法律声明

AI实战工程师

数学基础

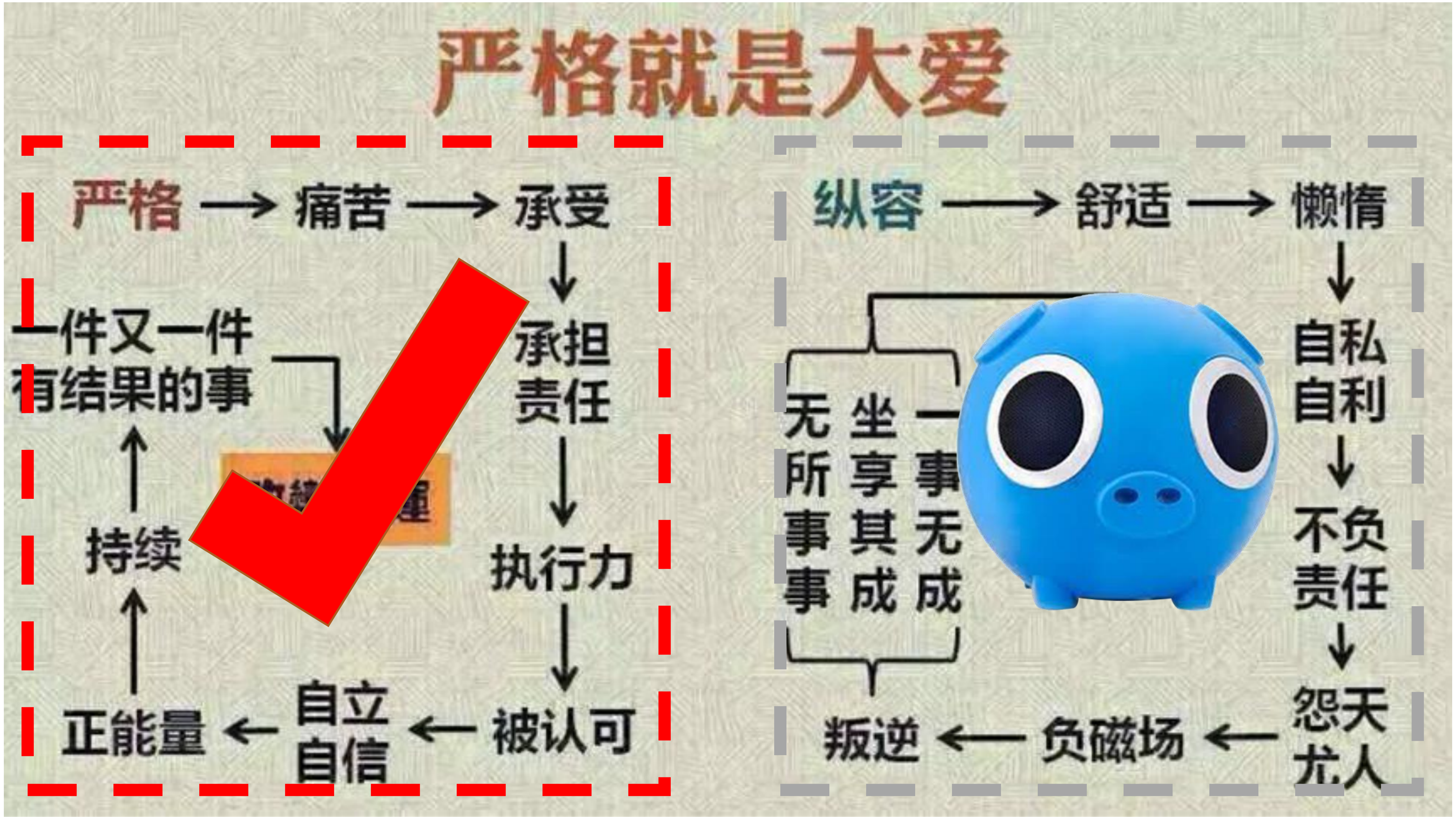
主讲人：June

上海育创网络科技有限公司

课程要求

- 课上课下“九字”真言
  - 认真听，善摘录，勤思考
  - **多温故，乐实践**，再发散
- 四不原则
  - 不懒散惰性，不迟到早退
  - 不请假旷课，不拖延作业

严格是大爱


严格就是大爱
严格
痛苦
承受
承担
责任
执行力
被认可
自立
自信
正能量
持续
一件又一件
有结果的事
纵容
舒适
懒惰
自私
自利
不负
责任
怨天
尤人
负磁场
叛逆
无所事事
坐享其成
一事无成

寄语

做别人不愿做的事，

做别人不敢做的事，

做别人做不到的事。

知识要点

1. 数学基础
   1) 数据分析
   2) 概率论
   3) 线性代数
   4) 矩阵向量

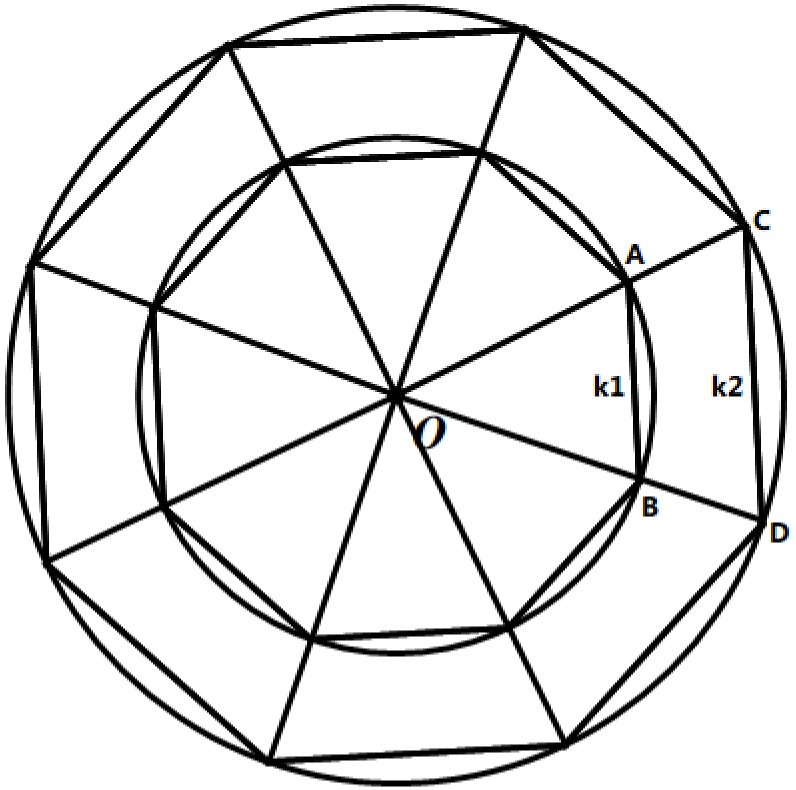
常数π

- **π(圆周率)**：圆的周长与直径的比值，是数学及物理学上普遍存在的一个数学常数，名称：pi(派)；π是一个无限不循环小数，通常使用3.1415代表圆周率来进行近似计算。

$$\pi = \frac{\mathrm{C}}{\mathrm{d}}$$

$$\frac{\pi}{4} = \lim_{n\to\infty}\sum_{i=1}^{n}(-1)^i\frac{1}{2i+1} \approx 0.785398$$

$$\pi = 3.141592$$



自然常数e

- **e(自然常数)**：简单来讲e表示的是**增长极限**，也就是单位时间内，持续的翻倍增长所能达到的极限值。e是自然对数的底数，是一个无理数，约等于2.718281828...
- 令： $f(x)=\log_a(x)$
- 则： $$\frac{f(x+\Delta x)-f(x)}{\Delta x}=\frac{\log_a(x+\Delta x)-\log_a(x)}{\Delta x}=\frac{\log_a\left(\frac{x+\Delta x}{x}\right)}{\Delta x}=\log_a\left(\frac{x+\Delta x}{x}\right)^{\frac{1}{\Delta x}}$$
$$\xrightarrow{令x=1}\log_a(1+\Delta x)^{\frac{1}{\Delta x}}\overset{令}{=\!=\!=}1\Rightarrow\lim_{\Delta x\to 0}(1+\Delta x)^{\frac{1}{\Delta x}}=a$$
- 求解： $\lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n=?$

自然常数e

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n=?$$

$$x_n=\left(1+\frac{1}{n}\right)^n$$

$$=1+C_n^1\frac{1}{n}+C_n^2\frac{1}{n^2}+C_n^3\frac{1}{n^3}+....+C_n^{n-1}\frac{1}{n^{n-1}}+\frac{1}{n^n}$$

$$=1+n\cdot\frac{1}{n}+\frac{n(n-1)}{2!}\cdot\frac{1}{n^2}+\frac{n(n-1)(n-2)}{3!}\cdot\frac{1}{n^3}+...+\frac{n(n-1)(n-2)...2}{(n-1)!}\frac{1}{n^{n-1}}+\frac{n(n-1)(n-2)...1}{\text{n!}}\frac{1}{n^n}$$

$$=1+1+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{n-1}{n}\right)$$

$$<1+1+\frac{1}{2!}+\frac{1}{3!}+...+\frac{1}{n!}$$

$$<1+1+\frac{1}{2}+\frac{1}{2^2}+...+\frac{1}{2^{n-1}}$$

$$<3-\frac{1}{2^{n-1}}$$

$$<3$$

自然常数e

- 假设某种类的单细胞生物每24小时分裂一次，在不考虑死亡与变异的情况下，那么很显然这群单细胞生物的总数量每天会增加一倍，也就是增长率为:

$$growth = (1 + 100\%) = 2$$

- 假设这种细胞每过12小时，平均会产生一半的原数量的新细胞，而且新细胞在之后的12小时已经存在分裂了，那么增长率为：

$$growth = \left(1 + \frac{100\%}{2}\right)^2 = 2.25$$

- 假设每过8小时分裂一次，那么增长率为：$growth = \left(1 + \frac{100\%}{3}\right)^3 = 2.37037...$

自然常数e

- 实际上，细胞的分裂是不间断的、连续的，每分每钟产生的新细胞，而且每个新细胞都会立即和母体细胞一起进行分裂操作，那么一天24小时的分裂的增长率为：

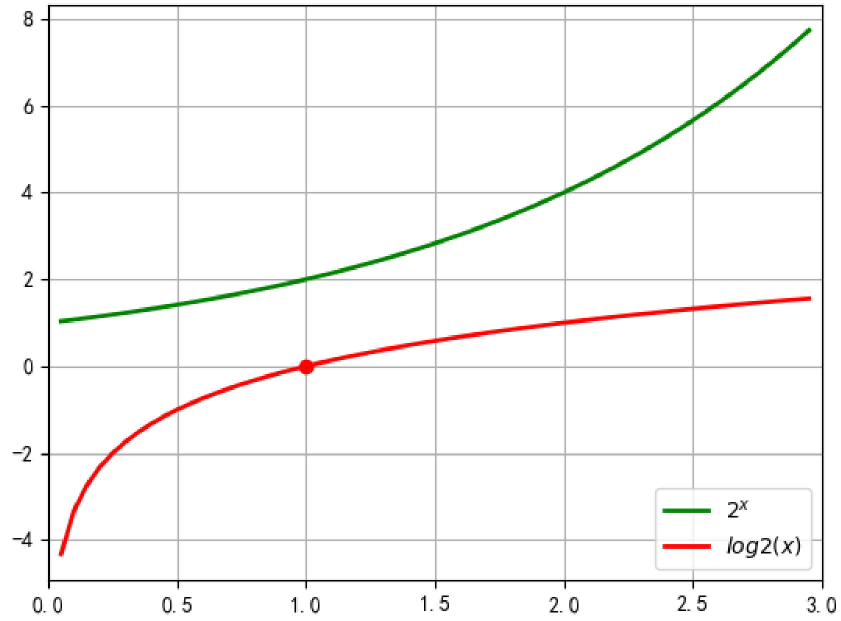
$$growth = \left(1 + \frac{100\%}{n}\right)^n = 2.71828\ldots$$

- 因此将该值称为e，表示单位时间内，持续翻倍增长所能达到的极限值，公式为：

$$e = \lim_{n \to \infty}\left(1 + \frac{1}{n}\right)^n$$

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： $y = a^x, a > 0 且 a \neq 1$



