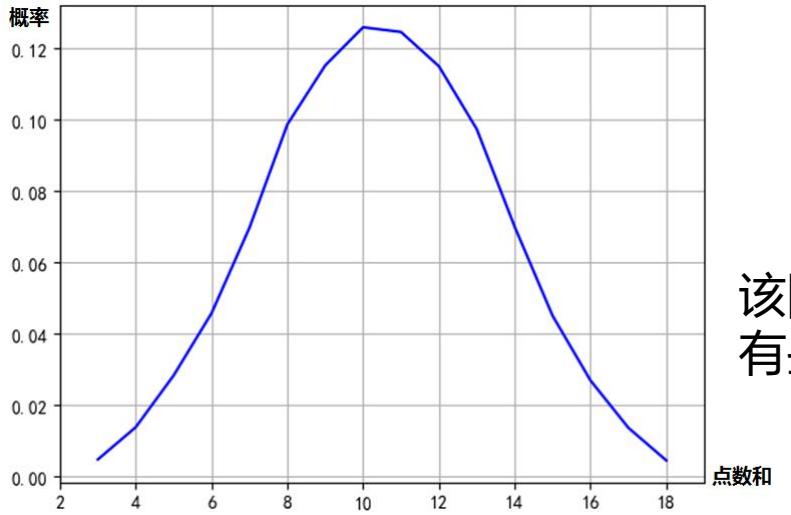
$$Y_n = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$

中心极限定理解释

- 用随机变量X表示投骰子出现的点数，左图为X的概率分布律。假设{$X_1, X_2,....,X_n$}为**独立同分布**的随机变量序列，Y_n为随机序列{$X_1, X_2,....,X_n$}的和，当n趋近于无穷时，Y_n近似地服从正态分布