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

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

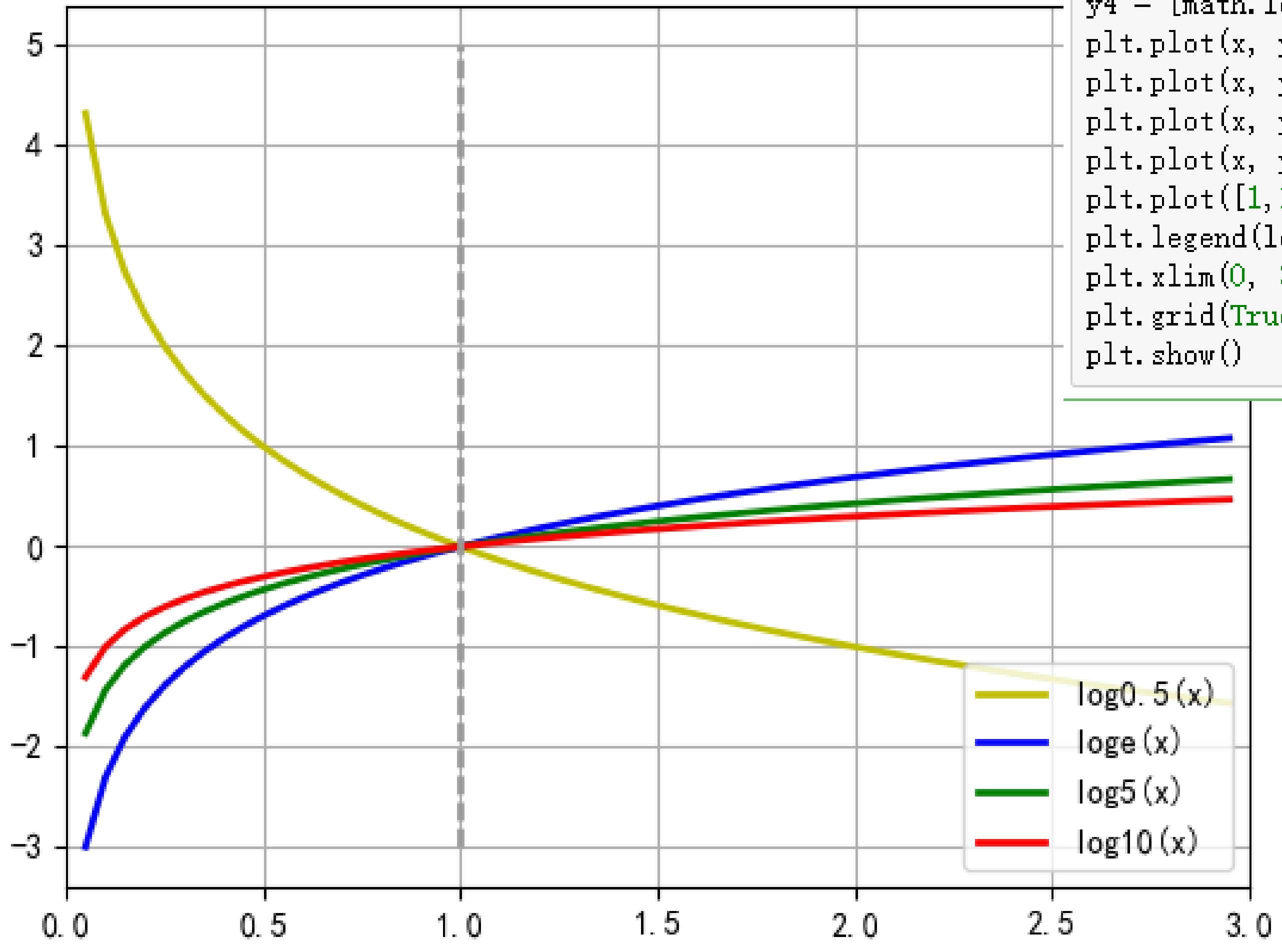
$$\log_a(1) = 0$$

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

$$a^0 = 1$$

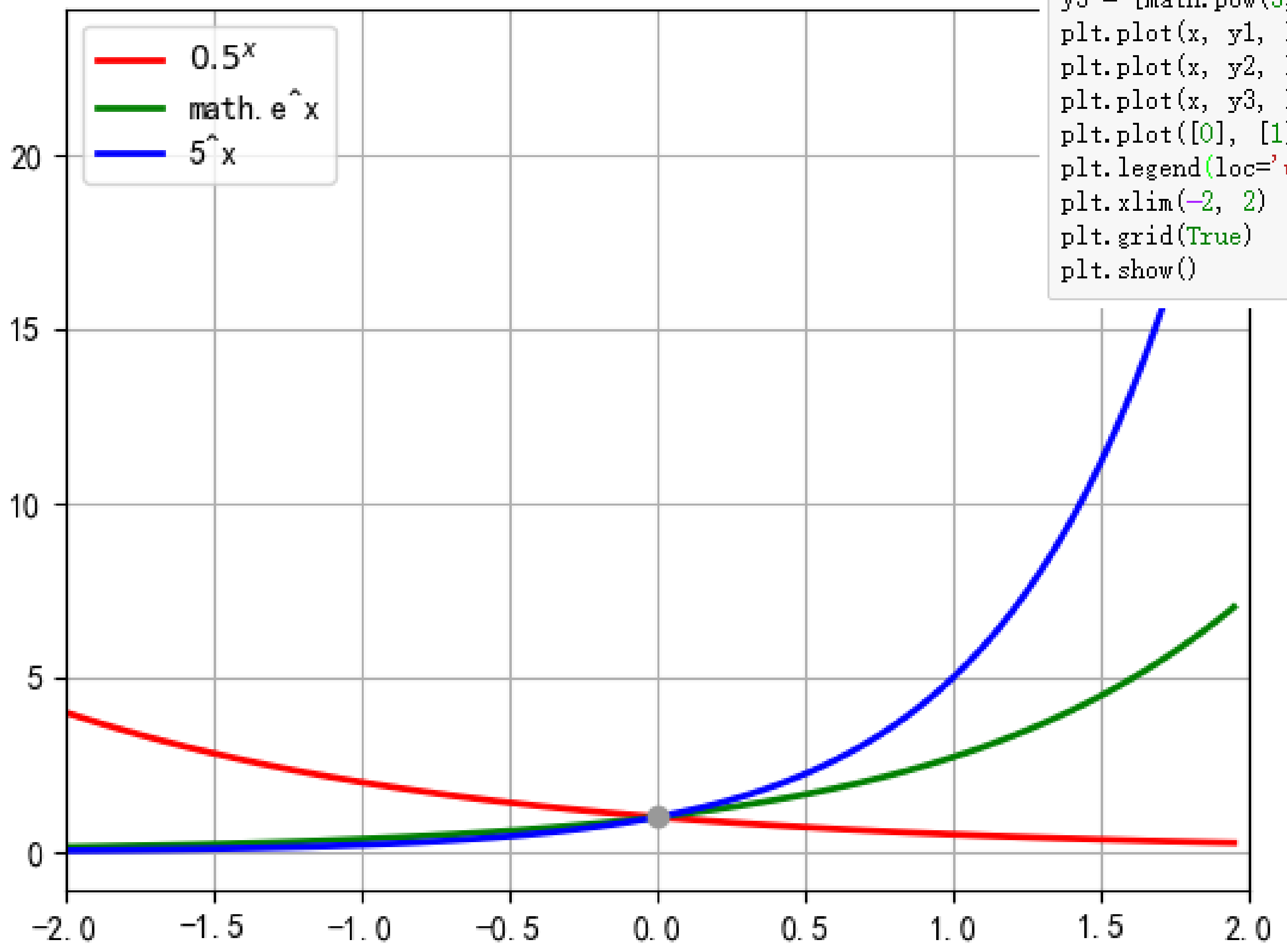
对数函数、指数函数

```
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

- **导数**就是曲线的斜率，是曲线变化快慢的反应。
- **二阶导数**是斜率变化的反应，表征曲线是**凹凸性**。
- 导数(Derivative)：当函数y=f(x)的自变量x在一点x_0上产生一个增量Δx时候，函数输出值的增量Δy与自变量的增量Δx的比值当Δx趋近于0时的极限值趋于a时，即a为函数f在x_0处的导数，记作$f'(x_0)$或$df(x_0)/dx$
- 导数是函数的局部性质，一个函数在某个点的导数描述了这个函数在这个点附近的变化率，也就是函数所代表的曲线在这个点的**切线斜率**。
- 可导的函数一定连续，不连续的函数一定不可导。

导数

$$y = f(x)$$

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

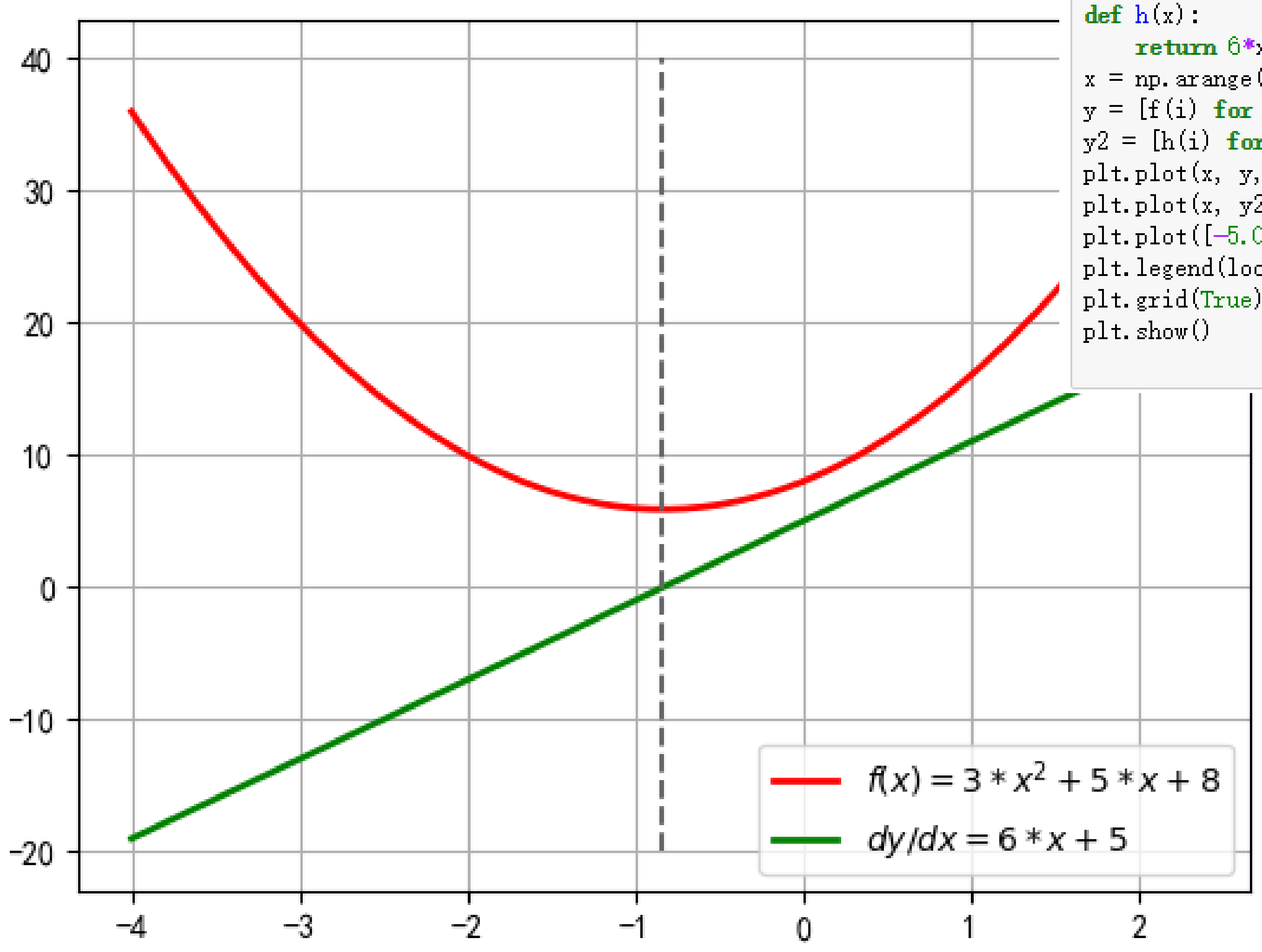
导数的性质：单调性

- 通过函数的导数的值，可以判断出函数的单调性、驻点以及极值点：
  - 若导数大于0，则**单调递增**；若导数小于0，则**单调递减**；导数等于零为函数**驻点**。
  - 如果函数的导函数在某一个区间内恒大于零(或恒小于零)，那么函数在这一个区间单调递增(或单调递减)，这种区间就叫做**单调区间**。
  - 导函数值等于零的点称为函数的**驻点**，在这类点上函数有可能取得极大值或者极小值(即极值可疑点)；对于极值点的判断需要判断驻点附近的导函数的值符号，如果存在使得之前区间上导函数值都大于零，而之后的区间上都小于等于零，那么这个点就是一个**极大值点**，反之则是一个**极小值点**。

导数的性质：单调性

```
def f(x):
    return 3 * x * x + 5 * x + 8

def h(x):
    return 6*x + 5
x = np.arange(-4, 2.4, 0.05)
y = [f(i) for i in x]
y2 = [h(i) for i in x]
plt.plot(x, y, linewidth=2, color='r', label='$f(x)=3*x^2 + 5*x+8$')
plt.plot(x, y2, linewidth=2, color='g', label='$dy/dx=6*x+5$')
plt.plot([-5.0/6, -5.0/6], [-20, 40], '--', color='#666666')
plt.legend(loc='lower right')
plt.grid(True)
plt.show()
```



凸集

- 集合中任意两点连线的点都在该集合中，则称该集合为凸集，否则为非凸集(凹集)。

Definition 2.1 *A set C is convex if, for any $x, y \in C$ and $\theta \in \mathbb{R}$ with $0 \le \theta \le 1$,*

$$\theta x + (1-\theta)y \in C.$$

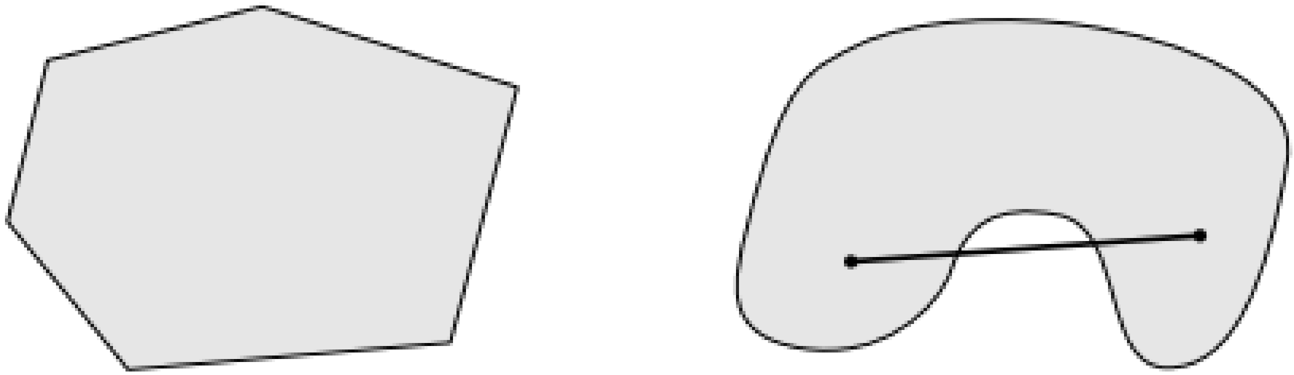
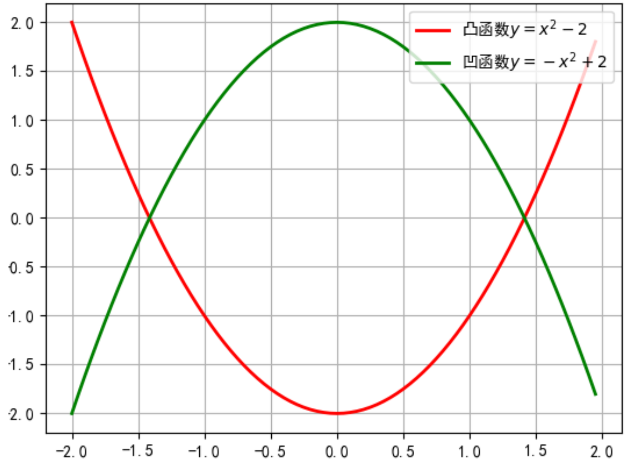


Figure 1: Examples of a convex set (a) and a non-convex set (b).

导数的性质：凹凸性

- 可导函数的凹凸性与其导数的单调性有关，如果函数的导函数在某个区间上单调递增，那么在这个区间上函数是向下凸的，反之则是向上凹的。如果二阶导函数存在，也可以用二阶导函数值的正负性来判断，如果在一个区间上恒大于零，那么在这个区间是上函数是**凸**的，反之函数是**凹**的；曲线的凹凸分界点称为曲线的**拐点**。



凸函数

- 若函数f的定义域C为凸集，而且满足条件：$\forall x, y \in C, 0 \le \theta \le 1,$

  且有：$f(\theta x+(1-\theta)y) \le \theta f(x)+(1-\theta)f(y)$

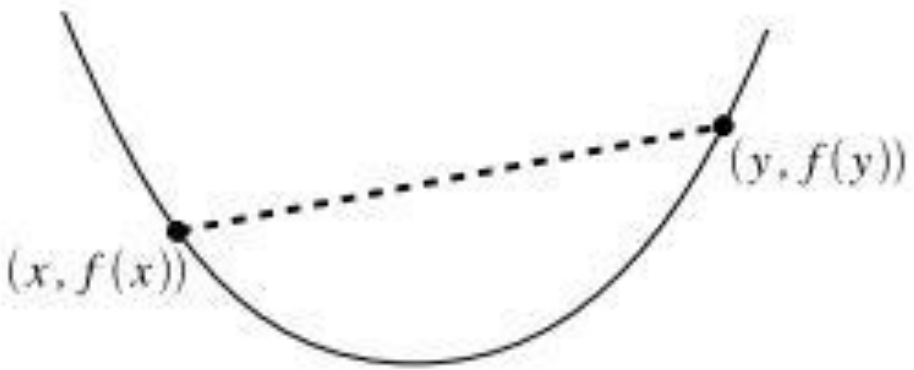
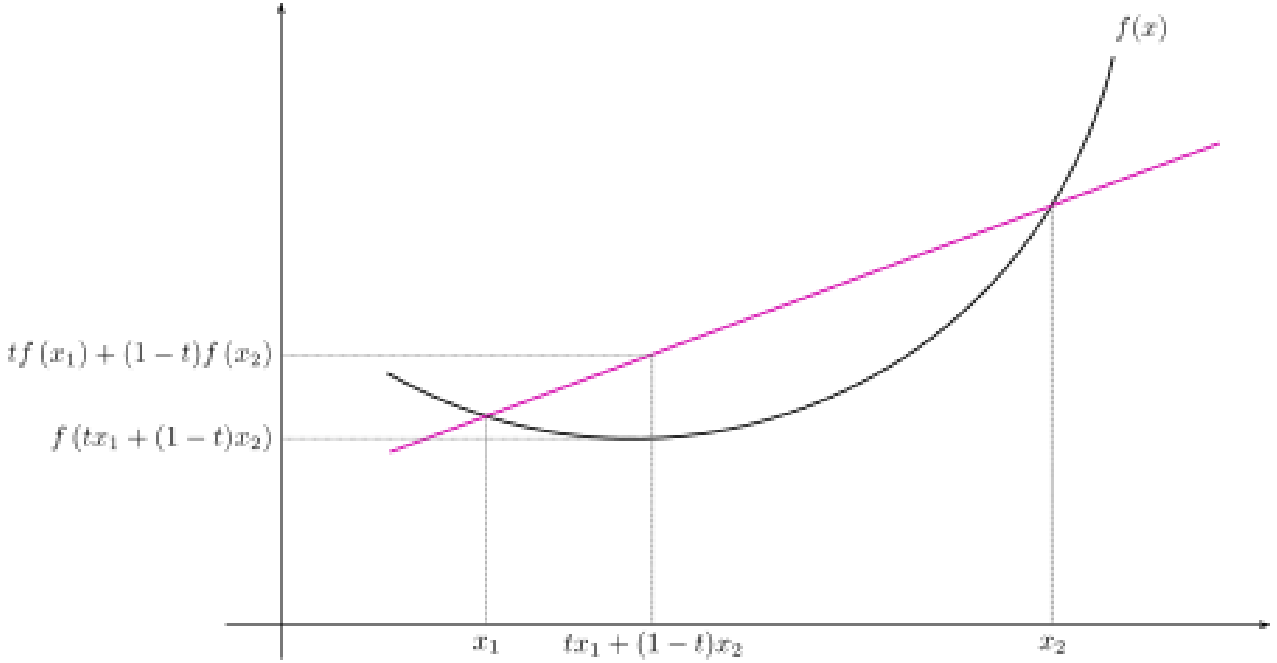


Figure 2: Graph of a convex function. By the definition of convex functions, the line connecting two points on the graph must lie above the function.

凸函数

- 凸函数(Convex Function)：对曲线上的任意两点的连线上的点对应的函数值不大于该两点对应函数值连线上的值。



凸函数

- 凸函数的一阶充要条件为，函数的一阶导数是递增函数，公式为：

$$f(y) \geq f(x) + \nabla_x f'(x)(y-x)$$

- 凸函数的二阶充要条件为，函数的二阶导数值大于0，公式为：

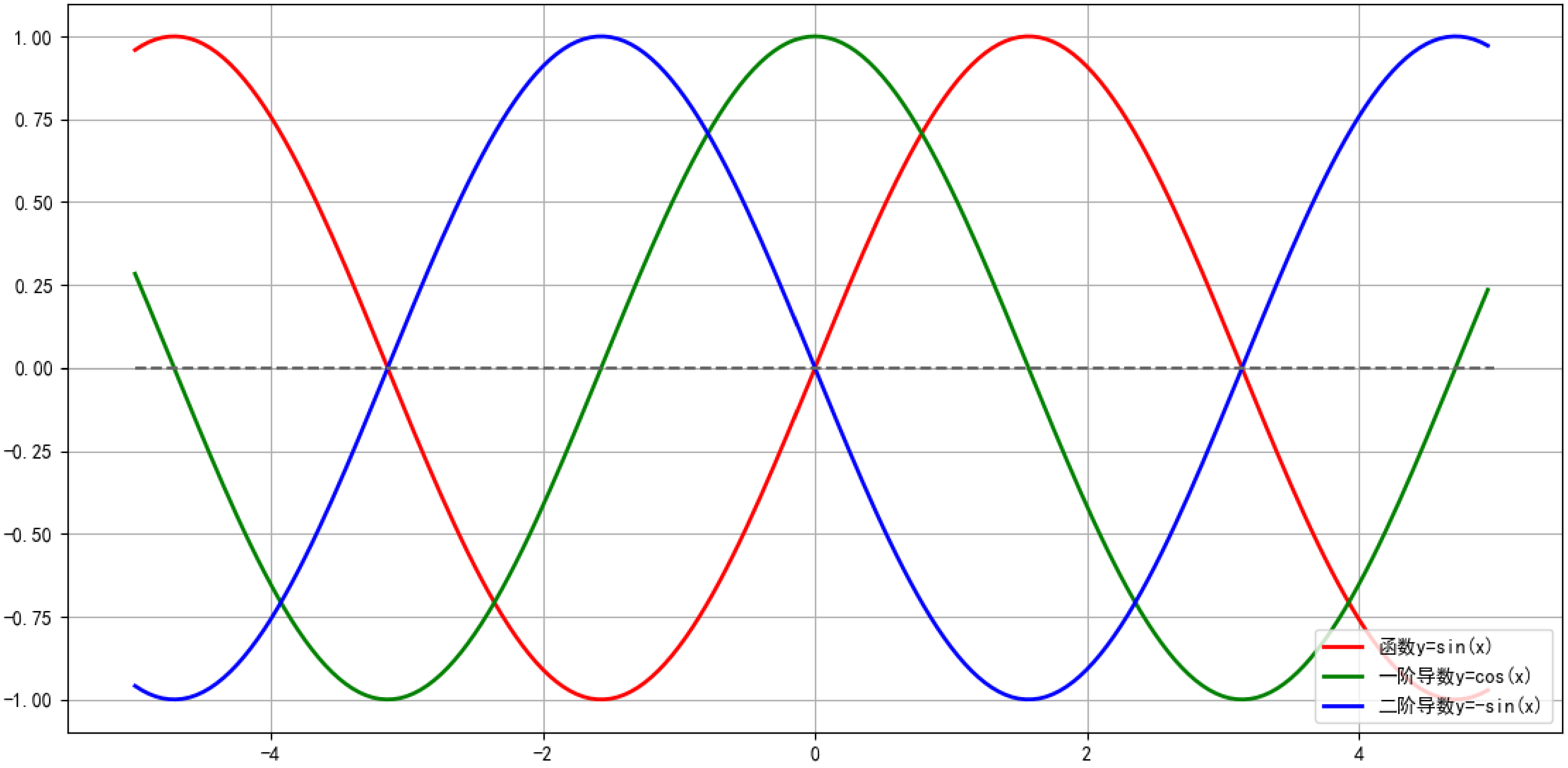
$$f''(x) \geq 0$$

- **注意：对于凸函数问题而言，局部最优解就是全局最优解**

凸函数

- 常见凸函数主要包括：
  - 指数函数 $f(x)=e^{ax}$
  - 幂函数 $f(x)=a^{x}, x\in R^{+}, a\geq 1或a\leq 0$
  - 负对数函数 $f(x)=-\ln x$
  - 负熵函数 $f(x)=x\ln x$
  - 二次函数 $f(x)=x^{2}$
  - 范数函数 $f(\vec{x})=\|\vec{x}\|$

导数的性质：凹凸性


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−4
−2
0
2
4
函数y=sin(x)
一阶导数y=cos(x)
二阶导数y=-sin(x)

常见导函数

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

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

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

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

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

常见导函数

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

可知$y>0$

$$y = x^n \quad \Rightarrow \quad \ln y = \ln x^n$$

令$t = \ln y, h = \ln x^n$

则$t' = h'$

$$t' = \frac{y'}{y}, h' = \frac{1}{x^n} \cdot n \cdot x^{n-1} = \frac{n}{x}$$

$$\Rightarrow y' = y \cdot \frac{n}{x} = x^n \cdot \frac{n}{x} = n \cdot x^{n-1}$$

导数的应用1

- 已知函数 $f(x)=x^x, x>0$
- 求f(x)的最小值

$$y=x^x$$

$$\to \ln y=\ln x^x \Rightarrow \ln y = x\ln x$$

两边同时对x求导

$$\to \quad \frac{y'}{y}=\ln x+1$$

令y′=0

$$\to \ln x+1=0$$

$$\to x=e^{-1}$$

$$\to t=e^{-\frac{1}{e}}$$

偏导数

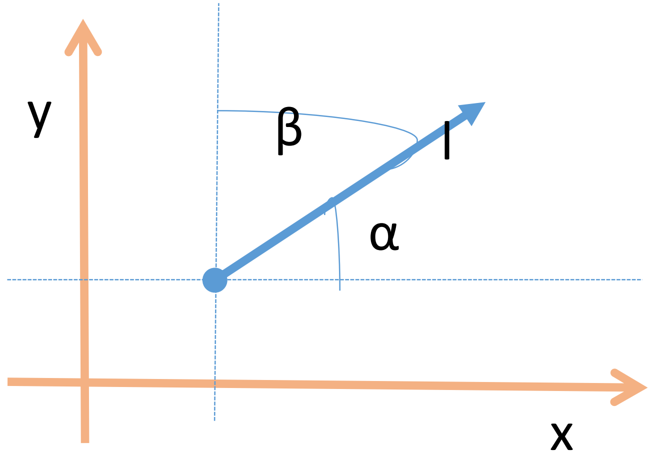
- 在一个多变量的函数中，偏导数就是关于其中一个变量的导数而保持其它变量恒定不变。假定二元函数z=f(x,y)，点(x_0,y_0)是其定义域内的一个点，将y固定在y_0上，而x在x_0上增量Δx，相应的函数z有增量Δz=f(x_0+Δx, y) - f(x_0,y_0)；Δz和Δx的比值当Δx的值趋近于0的时候，如果极限存在，那么此极限值称为函数z=f(x,y)在(x_0,y_0)处对x的偏导数(partial derivative)，记作：f'x(x_0,y_0)

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

方向导数

- **方向导数(directional derivative)**：不仅仅需要知道函数在坐标轴上的变化率(即偏导数)，而且还需要设法求得函数在其他特定方向上的变化率；而方向导数就是函数在其它特定方向上的变化率。
- 如果函数z=f(x,y)在点P(x,y)是可微分的(即可导的)，那么，函数在该点沿着任意一方向L的方向可微(即可导)都存在，且公式为：

$$\frac{\partial f}{\partial l}=\frac{\partial f}{\partial x}\cos\alpha+\frac{\partial f}{\partial y}\cos\beta$$



梯度

- 设函数z=f(x, y)在平面区域D内具有一阶连续偏导数，则对于每一个点P(x,y)∈D，向量：

$$\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right)$$

为函数z=f(x,y)在点P的梯度，记作gradf(x,y)

- 梯度的本意是一个向量(矢量)，表示某一函数在该点处的方向导数沿着该方向取得最大值，即函数在该点处沿着该方向变化最快，变化率最大。

梯度下降

- 梯度下降法(gradient descent)是一种最优化算法，通常被称为最速下降法，是解决无约束优化问题的最简单和最古老的方法之一，常用于机器学习算法和人工智能中递归的逼近最小偏差模型；最速下降法是用负梯度方向为搜索方向，越接近目标值，步长越小，前进越慢。
- 梯度下降法迭代公式为： $x_{k+1} = x_k + \alpha_k \bar{s}_k$

x为需要求解的值，s为梯度负方向，α为步长

- 缺点：靠近极小值的时候收敛速度比较慢；可能会”之字形”的下降；不太适合处理比较复杂的非线性函数问题。

梯度下降 $z = f(x,y) = x^2 + y^2$

```
## 原函数
def f(x, y):
    return x ** 2 + y ** 2
## 偏函数
def h(t):
    return 2 * t

X = []
Y = []
Z = []

x = 2
y = 2
f_change = x ** 2 + y ** 2
f_current = f(x, y)
step = 0.1
X.append(x)
Y.append(y)
Z.append(f_current)
while f_change > 1e-10:
    x = x - step * h(x)
    y = y - step * h(y)
    f_change = f_current - f(x, y)
    f_current = f(x, y)
    X.append(x)
    Y.append(y)
    Z.append(f_current)
print u"最终结果为:", (x, y)
```

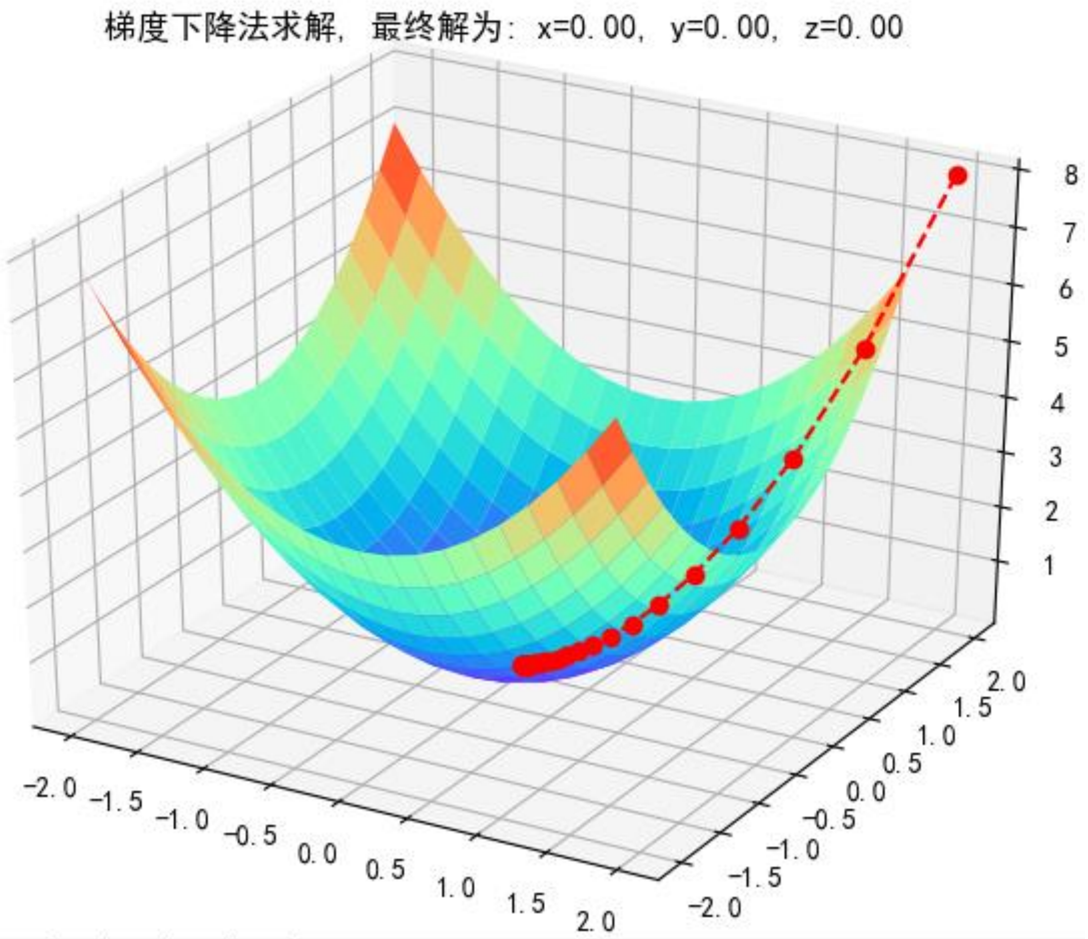
最终结果为：(9.353610478917782e-06, 9