$$Y_n = X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$

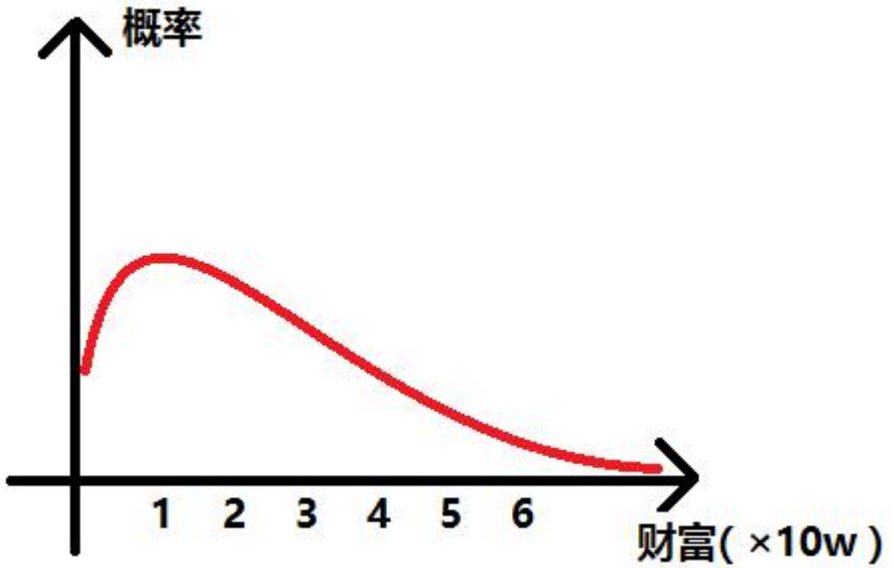




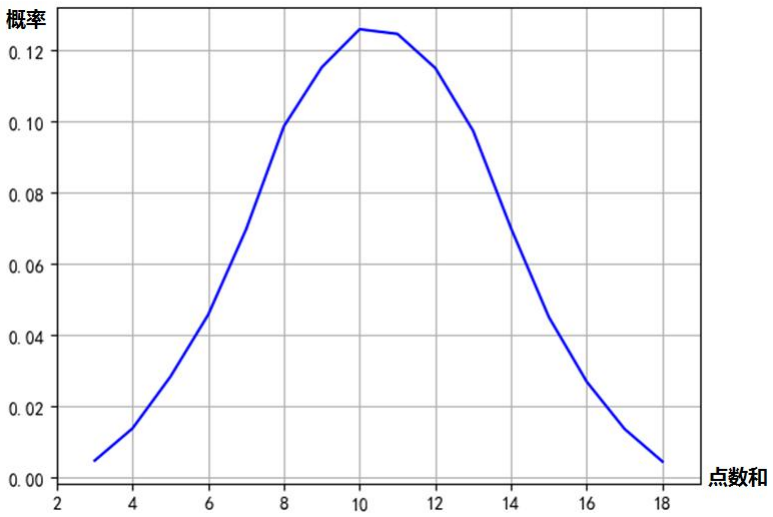
该图为：3个随机变量的和的概率分布有些正态分布的意味，但还差得远

中心极限定理思考

- 用随机变量X表示一个人拥有的财富，X服从图中的概率分布。假设$\{X_1, X_2,\ldots,X_n\}$是一个随机变量序列且都服从X的概率分布，Y_n为随机序列$\{X_1, X_2,\ldots,X_n\}$的和，当n趋近于无穷时，Y_n近似地服从正态分布吗？



模拟中心极限定理的程序




```python
if __name__ == '__main__':
    # 进行A事件多少次
    number1 = 10000000
    # 表示每次A事件抛几次骰子
    number2 = 3

    # 进行number1次事件A的操作，每次事件A都进行number2次抛骰子
    keys = [generate_mean(number2) for i in range(number1)]

    # 统计每个和数字出现的次数，eg: 和为3的出现多少次、和为10出现多少次....
    result = {}
    for key in keys:
        count = 1
        if key in result:
            count += result[key]
        result[key] = count

    # 获取x和y
    x = sorted(np.unique(list(result.keys())))
    y = []
    for key in x:
        # 将出现的次数进行一个百分比的计算
        y.append(result[key] / number1)

    # 画图
    plt.plot(x, y, 'b-')
    plt.xlim(x[0] - 1, x[-1] + 1)
    plt.grid(True)
    plt.show()
```

极大似然估计

- **极大似然估计(Maximum Likelihood Estimation, MLE)**也称最大似然估计，是一种具有理论性的参数估计方法。基本思想是：当从模型总体随机抽取n组样本观测值后，最合理的参数估计量应该使得从模型中抽取该n组样本观测值的概率最大；一般步骤如下：
  - 1. 写出似然函数；
  - 2. 对似然函数取对数，并整理；
  - 3. 求导数；
  - 4. 解似然方程

极大似然估计例题

- 一个暗箱里有三种球（1，2，3），其概率分布律如表所示，进行有放回的抽样，得到了 1 2 2 2 1 2 2 3 1 3，记为$x_1, x_2, \ldots, x_n$，现在想通过极大似然估计的方法，估计θ

X	1	2	3
P{X=?}	0.5θ	0.3+0.4θ	0.7-0.9θ

$$L(\theta)=\prod_{i=1}^{N}P\{X=x_i\}=\left[P\{X=1\}\right]^3\left[P\{X=2\}\right]^5\left[P\{X=3\}\right]^2$$

$$=\prod_{i=1}^{N}\left[P\{X=1\}\right]^{\mathrm{I}(x_i=1)}\left[P\{X=2\}\right]^{\mathrm{I}(x_i=2)}\left[P\{X=3\}\right]^{\mathrm{I}(x_i=3)}$$