```
fig = plt.figure()
ax = Axes3D(fig)
X2 = np.arange(-2, 2, 0.2)
Y2 = np.arange(-2, 2, 0.2)
X2, Y2 = np.meshgrid(X2, Y2)
Z2 = X2 ** 2 + Y2 ** 2

ax.plot_surface(X2, Y2, Z2, rstride=1, cstride=1, cmap='rainbow')
ax.plot(X, Y, Z, 'ro--')

ax.set_title(u'梯度下降法求解, 最终解为: x=%.2f, y=%.2f, z=%.2f' % (x, y, f_current))

plt.show()
```



梯度下降 $y = f(x) = x^2$

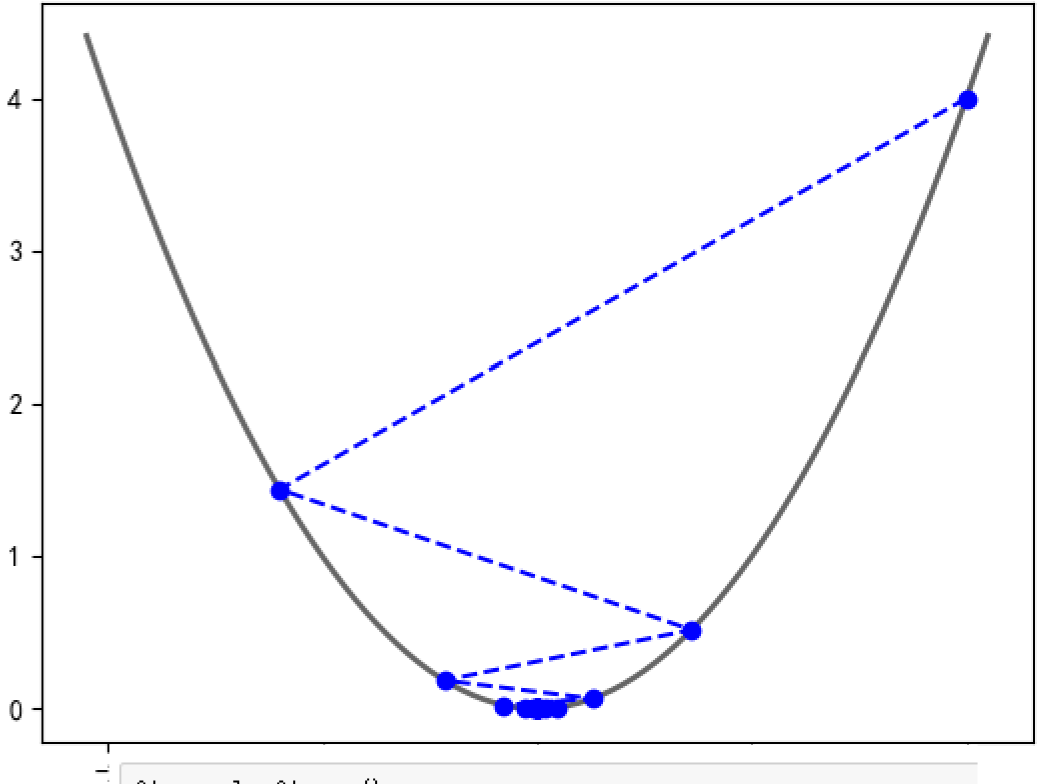
```
## 原函数
def f(x):
    return x ** 2
## 导数
def h(x):
    return 2 * x

X = []
Y = []

x = 2
step = 0.8
f_change = f(x)
f_current = f(x)
X.append(x)
Y.append(f_current)
while f_change > 1e-10:
    x = x - step * h(x)
    tmp = f(x)
    f_change = np.abs(f_current - tmp)
    f_current  = tmp
    X.append(x)
    Y.append(f_current)
print u"最终结果为:", (x, f_current)
```

最终结果为: (-5.686057605985963e-06, 3.233125109859082e-11)

$y = x^2$函数求解最小值，最终解为: x=-0.00, y=0.00




```
fig = plt.figure()
X2 = np.arange(-2.1, 2.15, 0.05)
Y2 = X2 ** 2

plt.plot(X2, Y2, '-', color='#666666', linewidth=2)
plt.plot(X, Y, 'bo--')
plt.title(u'$y=x^2$函数求解最小值，最终解为:x=%.2f,y=%.2f'
          % (x, f_current))
plt.show()
```

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：

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

其中，$f^{(n)}(x)$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

Taylor公式

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

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

$$0!=1 \qquad 1!=1 \qquad \mathrm{n}!=n*(n-1)!$$

Taylor公式-余项 $f(x)=\sum_{k=0}^{n}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k+R_n(x)$

- 佩亚诺(Peano)余项： $R_n(x)=o[(x-x_0)^n]$
- 拉格朗日(Lagrange)余项： $R_n(x)=f^{(n+1)}[x_0+\theta(x-x_0)]\frac{(x-x_0)^{n+1}}{(n+1)!}$
- 柯西(Cauchy)余项：$R_n(x)=f^{(n+1)}[x_0+\theta(x-x_0)]\frac{(1-\theta)^n(x-x_0)^{n+1}}{n!}$
- 积分余项：$R_n(x)=\frac{(-1)^n}{n!}\int_a^x(t-x)^n f^{(n+1)}(t)dt$

Taylor公式-拉格朗日余项θ值

$$R_n(x) = f^{(n+1)}(x_0 + \theta\Delta x)\frac{(\Delta x)^{n+1}}{(n+1)!}$$

- 在n阶Taylor公式中，如果余项为拉格朗日余项，那么当$f^{(n+2)}(x)$在x_0点连续，而且$f^{(n+2)}(x)y \neq 0$，有：

$$\lim_{\Delta x \to 0} \theta = \frac{1}{n+2}$$

Taylor公式-拉格朗日余项θ值

$$f_1(x)=\sum_{k=0}^{n}\frac{f^{(k)}(x_0)}{k!}(\Delta x)^k+f^{(n+1)}(x_0+\theta\Delta x)\frac{(\Delta x)^{n+1}}{(n+1)!}$$

$$f_2(x)=\sum_{k=0}^{n}\frac{f^{(k)}(x_0)}{k!}(\Delta x)^k+\frac{f^{(n+1)}(x_0)}{(n+1)!}(\Delta x)^{n+1}+f^{(n+2)}(x_0+\theta_1\Delta x)\frac{(\Delta x)^{n+2}}{(n+2)!}$$

$$f_2=f_1\Rightarrow f^{(n+1)}(x_0+\theta\Delta x)-f^{(n+1)}(x_0)=\frac{1}{n+2}f^{(n+2)}(x_0+\theta_1\Delta x)\Delta x$$

$$f^{(n+1)}(x_0+\theta\Delta x)-f^{(n+1)}(x_0)=f^{(n+2)}(x_0+\theta_2\theta\Delta x)(\theta\Delta x)$$

$$\theta=\frac{f^{(n+2)}(x_0+\theta_1\Delta x)}{(n+2)\cdot f^{(n+2)}(x_0+\theta_2\theta\Delta x)}\quad\xRightarrow{f^{(n+2)}(x_0)\neq 0}\quad\lim_{\Delta x\to 0}\theta=\frac{1}{n+2}$$

Taylor公式应用1

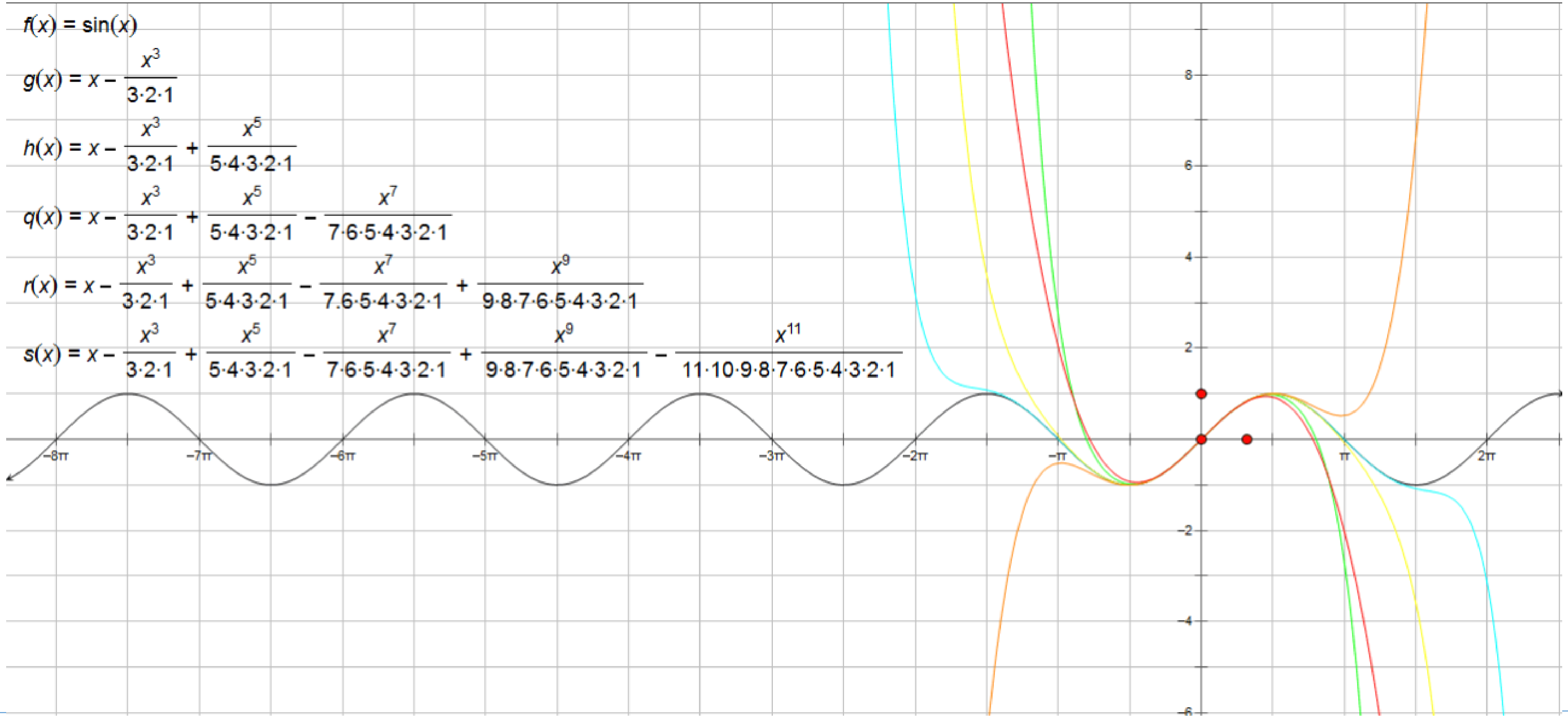
- 展开三角函数y=sin(x)（使用麦克劳林公式进行展开操作）

$$f^{(n)}(x)=\sin\left(x+\frac{n\pi}{2}\right)$$

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

Taylor公式应用1

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



Taylor公式应用2

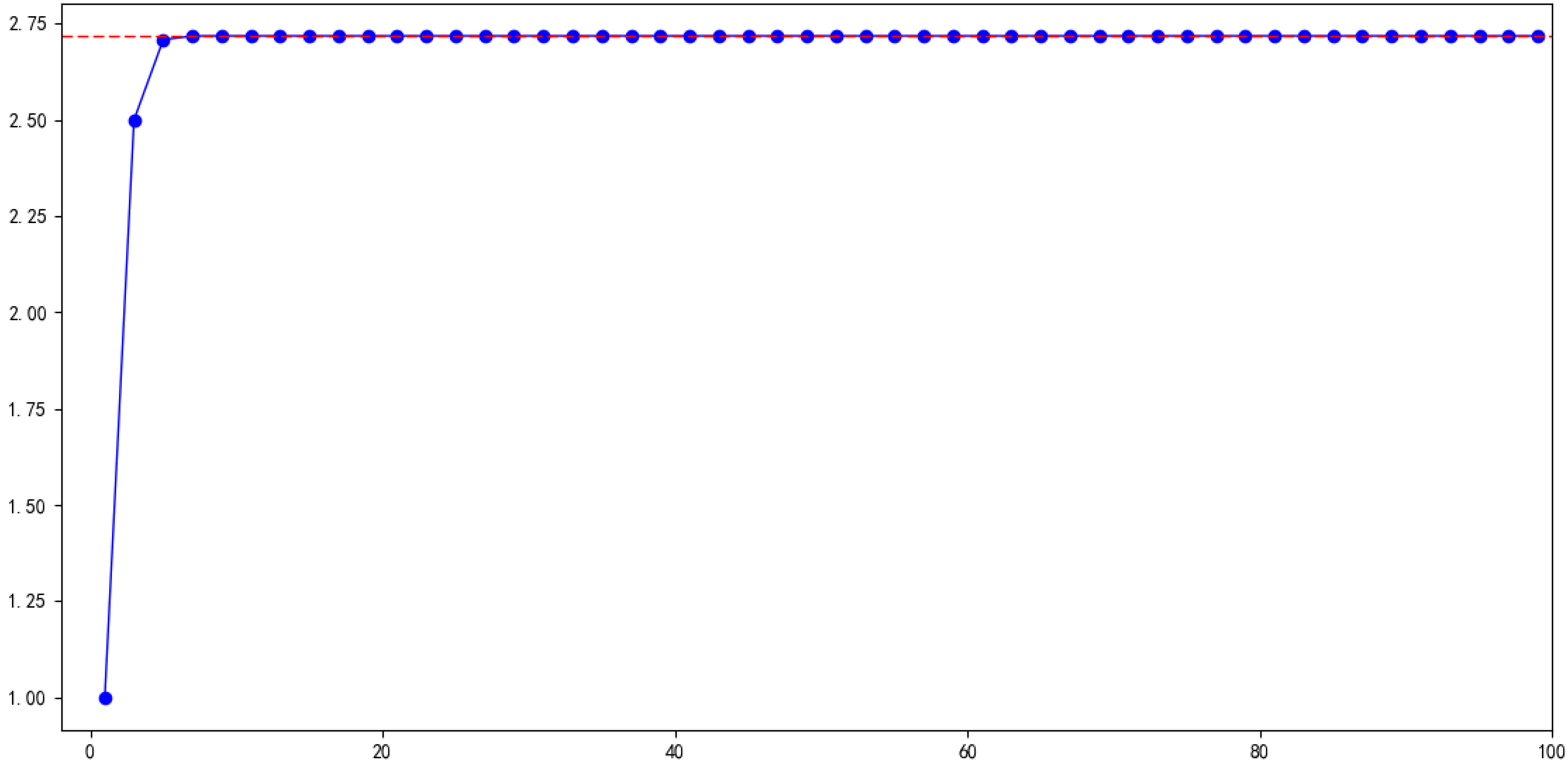
- 计算近似值 $e=\lim_{x\to\infty}\left(1+\frac{1}{n}\right)^n$，并估计误差值 $y=e^x \Rightarrow y'=y=e^x$

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

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

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

Taylor公式应用2



组合数

- 从m个不同元素中取出n(n ≤ m)个元素的所有组合的个数，叫做从m个不同元素中取出n个元素的组合数。记作:C(m,n)

$$C(m,n)=\frac{m!}{(m-n)!\cdot n!}$$

- 将m个不同的元素分为k组，每组元素数量为$m_1,m_2,\ldots m_k$;($m=m_1+m_2+\ldots+m_k$)，则不同的分组方式有:

$$C=C(m,m_1)\cdot C(m-m_1,m_2)\ldots C(m-m_1-m_2-\ldots-m_{k-1},m_k)$$

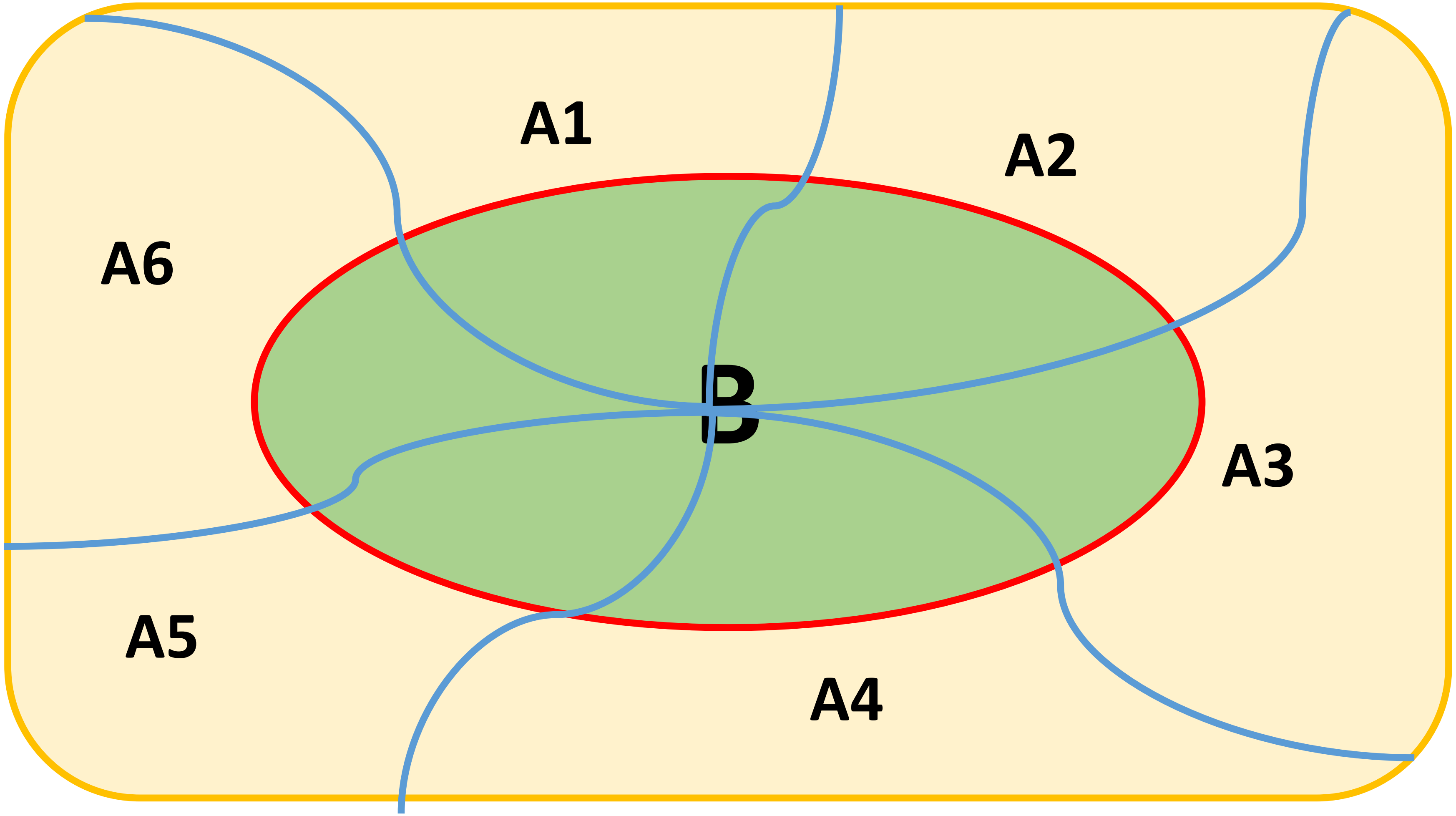
$$\Rightarrow C=\prod_{j=1}^{k}C\left(m-\sum_{i=1}^{j-1}m_i,m_j\right)\Rightarrow C=\frac{m!}{m_1!\cdot m_2!\ldots m_k!}$$

古典概率

- 概率是以假设为基础的，即假定随机现象所发生的事件是有限的、互不相容的，而且每个基本事件发生的可能性相等。一般来讲，如果在全部可能出现的基本事件范围内构成事件A的基本事件有a个，不构成事件A的有b个，那么事件A出现的概率为： $P(A)=\dfrac{a}{a+b}$
- 例如：投掷一枚平正的硬币，正面朝上与反面朝上是唯一可能出现的两个基本事件(不考虑可能出现竖着的情况)，且互不相容，将出现正面的事件记作A，那么事件A的概率:

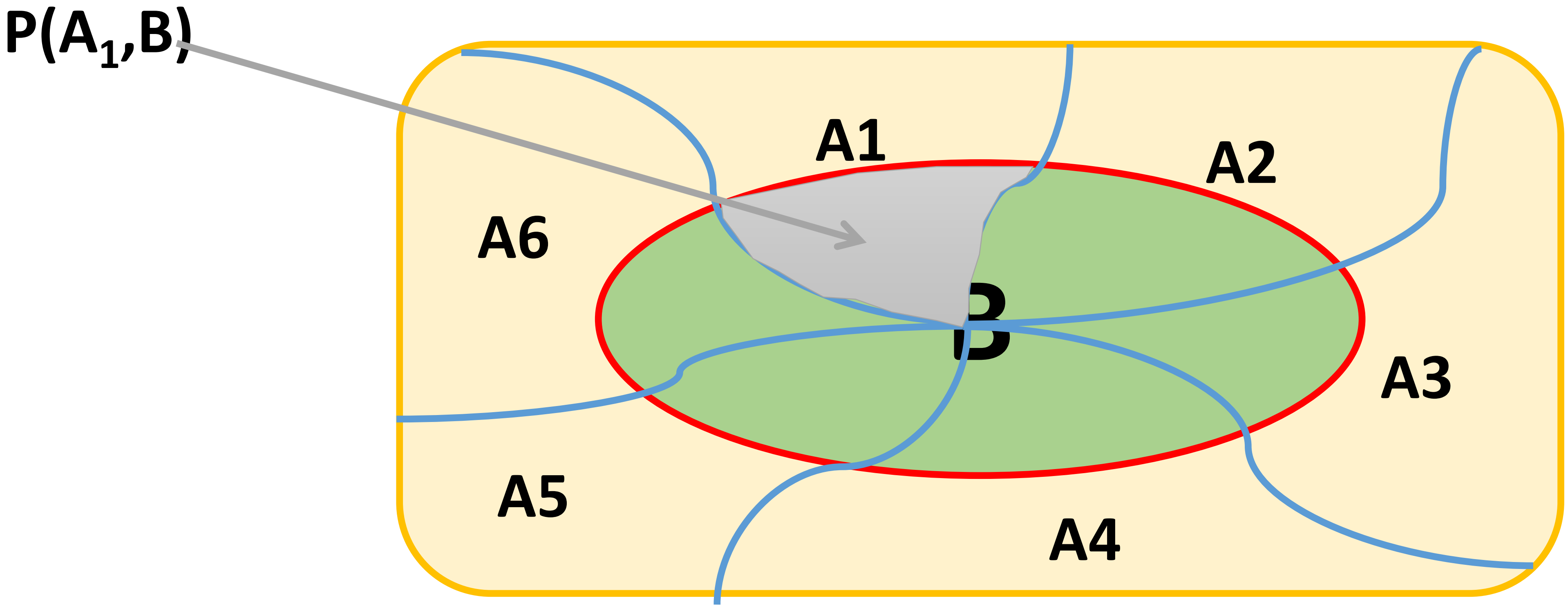
$$P(A)=\frac{1}{1+1}=\frac{1}{2}$$

概率公式



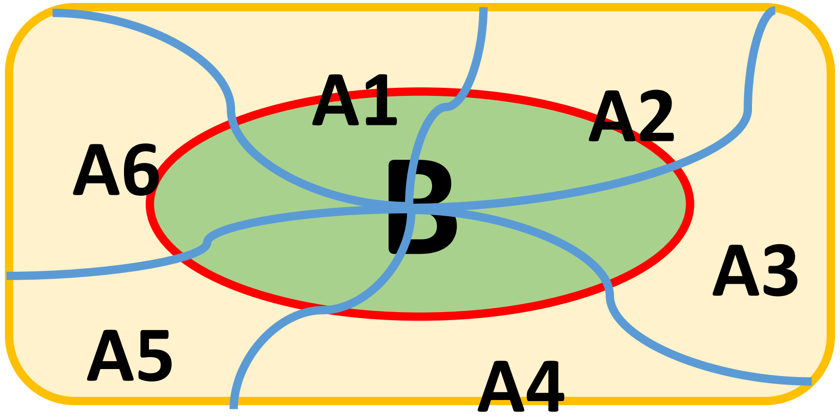
联合概率

- 表示两个事件共同发生的概率，事件A和事件B的共同概率记作：P(AB)、P(A,B)或者P(A∩B)，读作“事件A和事件B同时发生的概率”



条件概率

- 事件A在另外一个事件B已经发生的条件下的发生概率叫做条件概率，表示为P(A|B)，读作“在B条件下A发生的概率”
- 一般情况下P(A|B)≠P(A)，而且条件概率具有三个特性：
  - 非负性
  - 可列性
  - 可加性



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

条件概率

- 将条件概率公式由两个事件推广到任意有穷多个事件时，可以得到如下公式，假设A_1，A_2，....，A_n为n个任意事件(n≥2)，而且$P(A_1A_2...A_n)>0$，则：

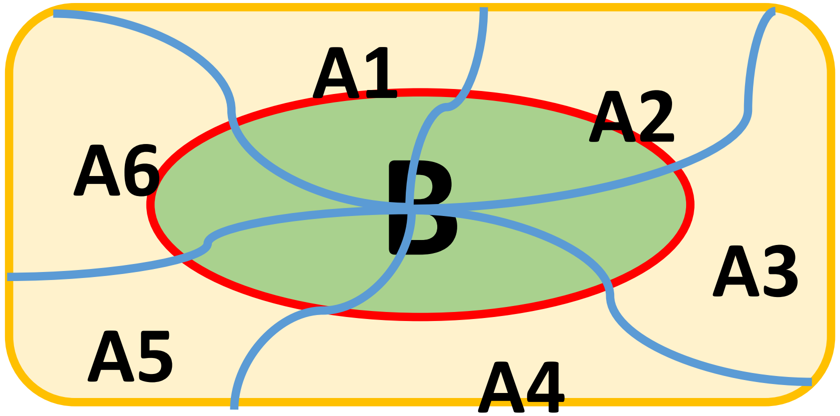
$$P(A_1A_2\ldots A_n)=P(A_1)P(A_2|A_1)\ldots P(A_n|A_1A_2\ldots A_{n-1})$$

全概率公式

- 样本空间Ω有一组事件B_1、$B_2 \ldots B_n$, 如果事件组满足下列两个条件，那么事件组称为样本空间的一个划分。

$$\forall i \neq j \in \{1,2\ldots,n\}, B_i B_j = \phi$$

$$B_1 \cup B_2 \ldots \cup B_n = \Omega$$

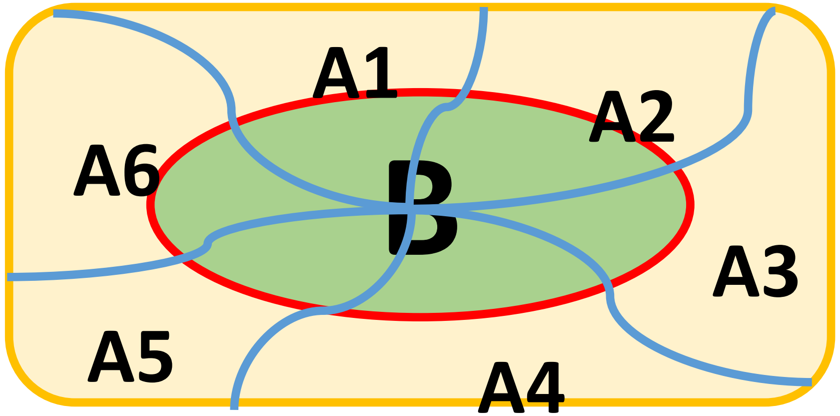


- 设事件$\{B_j\}$是样本空间Ω的一个划分，且$P(B_i)>0$，那么对于任意事件A，全概率公式为：$P(A)=\sum_{i=1}^{n} P(B_i)P(A|B_i)$

贝叶斯公式

- 设B_1、B_2...B_n是样本空间Ω的一个划分，如果对任意事件A而言，有P(A)>0，那么：

$$P(B_i \mid A) = \frac{P(A, B_i)}{P(A)} = \frac{P(B_i) \bullet P(A \mid B_i)}{\sum_{i=1}^{n} P(B_i) \bullet P(A \mid B_i)}$$



贝叶斯公式

- 一座房子在过去20年里一共发生过2次被盗案，房子的主人养了一条狗，狗平均每周晚上叫3次，在盗贼入侵时狗叫的概率估计为0.9，请求：在狗叫的时候发生入侵的概率是多少?

$$P(A)=\frac{3}{7} \qquad P(B)=\frac{2}{20*365}=\frac{2}{7300} \qquad P(A|B)=0.9$$

$$P(B|A)=\frac{P(B)P(A|B)}{P(A)}=\frac{0.9*\frac{2}{7300}}{\frac{3}{7}}=\frac{21}{36500}\approx 0.00058$$

贝叶斯公式 $P(A|B)=\frac{P(B|A)P(A)}{P(B)}$

- P(A)：在没有数据支持下，A发生的概率：**先验概率**或**边缘概率**。
- P(A|B)：在已知B发生后A的条件概率，也就是由于得自B的取值而被称为A的**后验概率**。
- P(B|A)：在已知A发生的情况下的概率分布：似然函数。

事件的独立性

- 给定A、B两个事件，如果概率存在P(A,B)=P(A)P(B)<A、B事件的联合分布概率等于事件各自分布概率的积>，则事件A和B相互独立。
- 如果事件A、B相互独立，互不影响，那么存在P(A|B)=P(A)，P(B|A)=P(B)

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。

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

- 离散型 $E(X)=\sum_{i} x_i p_i$

期望

- 假设C为一个常数，X和Y实两个随机变量，那么期望有一下性质：

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

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

如果X和Y相互独立，那么$E(XY)=E(X)E(Y)$

如果$E(XY)=E(X)E(Y)$,那么X和Y不相关

期望

- 甲乙两个人赌博，假设两人获胜的概率相等，比赛规则是先胜三局者为赢家，可以获得100元的奖励。当比赛进行了三局的时候，其中甲胜了两局，乙胜了一局，这个时候由于某些原因中止了比赛，请问如何分配这100元才比较公平？

$$P(\text{甲})=P(\text{乙})=\frac{1}{2}$$

$$P(\text{甲}=\text{赢})=P(\text{赢}|4)+P(\text{赢}|5,\ \text{输}|4)=\frac{1}{2}+\frac{1}{2}*\left(1-\frac{1}{2}\right)=\frac{3}{4}$$

$$P(\text{乙}=\text{赢})=P(\text{赢}|5,\ \text{赢}|4)=\frac{1}{2}*\frac{1}{2}=\frac{1}{4}$$

$$E(\text{甲})=100*P(\text{甲}=\text{赢})=100*\frac{3}{4}=75$$

$$E(\text{乙})=100*P(\text{乙}=\text{赢})=100*\frac{1}{4}=25$$

期望

- 某城市有10万个家庭，没有孩子的家庭有1000个，有一个孩子的有9万个，有两个孩子的家庭有6000个，有三个孩子的家庭有3000个，问此城市一个家庭平均有小孩多少个？

X	0	1	2	3
P	0.01	0.9	0.06	0.03

$$E(X)=\sum_{i=0}^{3} x_i p(x_i)=0*0.01+1*0.9+2*0.06+3*0.03=1.11$$

方差

- 方差(variance)实衡量随机变量或一组数据时离散程度的度量，是用来度量随机变量和其数学期望之间的偏离程度。即方差是衡量数据源数据和期望均值相差的度量值。

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

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

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

方差

- 假设C为一个常数，X和Y实两个随机变量，那么方差有一下性质：

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

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

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

如果X和Y不相关，那么$D(X\pm Y)=D(X)+D(Y)$

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。

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

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是统计独立的，那么二者之间的协方差为零。但是如果协方差为零，那么X和Y是不相关的。

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

协方差

- 假设C为一个常数，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

案例1

- 有甲乙两个单位愿意聘用你，而你能够获得的信息如下，请根据工资待遇的差异情况，您选择哪家单位？为什么？

甲单位不同职位月工资X_1元	1200	1400	1600	1800
获取该职位的概率P_1	0.4	0.3	0.2	0.1

乙单位不同职位月工资X_2元	1000	1400	1800	2200
获取该职位的概率P_2	0.4	0.3	0.2	0.1

$$E(X_1)=1400 \qquad E(X_2)=1400$$

$$D(X_1)=40000 \qquad D(X_2)=160000$$

案例2

- 已知随机变量X的分布列如下，分别求E(X)、E(2X+5)、D(X)、σ(X)的值

X	-2	1	3
P	0.16	0.44	0.40

$$E(X)=-2*0.16+1*0.44+3*0.40=1.32$$

$$E(2X+5)=2E(X)+5=2*1.32+5$$

$$D(X)=E(X^2)-(E(X))^2=(-2)^2*0.16+1^2*0.44+3^2*0.40-1.32^2=2.9376$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{2.9376}\approx 1.7139$$

案例3

- 一名射手击中靶心的概率为0.9，如果他在同样的条件下连续的射击10次，求他击中靶心次数的均值和方差

$$P_i=(1-0.9)^{10-i}*0.9^i$$

X	0	1	2	3	4	5	6	7	8	9	10
P	10^{-10}	$9*10^{-10}$	9^2*10^{-10}	9^3*10^{-10}	9^4*10^{-10}	9^5*10^{-10}	9^6*10^{-10}	9^7*10^{-10}	9^8*10^{-10}	9^9*10^{-10}	$9^{10}*10^{-10}$

$$E(X)=\sum_{i=0}^{10}x_i\cdot p(x_i)\approx 3.87 \qquad D(X)=E(X^2)-E^2(X)\approx 23.33$$

案例4

- 甲乙两名射手在同一条件下射击，所得环数X和X的分布如下，请比较这两个射手的射击水平。

X_1	6	7	8	9	10
P_1	0.16	0.14	0.42	0.1	0.18

X_2	6	7	8	9	10
P_2	0.19	0.24	0.12	0.28	0.17

$$E(X_1)=8 \qquad E(X_2)=8$$

$$D(X_1)=1.6 \qquad D(X_2)=1.96$$

案例5