$$=(0.5\theta)^3(0.3+0.4\theta)^5(0.7-0.9\theta)^2$$

$$\ln L(\theta)=3\ln 0.5\theta+5\ln(0.3+0.4\theta)+2\ln(0.7-0.9\theta)$$

求导，令$=0$,求出$\theta=0.5598$

线性代数

向量的运算

设两向量为：$\vec{a}=(x_1,y_1)$ $\vec{b}=(x_2,y_2)$

向量的加法/减法满足平行四边形法则和三角形法则

$$\vec{a}+\vec{b}=(x_1+x_2,y_1+y_2)$$

$$\vec{a}-\vec{b}=(x_1-x_2,y_1-y_2)$$

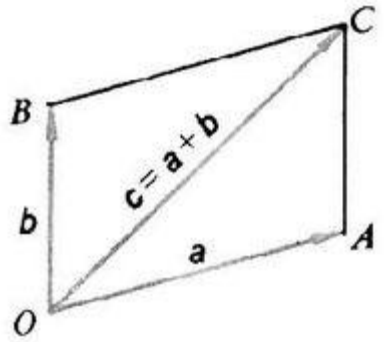


图 4　向量的加法

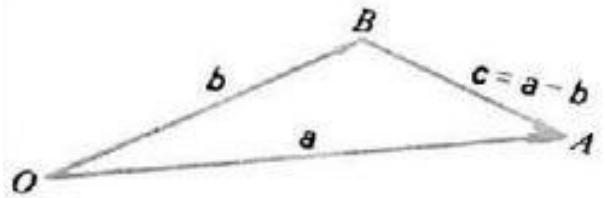


图 5　向量的减法

数乘：实数λ和向量a的乘积还是一个向量，记作λa，且|λa|=λ|a|；数乘的几何意义是将向量a进行伸长或者压缩操作

$$\lambda\vec{a}=(\lambda x_1,\lambda y_1)$$

向量的运算

设两向量为：$\vec{a}=(x_1,y_1)$ $\vec{b}=(x_2,y_2)$并且a和b之间的夹角为θ

数量积：两个向量的数量积(内积、点积)是一个数量/实数，记作 $\vec{a}\cdot\vec{b}$

$$\vec{a}\cdot\vec{b}=\left|\vec{a}\right|*\left|\vec{b}\right|*\cos\theta$$

向量积：两个向量的向量积(外积、叉积)是一个向量，记作 $\vec{a}\times\vec{b}$；向量积即两个不共线非零向量所在平面的一组法向量。

$$\left|\vec{a}\times\vec{b}\right|=\left|\vec{a}\right|*\left|\vec{b}\right|*\sin\theta$$

矩阵的直观表示

- 数域F中m*n个数排成m行n列，并括以圆括弧(或方括弧)的数表示成为数域F上的矩阵，通常用大写字母记作A或者A_{m*n}，有时也记作$A=(a_{ij})_{m*n}(i=1,2...,m;j=1,2,...n)$，其中$a_{ij}$表示矩阵A的第i行的第j列元素

$$A=\begin{Bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Bmatrix}$$

矩阵的加减法

- 矩阵的加法与减法要求进行操作的两个矩阵A和B具有相同的阶，假设A为m*n阶矩阵，B为m*n阶矩阵，那么C=A ± B也是m*n阶的矩阵，并且矩阵C的元素满足：$c_{ij} = a_{ij} \pm b_{ij}$

$$A = \begin{Bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Bmatrix} \quad B = \begin{Bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{Bmatrix}$$

$$C = A \pm B = \begin{Bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1n} \pm b_{1n} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2n} \pm b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} \pm b_{m1} & a_{m2} \pm b_{m2} & \cdots & a_{mn} \pm b_{mn} \end{Bmatrix}$$

矩阵的加减法

- 记 $A_{m\times n}=(a_{ij})\quad B_{m\times n}=(b_{ij})$

加法： $A_{m\times n}+B_{m\times n}=(a_{ij}+b_{ij})$

$$\begin{bmatrix}1&2\\5&6\end{bmatrix}+\begin{bmatrix}3&4\\7&8\end{bmatrix}=\begin{bmatrix}4&6\\12&14\end{bmatrix}$$

减法：

$$A_{m\times n}-B_{m\times n}=(a_{ij}-b_{ij})$$

$$\begin{bmatrix}6&8\\11&14\end{bmatrix}-\begin{bmatrix}3&4\\7&8\end{bmatrix}=\begin{bmatrix}3&4\\4&6\end{bmatrix}$$

运算律：

交换律 $A+B=B+A$

结合律 $(A+B)+C=A+(B+C)$

矩阵与数的乘法

- 数乘：将数λ与矩阵A相乘，就是将数λ与矩阵A中的每一个元素相乘，记作λA；结果C=λA，并且C中的元素满足： $c_{ij} = \lambda a_{ij}$

$$A = \begin{Bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Bmatrix} \qquad \lambda A = \begin{Bmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{Bmatrix}$$