- 一台机器在一天内发生故障的概率为0.1，若这台机器一周5个工作日不发生故障，可获利5万元，发生一次故障仍可获利2.5万元，发生两次故障的利润为0万元，发生3次及3次以上的故障要亏损1万元，问这台机器一周内可获利的均值是多少？

故障次数	0	1	2	3	4	5
概率P	0.9^5	$0.9^4*0.1$	$0.9^3*0.1^2$	$0.9^2*0.1^3$	$0.9*0.1^4$	0.1^5
利润X	5	2.5	0	-1	-1	-1

$$E(X)=\sum_{i=0}^{5}x_i*p(x_i)=5*0.9^5+2.5*0.1*0.9^4+0-\left(0.9^2*0.1^3+0.9*0.1^4+0.1^5\right)=3.115565\text{万}$$

协方差矩阵

- 对于n个随机向量($X_1,X_2,X_3....X_n$), 任意两个元素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。

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

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

Pearson相关系数

- 协方差可以描述X和Y的相关程度，但是协方差的值和X/Y的值采用的是不同的量纲，导致协方差在数值上表现出比较大的差异，因此可以引入相关系数来表示X和Y的相关性。

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$$

绝对值范围	含义
0.8-1.0	极强相关
0.6-0.8	强相关
0.4-0.6	中等程度相关
0.2-0.4	弱相关
0-0.2	极弱相关或无相关

- 当p(X,Y)=0的时候，称X和Y不线性相关。
- Pearson相关系数取值范围为[-1,1]

中心矩、原点矩

- 假设X和Y是随机变量，若$E(X^k)$,k=1,2,.....存在，则称它为X的k阶原点矩，简称k阶矩。
- 若$E\{[X-E(X)]^k\}$，k=1,2...存在，则称它为X的k阶中心矩。
- 若$E\{[X-c]^k\}$，k=1,2...存在，则称它为X关于点c的k阶矩。
- 若$E\{X^kY^p\}$，k、p=1,2...存在，则称它为X和Y的k+p阶混合原点矩。
- 若$E\{[X-E(X)]^k [Y-E(Y)]^p\}$，k、p=1,2...存在，则称它为X和Y的k+p阶混合中心矩。

中心矩、原点矩

- X的数学期望E(X)是X的一阶原点矩。
- X的方差D(X)是X的二阶中心矩。
- X和Y的协方差Cov(X,Y)是X和Y的二阶混合中心矩。

峰度

- 峰度(peakedness; kurtosis)又称峰态系数。表示了概率密度分布曲线在平均值处峰值高低的特征数，直观来讲，峰度反映的是峰部的尖度。
- 样本的峰度是和正态分布相比较而言的统计量，如果峰度值大于三，那么峰的形状比较尖，比正态分布峰要陡峭。反之亦然。
- 峰度计算公式：随机变量的四阶中心矩与方差平方的比值。

$$kurtosis=\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^4}{\left(N-1\right)\bullet s^4}$$

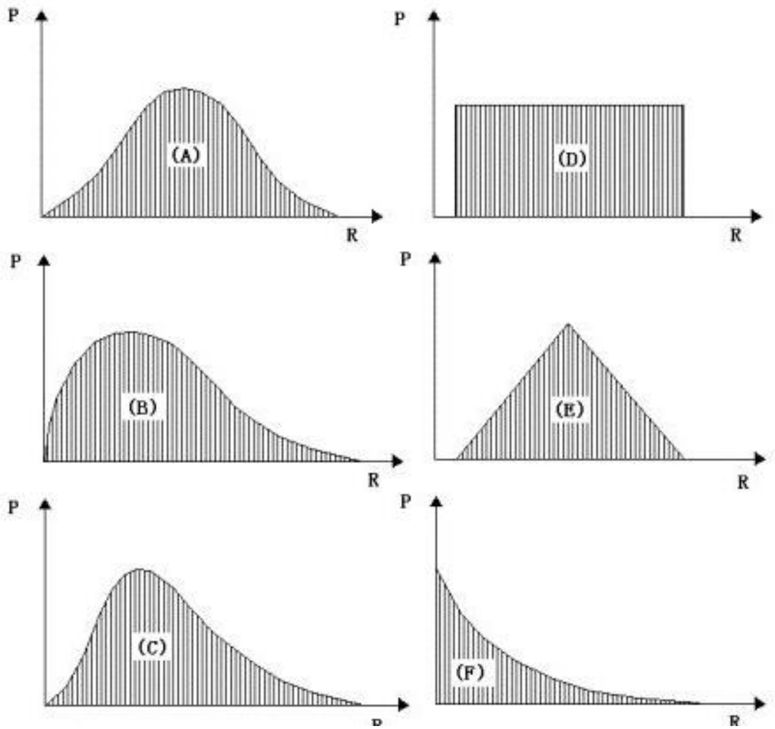
偏度

- 偏度系数(skewness)是描述分布偏离对称性程度的一个特征数，当分布左右对称的时候，偏度系数为0，当偏度系数大于0时候，即重尾在右侧时，该分布为右偏；当偏度系数小于0时候，即重尾在左侧时，该分布为左偏。
- 偏度计算公式：随机变量的三阶中心矩与样本的平均离均差立方和的比值。

$$kurtosis=\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^3}{\left(N-1\right)\bullet s^3}$$

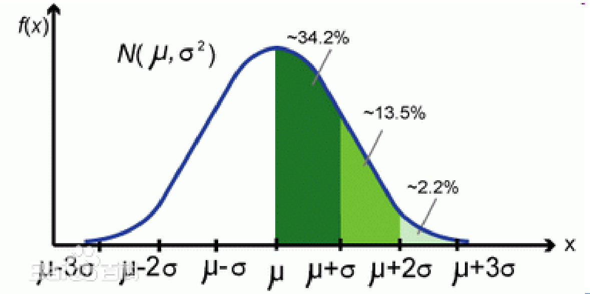
概率密度函数

- 概率密度函数是用于描述一个随机变量的输出值在某个确定取值点附近的可能性的函数，表示瞬时幅度落在某个指定范围内的概率，因此，它的取值范围随幅度而变化。



正态分布

- **正态分布(Normal Distribution)**，又称常态分布、高斯分布(Gaussian Distribution)；若随机变量X服从一个数学期望为μ、方差为σ^2的分布，那么认为随机变量X服从正态分布，记为N(μ,σ^2)；其概率密度函数为正态分布的期望值μ决定其位置，其标准差σ^2决定了分布的幅度。当σ=1，μ=0时正态分布为**标准正态分布**。
- 正态曲线呈钟型，两头低、中间高，左右对称，故又称钟形曲线。



正态分布

- 值域：(-∞,+∞)
- 概率密度函数： $f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

- 期望：μ,σ^2
- 方差：σ^2
- 中位数：μ
- 众数：μ
- 偏度：0
- 峰度：3

两点分布(0-1分布、伯努利分布) $X \sim B(x, p)$

- 已知随机变量X的分布概率为：

X	1	0
P	p	1-p或者q

$$\Rightarrow P(X=1)+P(X=0)=1$$

- 则有：

$$E(X)=1*p+0*q=p$$

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

$$=1^2*p+0^2*(1-p)-p^2=pq$$

二项分布(Bernoulli Distribution) $\xi \sim B(n, p)$

- 二项分布即重复n次独立的重复实验中(伯努利实验)，每次实验只有两种可能的结果，并且两种结果发生与否互相独立，并且与其它实验结果无关，也就是事件发生与否的概率在每一次独立实验中保持不变；用ξ表示事件A的发生次数，使用p表示发生的概率，使用q表示不发生的概率，那么N次独立实验中发生K次的概率为：

$$P(\xi = K) = C_n^k * p^k * q^{n-k}$$

- 注意当n为1的时候，二项分布即两点分布
- 二项分布期望： $E(\xi) = np$ ，方差： $D(\xi) = npq$

二项分布(Bernoulli Distribution) $\xi \sim B(n,p)$

- n次事件独立互不影响，所以期望为n个事件期望的和

$$E(X)=\sum_{i}^{n}E(X_i)=\sum_{i}^{n}p=np$$

$$D(X)=\sum_{i=1}^{n}D(X_i)=\sum_{i=1}^{n}pq=npq$$

几何分布(Geometric Distribution) $X \sim B(p)$

- 几何分布是指在n次伯努利实验中，实验k次才得到第一次成功的几率，其中前k-1次实验均失败，第k次实验成功。
- 如果事件发生的概率为p，不发生的概率为q，则第k次事件成功的概率为：

$$P(\xi = k) = q^{k-1} * p$$

- 几何分布期望为：$E(\xi) = \frac{1}{p}$ ，方差为：$D(\xi) = \frac{1-p}{p^2}$

几何分布(Geometric Distribution) $X \sim B(p)$

- 有概率函数 $P(\xi = k) = q^{k-1} * p$ ，可得期望为：

$$E(\xi) = \sum_{i=1}^{k} i * P(\xi = i) = p + 2pq + 3pq^2 + \ldots + kpq^{k-1} = p \cdot S_k$$

$$S_k = 1 + 2q + 3q^2 + \ldots + kq^{k-1}$$

$$qS_k = q + 2q^2 + 3q^3 + \ldots + (k-1)q^{k-1} + kq^k$$

$$S_k - qS_k = 0 \Rightarrow (1-q)S_k = 1 + q + q^2 + \ldots + q^{k-1} - kq^k$$

$$\Rightarrow \quad S_k = \frac{1-q^k}{(1-q)^2} - \frac{kq^k}{1-q} \quad \lim_{k \to \infty} q^k = 0 \quad \Rightarrow E(\xi) = p \cdot S_k = p \cdot \left(\frac{1-q^k}{(1-q)^2} - \frac{kq^k}{1-q} \right) = \frac{1}{p}$$

几何分布(Geometric Distribution) $X \sim B(p)$

- 有概率函数 $P(\xi = k) = q^{k-1} * p$ ，可得方差为：

$$E(\xi^2) = \frac{2-p}{p^2} \qquad E(\xi) = \frac{1}{p}$$

$$D(\xi) = E(\xi^2) - [E(\xi)]^2$$

$$= \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

泊松分布(Poisson Distribution) $X \sim \pi(\lambda)$

- **泊松分布**常用于描述单位时间内随机事件发生的次数，使用参数λ表示单位时间(或单位面积)内随机事件的平均发生率。
- 当二项分布的n很大，而p很小的时候，泊松分布可以看做是二项分布的近似，一般当n>20,p<0.05时候，就可以近似计算了。
- 概率分布函数为：$P(X=k)=\frac{\lambda^k}{k!}\cdot e^{-\lambda}, k=0,1,2\ldots, \lambda>0$
- 泊松分布的期望和方差均为λ

泊松分布(Poisson Distribution)

$$X \sim \pi(\lambda)$$

$$P(X=k)=\frac{\lambda^k}{k!}\cdot e^{-\lambda}, k=0,1,2\ldots., \lambda>0 \quad f(x)=\sum_{k=0}^{\infty}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \Rightarrow e^x=\sum_{k=0}^{\infty}\frac{x^k}{k!}$$

$$E(X)=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\lambda\cdot e^{-\lambda}\cdot\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda\cdot e^{-\lambda}\cdot e^{\lambda}=\lambda$$

$$E(X^2)=E[X(X-1)+X]=\sum_{k=0}^{\infty}k(k-1)\frac{\lambda^k}{k!}e^{-\lambda}+\lambda=\lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}+\lambda=\lambda^2+\lambda$$

$$D(X)=E(X^2)-[E(X)]^2=\lambda^2+\lambda-\lambda^2=\lambda$$

泊松分布(Poisson Distribution) $X \sim \pi(\lambda)$

- 泊松分布特别适合用于描述单位时间(或单位空间)内随机事件发生的次数，例如：
  - 某一服务设施在一定时间内到达的人数
  - 电话交换机接到呼叫的次数
  - 汽车站台的候客人数
  - 机器出现的故障数
  - 自然灾害发生的次数
  - 一个产品上的缺陷数
  - 显微镜下单位分区内的细菌分布数

均匀分布 $X \sim U(a,b)$

- 概率密度函数为：

$$f(x)=\begin{cases}\dfrac{1}{b-a}, a<x<b \\ 0, \text{其它x取值}\end{cases}$$

- 期望为：

$$E(x)=\int_{-\infty}^{+\infty} xf(x)dx=\int_a^b \frac{1}{b-a}xdx=\frac{1}{b-a}\cdot\frac{1}{2}\cdot\left(b^2-a^2\right)=\frac{1}{2}(a+b)$$

- 方差为：

$$(b-a)^3=(b-a)\left(b^2+ab+a^2\right)$$

$$E\left(x^2\right)=\int_a^b x^2\frac{1}{b-a}dx-\left(\frac{a+b}{2}\right)^2=\frac{(b-a)^2}{12}$$

指数分布(Exponential Distribution) $X \sim E(\lambda)$

- 概率密度函数为：$f(x)=\begin{cases}\lambda e^{-\lambda x}, x>0\\ 0, x\le 0\end{cases}$,其中$\lambda>0$

- 期望为：$E(x)=\int_{-\infty}^{\infty} xf(x)dx=-xe^{-\lambda x}\Big|_0^{+\infty}+\int_0^{\infty} e^{-\lambda x}dx=-\frac{1}{\lambda}\int_0^{\infty} e^{-\lambda x}d(-\lambda x)=\frac{1}{\lambda}$

- 方差为：

$$D(x)=E(x^2)-[E(x)]^2=\int_0^{+\infty} x^2\lambda e^{-\lambda x}dx-\frac{1}{\lambda^2}=2\frac{1}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}$$

指数分布(Exponential Distribution) $X \sim E(\lambda)$

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, x>0 \\ 0, x\le 0\end{cases}, 其中\lambda>0$$

- λ>0是分布的一个参数，被称为率参数(rate parameter)，即每单位时间内发生某事件的次数，指数分布的区间是[0,+∞)。
- 指数分布是一种连续概率分布，常用来表示独立随机事件发生的时间间隔，比如旅客进机场的时间间隔等等。
- 许多电子产品的寿命分布一般都服从指数分布，指数分布是可靠性研究中最常用的一种分布形式。

指数分布(Exponential Distribution) $X \sim E(\lambda)$

- 指数分布的的一个最重要的特性就是无记忆性(Memoryless Property， 又称遗失记忆性)，这表示如果一个随机变量呈指数分布，那么有：

$$P(T > s + t \mid T > s) = p(T > t), \quad s, t \geq 0$$

即如果T是某一个元件的寿命，已知元件使用了s小时，那么它总共使用至少s+t小时的条件概率，与从开始使用时算起到它使用至少t小时的概率相等。

切比雪夫不等式/切比雪夫定理

- 设随机变量X的期望为μ，方差为σ^2，对于任意的正数ε，有：

$$P\left\{\left|X-\mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{\varepsilon^2}$$

- 切比雪夫不等式的含义是：DX(方差)越小，时间{|X-μ|<ε}发生的概率就越大，即：X取的值基本上集中在期望μ附近

大数定律

- 设随机变量$X_1,X_2,\ldots,X_n$是一列相互独立的随机变量(或者两两不相关)，并且分别存在期望$E(X_k)$和方差$D(X_k)$，对于任意小的正数ε，有：

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E\left(X_k\right)\right| < \varepsilon\right\} = 1$$

- 当具有相同期望μ和方差为σ^2的时候，对随机变量的均值: $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

则有： $\lim_{n\to\infty} P\left\{\left|Y_n - \mu\right| < \varepsilon\right\} = 1$

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。

中心极限定理

- 中心极限定理(Central Limit Theorem)；假设$\{X_n\}$为独立同分布的随机变量序列，并具有相同的期望μ和方差为σ^2，则$\{X_n\}$服从中心极限定理，且Z_n为随机序列$\{X_n\}$的规范和：

$$Y_n = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$$

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - n\mu}{\sqrt{n}\sigma} \to N(0,1)$$

中心极限定理

- 中心极限定理的意义：设从均值为μ、方差为σ^2有限的任意一个总体中抽取样本量为n的样本，当n充分大时，样本均值的抽样分布近似服从均值为μ、方差为σ^2/n 的正态分布。
- 中心极限定理就是一般在同分布的情况下，抽样样本值的和在总体数量趋于无穷时的极限分布近似与正态分布。

最大似然法

- 最大似然法(Maximum Likelihood, ML)也称为最大概似估计、极大似然估计，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。