矩阵与数的乘法

- 数乘： $\mathrm{kA} = (ka_{ij})$

$$3 \times \begin{bmatrix} 1 & 2 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 3 \times 1 & 3 \times 2 \\ 3 \times 5 & 3 \times 6 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 15 & 18 \end{bmatrix}$$

- 数乘运算律

结合律 $(\lambda\mu)A = \lambda(\mu A)$

分配律 $(\lambda + \mu)A = \lambda A + \mu A$

$\lambda(A + B) = \lambda A + \lambda B$

矩阵与向量的乘法

- 假设A为m*n阶矩阵，x为n*1的列向量，则Ax为m*1的列向量，记作：$\vec{y} = A\vec{x}$

$$A = \begin{Bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Bmatrix} \qquad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} \qquad \vec{y} = A\vec{x} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{pmatrix}$$

矩阵与矩阵的乘法

- 矩阵的乘法仅当第一个矩阵A的列数和第二个矩阵B的行数相等时才能够定义，假设A为m*s阶矩阵，B为s*n阶矩阵，那么C=AB是m*n阶矩阵

$$C = AB = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 1\times1+2\times2+3\times3 & 1\times4+2\times5+3\times6 \\ 4\times1+5\times2+6\times3 & 4\times4+5\times5+6\times6 \end{pmatrix} = \begin{pmatrix} 14 & 32 \\ 32 & 77 \end{pmatrix}$$

矩阵与矩阵的乘法

- 例：

$$A=\begin{bmatrix}1&0&-1&2\\-1&1&3&0\\0&5&-1&4\end{bmatrix}\qquad B=\begin{bmatrix}0&3&4\\1&2&1\\3&1&-1\\-1&2&1\end{bmatrix}$$

求C=AB

$$C=AB=\begin{bmatrix}1&0&-1&2\\-1&1&3&0\\0&5&-1&4\end{bmatrix}\begin{bmatrix}0&3&4\\1&2&1\\3&1&-1\\-1&2&1\end{bmatrix}=\begin{bmatrix}-5&6&7\\10&2&-6\\-2&17&10\end{bmatrix}$$

$$(AB)C=A(BC)\quad (A+B)C=AC+BC\quad C(A+B)=CA+CB$$

矩阵的转置

- 矩阵的转置：把矩阵A的行和列互相交换所产生的矩阵称为A的转置矩阵，这一过程叫做矩阵的转置。 使用A^T表示A的转置

$$A=\begin{Bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{Bmatrix} \qquad A^T=\begin{Bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{Bmatrix}$$

矩阵的转置

- 例：$A=\begin{bmatrix}1 & 2\\ 5 & 6\end{bmatrix}$，A的转置为 $A^T=\begin{bmatrix}1 & 5\\ 2 & 6\end{bmatrix}$
- 转置的运算性质：

$$(A^T)^T=A$$

$$(\lambda A)^T=\lambda A^T$$

$$(AB)^T=B^TA^T$$

$$(A+B)^T=A^T+B^T$$

数学知识回顾

- 常见函数
- **导数、梯度《求导的方式、导数/梯度的含义/作用》**
- Taylor公式（SVM高斯核函数、XGboost）
- **联合概率、条件概率、全概率公式、贝叶斯公式**
- 期望、方差、协方差《了解这三个东西表示数据具有什么样的特性》
- 大数定律、中心极限定理
- **最大似然估计(MLE)《最大似然估计必须掌握》**
- **向量、矩阵的运算**

PythonApi回顾

Python科学计算库回顾

- Python科学计算库主要是为机器学习提供了一些便捷、封装好的API，那么在实际工作中，主要是将其应用在机器学习的特征工程阶段，主要涉及到的库有以下几个：
  - **NumPy**-数学计算基础库：N维数组、线性代数计算、傅立叶变换、随机数等；
  - **SciPy**-数值计算库：线性代数、拟合与优化、插值、数值积分、稀疏矩阵、图像处理、统计等；
  - **Pandas**-数据分析库：数据导入、整理、处理、分析等；
  - **Matplotlib**-会图库：绘制二维图形和图表。

Python科学计算库回顾

- 官网：https://scipy.org/

- 参考文档：
  - http://python.usyiyi.cn/
  - https://docs.scipy.org/doc/
  - http://pandas.pydata.org/pandas-docs/stable/index.html


北风网
IBEIFENG.COM