- 最大似然法估计值的一般步骤如下：
  - 写出似然函数
  - 对似然函数取对数，并整理
  - 求导数
  - 解似然方程

最大似然法

- 设总体分布为f(x,θ), $\{X_n\}$为该总体采样得到的样本。因为随机序列$\{X_n\}$独立同分布，则它们的联合密度函数为：

$$L(x_1,x_2,\ldots,x_n;\theta_1,\theta_2,\ldots,\theta_n)=\prod_{i=1}^{n}f(x_i;\theta_1,\theta_2,\ldots,\theta_n)$$

- 这里θ被看做固定但是未知的参数，反过来，因为样本已经存在，可以看做$\{X_n\}$是固定的，L(x,θ)是关于θ的函数，即似然函数。
- 求参数θ的值，使得似然函数取最大值，这种方法叫做最大似然估计法。

最大似然法

- 若给定一组样本{X_n}，已知随机样本符合高斯分布N(μ,σ^2)，试估计σ和μ的值。

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad L(x)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$l(x)=\log(L(x))=\log\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$=\sum_{i=1}^{n}\log\left(\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)=\sum_{i=1}^{n}\log\left(\frac{1}{\sqrt{2\pi}\sigma}\right)-\sum_{i=1}^{n}-\frac{(x-\mu)^2}{2\sigma^2}=-\frac{n}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{i}(x-\mu)^2$$

最大似然法

$$l(x)=-\frac{n}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\sum_i(x-\mu)^2$$

$$=-\frac{n}{2}\log(2\pi)-\left[n\log(\sigma)+\frac{1}{2\sigma^2}\sum_i(x-\mu)^2\right]=-\frac{n}{2}\log(2\pi)-l_2(x)$$

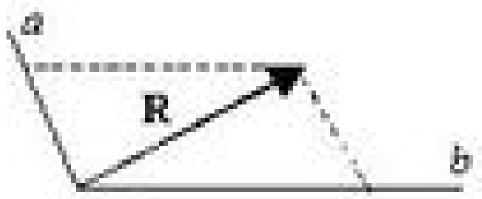
- 要求似然函数l(x)最大，即l(x)求极值即可，将似然函数对参数μ和σ分别求偏导数

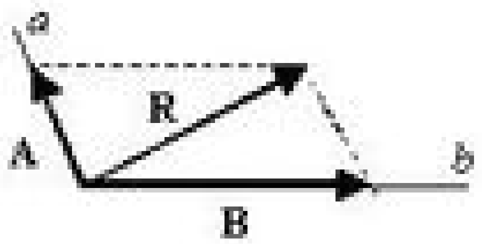
$$\frac{l(x)}{d\mu}=-\frac{1}{\sigma^2}\sum_i\left(x_i-\mu\right)=0\Rightarrow\mu=\frac{1}{n}\sum_i x_i$$

$$\frac{l_2(x)}{d\sigma}=\frac{\pi}{\sigma}-\frac{\sum_i\left(x_i-\mu\right)^2}{\sigma^3}=0\Rightarrow\sigma^2=\frac{1}{n}\sum_i\left(x_i-\mu\right)^2$$

向量

- 向量：是指具有n个互相独立的性质(维度)的对象的表示，向量常使用字母+箭头的形式进行表示，也可以使用几何坐标来表示向量，比如 $\vec{a} = \overrightarrow{OP} = xi + yj + zk$ ，可以用坐标(i,j,k)表示向量a
- 向量的模：向量的大小，也就是向量的长度，向量坐标到原点的距离，常记作|a|
- 单位向量：长度为一个单位(即模为1)的向量就叫做单位向量





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

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

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

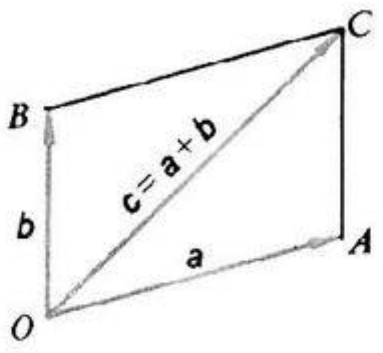


图 4 向量的加法

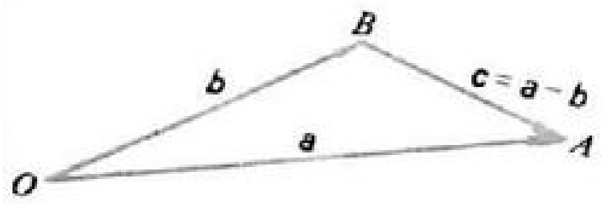


图 5 向量的减法

- 数乘：实数λ和向量a的叉乘乘积还是一个向量，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。

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

正交向量

- 正交向量：如果两个向量的点积为零，那么这两个向量互为正交向量；在几何意义上来讲，正交向量在二维/三维空间上其实就是两个向量垂直。
- 如果两个或者多个向量，它们的点积均为0，那么它们互相称为正交向量。

线性代数

- 线性(linear)指量(变量)与量(变量)之间按比例、成直线关系，在数学上可以理解为一阶导数为常数的函数；而非线性(non-linear)是指不成比例、没有直线关系，一阶导数不是常数的函数。
- 线性代数中的基本量指的是向量，基本关系是严格的线性关系；也就是可以简单的将线性代数理解为向量与向量之间的线性关系的映射。

矩阵

- 矩阵：即描述线性代数中线性关系的参数，即矩阵是一个线性变换，可以将一些向量转换为另一些向量。
- 初等代数中，y=ax表示的是x到y的一种映射关系，其中a是描述这中关系的参数。
- 线性代数中，Y=AX表示的是向量X和Y的一种映射关系，其中A是描述这种关系的参数。

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，当F为实数域R时，A叫做实矩阵，当F为复数域C时，A叫做复矩阵。

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

方阵、零矩阵、单位矩阵

- 如果矩阵A中的所有元素(m*n个)均为0，那么此时矩阵A叫做零矩阵，可以记作0。
- 如果矩阵A中m等于n，那么称矩阵A为n阶矩阵(或n阶方阵)。
- 单位矩阵：n阶方阵中除了主对角线上的元素外，其它元素均为0，主对角线元素均为1，那么此时的n阶方阵叫做n阶单位矩阵。单位矩阵常用E或者I表示。

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：$c_{ij} = a_{ij} \pm b_{ij}$

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

$$A+B=B+A \qquad (A+B)+C=A+(B+C)$$

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： $c_{ij} = \lambda a_{ij}$

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

$$(\lambda\mu)A = \lambda(\mu A) \qquad (\lambda + \mu)A = \lambda A + \mu A$$

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

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：$\vec{y} = A\vec{x}$

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

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

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=A*B是m*n阶矩阵，并且矩阵C中的元素满足: $c_{ij}=\sum_{k=1}^{s}a_{ik}b_{kj}$

$$A=\begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1s} \\ a_{21} & a_{22} & \dots & a_{2s} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{ms} \end{Bmatrix} \qquad B=\begin{Bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{s1} & b_{s2} & \dots & b_{sn} \end{Bmatrix} \qquad C=A*B=\begin{Bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{Bmatrix}$$

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

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

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

$$\left(A^T\right)^T = A \qquad (\lambda A)^T = \lambda A^T \qquad (AB)^T = B^T A^T$$

矩阵的秩

- 在m*n矩阵A中，任取k行k列，不改变这k^2个元素的在A中的次序，得到k阶方阵，称为矩阵A的k阶子式。
  - m*n阶矩阵A的k阶子式有 $C_m^k C_n^k$ 个
- 设在矩阵A中有一个不等于0的r阶子式D，且所有r+1阶子式(如果存在)全等于0，那么D称为矩阵A的最高阶非零子式，r称为矩阵A的秩，记作R(A)=r
  - n*n的可逆矩阵，秩为n
  - 可逆矩阵又称为满秩矩阵
  - 矩阵的秩等于它行(列)向量组的秩

方阵行列式

- 行列式是数学的一个函数，可以看做在几何空间中，一个线性变换对“面积”/“体积”的影响。
- 方阵行列式，n阶方阵A的方阵行列式表示为|A|或者det(A)
    - 1X1的方阵，其行列式等于该元素本身。 $A=\left(a_{11}\right) \quad |A|=a_{11}$
    - 2X2的方阵，其行列式用主对角线元素乘积减去次对角线元素的乘积。

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad |A|=a_{11}*a_{22}-a_{12}*a_{22}$$

方阵行列式

- 方阵行列式，n阶方阵A的方阵行列式表示为|A|或者det(A)；n阶方阵A的行列式计算规则为：主对角线元素乘积和减去次对角线元素乘积和。设r_i为第i个主对角线的积，l_i为第i个次对角线的积。0≤i≤n

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

$$r_i=\prod_{k=1}^{i} a_{k(n+k-i)} * \prod_{k=i+1}^{n} a_{k(k-i)}$$

$$l_i=\prod_{k=1}^{i} a_{k(i-k+1)} * \prod_{k=i+1}^{n} a_{k(n-k+i+1)}$$

$$|A|=\sum_{i=1}^{n} r_i - \sum_{i=1}^{n} l_i$$

方阵行列式

- 根据方阵行列式的计算规则可以得到三阶方阵A的行列式为：

$$A=\begin{Bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{Bmatrix}$$

$$r_i=\prod_{k=1}^{i} a_{k(3-i+k)} * \prod_{k=i+1}^{3} a_{k(k-i)}$$

$$l_i=\prod_{k=1}^{i} a_{k(i-k+1)} * \prod_{k=i+1}^{3} a_{k(4-k+i)}$$

$$|A|=\sum_{i=1}^{3} r_i-\sum_{i=1}^{3} l_i$$

$$=a_{13}a_{21}a_{32}+a_{12}a_{23}a_{31}+a_{11}a_{22}a_{33}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}-a_{13}a_{22}a_{31}$$

代数余子式

- 在一个n阶的行列式A中，把元素a_{ij}(i,j=1,2,3,....,n)所在的行和列划去后，剩下的$(n-1)^2$个元素按照原来的次序组成的一个n-1阶行列式M_{ij}，称为元素a_{ij}的余子式。M_{ij}带上符号$(-1)^{(i+j)}$称为a_{ij}的代数余子式，记作：$A_{ij}=(-1)^{(i+j)}M_{ij}$

$$\forall 1 \le j \le n, |A| = \sum_{i=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij}$$

$$\forall 1 \le i \le n, |A| = \sum_{j=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij}$$

$$A = \begin{Bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{Bmatrix}$$

$$|A| = a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} + a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

伴随矩阵

- 对于n阶方阵的任意元素a_{ij}都有各自的代数余子式$A_{ij}=(-1)^{(i+j)}M_{ij}$，将所有的代数余子式按照次序进行排列，可以得到一个n阶的方阵A^*；那么A^*称为矩阵A的伴随矩阵。
  - 注意：A_{ij}位于A^*的第j行第i列

$$A \cdot A^* = |A| \cdot E$$

$$A^* = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

方阵的逆

- 设A是数域上的一个n阶方阵，若在相同的数域上存在另一个n阶方阵B，使得AB=BA=E，那么称B为A的逆矩阵，而A被称为可逆矩阵或非奇异矩阵。如果A不存在逆矩阵，那么A称为奇异矩阵。A的逆矩阵记作：A^{-1}
- 具有以下性质：
  - 如果矩阵A是可逆的，那么矩阵A的逆矩阵是唯一的。
  - A的逆矩阵的逆矩阵还是A，记作$(A^{-1})^{-1}=A$
  - 可逆矩阵A的转置矩阵A^T也可逆，并且$(A^T)^{-1}=(A^{-1})^T$
  - 若矩阵A可逆，则矩阵A满足消去律，即AB=BC => B=C
  - 矩阵A可逆的充要条件是行列式|A|不等于0

方阵的逆

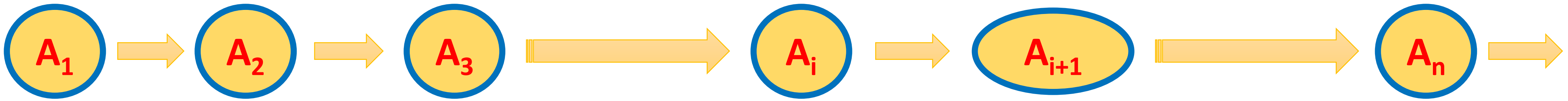
$$\forall 1 \le i \le n, |A| = \sum_{j=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij}$$

$$A = \begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{Bmatrix} \qquad A^* = \begin{Bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{Bmatrix}$$

$$A \cdot A^* = \begin{Bmatrix} |A| & 0 & \dots & 0 \\ 0 & |A| & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & |A| \end{Bmatrix} = |A|E \Rightarrow A^{-1} = \frac{1}{|A|} A^*$$

状态转移模型

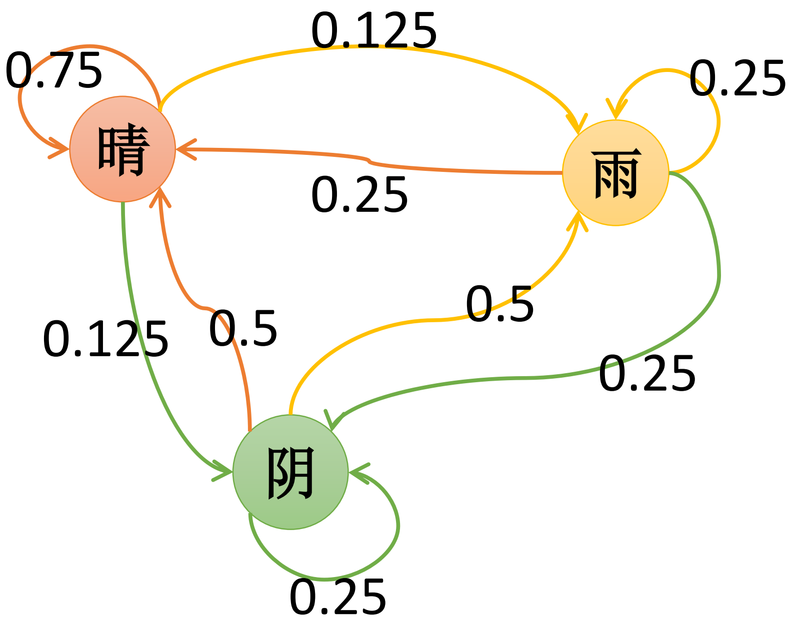
- 考虑某随机过程π，它的状态有n个，使用1~n表示。记在当前时刻t时位于i状态，它在t+1时刻位于j状态的概率为P(i,j)=P(j|i)；
  - 即状态转移的概率只依赖上一个状态值



状态转移模型

- 设将天气状态分为晴、阴、雨三种状态，假定某天的天气状态只和上一天的天气状态有关，状态使用1(晴)、2(阴)、3(雨)表示，转移概率矩阵如下：

今/明	晴	阴	雨
晴	0.75	0.125	0.125
阴	0.5	0.25	0.25
雨	0.25	0.5	0.25



概率转移矩阵

- 第n+1天天气状态为j的概率为：

$$\pi\left(X_{n+1}=j\right)=\sum_{i=1}^{K}\pi\left(X_{n}=i\right)\cdot P\left(X_{n+1}=j \mid X_{n}=i\right)$$

$$\Rightarrow \pi^{n+1}=\pi^{n}\cdot P$$

- 因此，矩阵P即为条件概率转移矩阵。
    - 矩阵P的第i行元素表示，在上一个状态为i的时候的分布概率，即每行元素的和必须为1

初始概率π[0.5,0.3,0.2]

$$P=\begin{bmatrix} 0.75 & 0.125 & 0.125 \\ 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

第n天	晴	阴	雨
0	0.5	0.3	0.2
1	0.575	0.2375	0.1875
2	0.5969	0.225	0.1781
3	0.6047	0.2199	0.1754
4	0.6073	0.2183	0.1744
5	0.6082	0.2177	0.1741
6	0.6085	0.2175	0.174
7	0.6086	0.2174	0.1739
8	0.6087	0.2174	0.1739
9	0.6087	0.2174	0.1739
10	0.6087	0.2174	0.1739
11	0.6087	0.2174	0.1739

初始概率π[0.1,0.6,0.3]

$$P=\begin{bmatrix}0.75 & 0.125 & 0.125\\ 0.5 & 0.25 & 0.25\\ 0.25 & 0.5 & 0.25\end{bmatrix}$$

第n天	晴	阴	雨
0	0.1	0.6	0.3
1	0.45	0.3125	0.2375
2	0.5531	0.2531	0.1937
3	0.5898	0.2293	0.1809
4	0.6022	0.2215	0.1763
5	0.6065	0.2188	0.1747
6	0.6079	0.2179	0.1742
7	0.6084	0.2176	0.174
8	0.6086	0.2174	0.1739
9	0.6087	0.2174	0.1739
10	0.6087	0.2174	0.1739
11	0.6087	0.2174	0.1739

平稳分布

- 初始概率不同，但经过若干次迭代后，π最终会收敛在某个分布上。
- 从而，这是转移概率矩阵P的性质，而非初始分布的性质，事实上，上述天气的转移矩阵P的n次幂，最终会收敛到(0.6087, 0.2174, 0.1739)，n>10

向量组

- 向量组：有限个相同维数的行向量或列向量组合成的一个集合就叫做向量组

$$A=\left(\vec{a}_1,\vec{a}_2,\vec{a}_2.....\vec{a}_n....\right)$$

- 向量组是有多个向量构成，可以表示为矩阵

向量组等价

- 向量b能由向量组A:$a_1,a_2,...a_m$线性表示的充要条件为矩阵$A=(a_1,a_2,...a_m)$的秩等于矩阵$B=(a_1,a_2,...a_m,b)$的秩。
- 设有两个向量组A:$a_1,a_2,...a_m$及B:$b_1,b_2,...b_m$，若B组的向量都能由向量组A线性表示，则称向量组B能由向量组A线性表示。若向量组A和向量组B能相互线性表示，则称两个向量组等价。

系数矩阵

- 将向量组A和B所构成的矩阵依次记作$A=(a_1,a_2,\ldots a_m)$和$B=(b_1,b_2,\ldots b_n)$，且组B能由向量组A线性表示。即对每个向量b_j，存在$k_{1j},k_{2j},\ldots,k_{mj}$；使得

$$b_j = k_{1j}a_1 + k_{2j}a_2 + \ldots + k_{mj}a_m = \begin{pmatrix} a_1 & a_2 & \ldots & a_m \end{pmatrix} \begin{pmatrix} k_{1j} \\ k_{2j} \\ \ldots \\ k_{mj} \end{pmatrix}$$

- 从而可以得到系数矩阵K

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_m \end{pmatrix} \begin{bmatrix} k_{11} & k_{12} & \ldots & k_{1n} \\ k_{21} & k_{22} & \ldots & k_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ k_{m1} & k_{m2} & \ldots & k_{mn} \end{bmatrix} = \begin{pmatrix} b_1 & b_2 & \ldots & b_n \end{pmatrix}$$

正交矩阵

- 若n阶方阵A满足$A^TA=E$，则称A为正交矩阵，简称正交阵。
  - A是正交阵的充要条件：A的列(行)向量都是单位向量，且两两正交。
- 若A为正交矩阵，x为向量，则Ax称为正交变换。
  - 正交变换不改变向量的长度。
- 正交矩阵的性质：
  - 若A为正交矩阵，则逆矩阵A^{-1}也为正交矩阵
  - 若P、Q为正交矩阵，那么P*Q也为正交矩阵

对称矩阵

- 元素以对角线为对称轴对应相等的矩阵就叫做对称矩阵
- 对称矩阵具有的特性：
  - 对称矩阵中a_{ij}等于a_{ji}
  - 对称矩阵一定是方阵；并且对于任何的方阵A，$A+A^T$是对称矩阵
  - 除对角线外的其它元素均为0的矩阵叫做对角矩阵
  - 矩阵中的每个元素都是实数的对称矩阵叫做实对称矩阵

$$A=\begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{Bmatrix}$$

$$a_{12}=a_{21}$$
$$a_{ij}=a_{ji}$$
$$a_{1n}=a_{n1}$$

特征值和特征向量

- A为n阶矩阵，若数λ和n维非0列向量x满足Ax=λx，那么数λ称为A的**特征值**，x称为A的对应于特征值λ的**特征向量**。并且|λE-A|叫做A的**特征多项式**。当特征多项式等于0的时候，称为A的特征方程，特征方程是一个齐次线性方程组，求解特征值的过程其实就是求解特征方程的解。

$$
\begin{aligned}
&Ax = \lambda x \\
&\Rightarrow Ax = \lambda E x \\
&\Rightarrow (\lambda E - A)x = 0
\end{aligned}
$$

$$
|\lambda E - A| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} = 0
$$

特征值和特征向量求解1

$$A=\begin{vmatrix}4 & 2 & -5\\ 6 & 4 & -9\\ 5 & 3 & -7\end{vmatrix}$$

$$|\lambda E-A|=\begin{vmatrix}\lambda-4 & -2 & 5\\ -6 & \lambda-4 & 9\\ -5 & -3 & \lambda+7\end{vmatrix}=\lambda^2(\lambda-1)=0 \Rightarrow \lambda_1=1,\ \lambda_2=\lambda_3=0$$

$$\lambda_1=1\Rightarrow(E-A)x=0$$

$$E-A=\begin{bmatrix}-3 & -2 & 5\\ -6 & -3 & 9\\ -5 & -3 & 8\end{bmatrix}\to\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & -1\\ 0 & 0 & 0\end{bmatrix}$$

$$\Rightarrow\begin{cases}x_1-x_3=0\\ x_2-x_3=0\end{cases}\Rightarrow\begin{cases}x_1=x_3\\ x_2=x_3\end{cases}\xRightarrow{\text{令}x_3=1}\xi_1=\begin{bmatrix}1\\1\\1\end{bmatrix}$$

$$\xi_2=\xi_3=\begin{bmatrix}1\\3\\2\end{bmatrix}$$

特征值和特征向量求解2

$$A=\begin{vmatrix}5&0&0\\0&2&0\\0&0&6\end{vmatrix}$$

$$|\lambda E-A|=\begin{vmatrix}\lambda-5&0&0\\0&\lambda-2&0\\0&0&\lambda-6\end{vmatrix}=(\lambda-5)(\lambda-2)(\lambda-6)=0$$

$$\Rightarrow \lambda_1=5,\ \lambda_2=2,\ \lambda_3=6$$

$$\lambda_1=1\Rightarrow (E-A)x=0$$

$$E-A=\begin{bmatrix}0&0&0\\0&3&0\\0&0&-1\end{bmatrix}\Rightarrow\begin{cases}x_2=0\\x_3=0\end{cases}\xRightarrow{\text{令}x_1=1}\xi_1=\begin{bmatrix}1\\0\\0\end{bmatrix}\quad \xi_2=\begin{bmatrix}0\\1\\0\end{bmatrix}\quad \xi_3=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

特征值的性质

- n阶方阵$A=(a_{ij})$的所有特征根λ_1、$\lambda_2.....\lambda_n$，则有

$$\lambda_1+\lambda_2+\ldots+\lambda_n=\sum_{i=1}^{n}a_{ii} \qquad \lambda_1\lambda_2\ldots\lambda_n=|A|$$

- 若λ是可逆矩阵A的一个特征根，x为对应的特征向量：
  - 则1/λ是矩阵A^{-1}的一个特征根，x仍为对应的特征向量。
  - 则λ^m是矩阵A^m的一个特征根，x仍未对应的特征向量。
- 设λ_1、$\lambda_2.....\lambda_n$是方阵A的互不相同的特征值，x_i是λ_i的特征向量，则$x_1,x_2...x_n$线性无关，即不相同特征值的特征向量线性无关。

可对角化矩阵

- 如果一个n阶方阵A相似于对角矩阵，也就是，如果存在一个可逆矩阵P使得PAP是对角矩阵，则称矩阵A为可对角化矩阵。并且最终对角矩阵的特征值就是矩阵A的特征值。

$$P^{-1}AP = \Lambda$$

- 矩阵对角化在PCA、白化等机器学习领域应用的比较多。

正定矩阵

- 对于n阶方阵A，若任意n阶向量x，都有$x^TAx>0$，则称矩阵A为正定矩阵。
- 若$x^TAx>=0$，则矩阵A为半正定矩阵。
- 性质：
  - 正定矩阵的任意主子矩阵也是正定矩阵。
  - 若A为n阶正定矩阵，则A为n阶可逆矩阵。

例题

- 给定凸维C的定义如下：

$$\forall x_1, x_2 \in C, \theta_1, \theta_2 \geq 0, \text{有}\ \theta_1 x_1 + \theta_2 x_2 \in C$$

- 证明：n阶半正定方阵的集合为凸维

$$\forall \vec{z}, \vec{z}^T A \vec{z} \geq 0, \vec{z}^T B \vec{z} \geq 0 \qquad \Rightarrow \vec{z}^T \left(\theta_1 A + \theta_2 B\right) \vec{z} = \vec{z}^T \theta_1 A \vec{z} + \vec{z}^T \theta_2 B \vec{z}$$

$$\forall \theta_1, \theta_2 \geq 0, \qquad \Rightarrow = \theta_1 \vec{z}^T A \vec{z} + \theta_2 \vec{z}^T B \vec{z} \geq 0$$

- 即：$\forall \theta_1, \theta_2 \geq 0, \theta_1 A + \theta_2 B$ 为半正定阵。从而，n阶半正定阵的集合为凸维

奇异矩阵

- 若方阵A的行列式的值等于0，那么方阵A叫做奇异矩阵，否则叫做非奇异矩阵。
- 可逆矩阵就是非奇异矩阵，非奇异矩阵也是可逆矩阵。
- 若A为奇异矩阵，则Ax=0有无穷解，Ax=b有无穷阶或者无解。
- 若A为非奇异矩阵，则Ax=0有且只有唯一零解，Ax=b有唯一解。

QR分解

- 对于m*n的列满秩矩阵A，必有：$A_{m*n}=Q_{m*n}\cdot R_{n*n}$
- 其中Q为列正价矩阵，R为非奇异上三角矩阵，当要求R的对角线元素为正的时候，该分解唯一。
- 该分解叫做QR分解，常用于求解A的特征值、A的逆等问题。

$$A=QR\Rightarrow A_1=Q^TAQ=RQ$$

$$\ldots.\qquad\Rightarrow A_k\rightarrow diag\{\lambda_1,\lambda_2\ldots.\lambda_n\}$$

$$A_k=Q_kR_k\Rightarrow A_{k+1}=R_kQ_k$$

SVD

- 奇异值分解(Singular Value Decomposition)是一种重要的矩阵分解方法，可以看做是对称方阵在任意矩阵上的推广。
- 假设A为一个m*n阶实矩阵，则存在一个分解使得：

$$A_{m*n} = U_{m*m} \Sigma_{m*n} V_{n*n}^{T}$$

  - 通常将奇异值由大到小排列，这样Σ便能由A唯一确定了。
- 与特征值、特征向量的概念相对应，则：
  - Σ对角线上的元素称为矩阵A的奇异值
  - U和V称为A的左/右奇异向量矩阵

向量的导数

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

$$A=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \qquad \vec{x}=\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \qquad \vec{y}=\begin{pmatrix} a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n \\ a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n \\ \ldots. \\ a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n \end{pmatrix}$$

$$\frac{\partial \vec{y}}{\partial \vec{x}} = \frac{\partial A \vec{x}}{\partial \vec{x}} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix} = A^T$$

向量的导数

- 向量偏导公式

$$\frac{\partial A\vec{x}}{\partial \vec{x}} = A^T$$

$$\frac{\partial A\vec{x}}{\partial \vec{x}^T} = A$$

$$\frac{\partial \left(\vec{x}^T A\right)}{\partial \vec{x}} = A$$

标量对向量的导数

- A为n*n的矩阵，x为n*1的列向量，记 $y = \vec{x}^T \cdot A \cdot \vec{x}$
- 同理可得： $\frac{\partial y}{\partial \vec{x}} = \frac{\partial\left(\vec{x}^T \cdot A \cdot \vec{x}\right)}{\partial \vec{x}} = \left(A^T + A\right) \cdot \vec{x}$
- 若A为对称矩阵，则有 $\frac{\partial\left(\vec{x}^T A \vec{x}\right)}{\partial \vec{x}} = 2A\vec{x}$

标量对向量的导数

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

$$\vec{x}^T \cdot A \cdot \vec{x} = (x_1, x_2 \ldots x_n)\left(\sum_{j=1}^n a_{1j}x_j, \sum_{j=1}^n a_{2j}x_j, \ldots \sum_{j=1}^n a_{nj}x_j\right)^T = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j$$

$$\frac{\partial\left(\vec{x}^T \cdot A \cdot \vec{x}\right)}{\partial \vec{x}} = \left(\sum_{j=1}^n a_{ij}x_j\right) + \left(\sum_{j=1}^n a_{ji}x_j\right) = \sum_{j=1}^n \left(a_{ij} + a_{ji}\right)x_j$$

标量对方阵的导数

- A为n*n的矩阵，|A|为A的行列式，计算 $\frac{\partial |A|}{\partial A}$

$$\forall 1 \le i \le n, |A| = \sum_{j=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij}$$

$$\frac{\partial |A|}{\partial a_{ij}} = \frac{\partial \left(\sum_{j=1}^{n} a_{ij} \cdot (-1)^{i+j} M_{ij} \right)}{\partial a_{ij}} = (-1)^{i+j} M_{ij} = A^*{}_{ji}$$

$$\frac{\partial |A|}{\partial A} = \left(A^*\right)^T = |A| \cdot \left(A^{-1}\right)^T$$

北风网
IBEIFENG.COM

THANK YOU

上海育创网络科技有限公司