1 高数

1.1 函数与极限

连续和一直连续的定义和区别
极限的定义

映射：映射是一个法则，这个法则使一个集合中的数在另一个集合中都有唯一确定的对应数，第一个集合叫做定义域，第二个集合叫做值域。
函数：一种特殊类型的映射，它描述了一个输入值与一个输出值之间的关系，他的特殊性在于每个输入都有唯一的输出；定义域是输入值的集合，值域是输出值的集合；

极限：对于函数 $f(x)$，有两种极限，一种是自变量 x 趋近于有限值 x_0，另一种是自变量 x 趋近于无穷 ∞

1. 自变量 x 趋近于有限值 x_0

   $x \to x_0$ 时：对应函数值 $f(x)$ 无限接近于常数 A，则 A 是 $f(x)$ 当 $x \to x_0$ 时的极限，记作

   $\lim\limits_{x\to x_0} f(x) = A$ 或 $f(x) \to A\left(x \to x_0\right)$ 用数学来表示就是 $|f(x) - A| < \varepsilon$，ε 是任意小的正数。

   左极限：x 从左侧趋近 x_0；

   右极限：x 从右侧趋近 x_0；

   函数极限与单侧极限的关系：$x \to x_0$ 时，函数 $f(x)$ 极限存在的充要条件是左右极限存在且相等。

2. 自变量 x 趋近于无穷 ∞

   将上述的 x_0 替换为 ∞ 就行了；左极限对应 $x \to -\infty$，右极限对应 $x \to \infty$

   函数 $f(x)$ 极限存在的充要条件是左右极限存在且相等：$\lim\limits_{x\to\infty} f(x) = \lim\limits_{x\to-\infty} f(x) = A$

函数的连续性：当自变量趋于该点时，函数值的极限与函数在该点所取的值一致。即左右极限都存在且等于 $f(x_0)$

函数的间断点：有了连续的概念，那就有了间断的概念

第一类间断点

可去间断点：左右极限存在且相等但是不等于 $f(x_0)$（f 在 x_0 处无定义或者有定义但是 $f(x_0)$¡ÙA）

跳跃间断点：f 在 x_0 处的左右极限都存在，但是左右极限不相等

第二类间断点：函数 f 至少有一侧的极限不存在

1.2 微分

可导和可微区别与联系（一元函数、多元函数）
可导 与连续的关系是什么？
中值定理
微分中值三大定理之间关系以及费马引理与其之间的联系
罗尔定理
拉格朗日中值定理
洛必达法则
判断极值的方法
怎么判断一个函数的凹凸性
用数学语言描述高数中的一些定理；
拉格朗日因子；

1.2.1 导数与微分

可导（左右导数均存在且相等）一定连续，连续不一定可导。（$\sqrt{x}$ 连续不可导）

导数：自变量和因变量都有微小变化时，变化前后两点间的斜率 $f'(x_0)=\lim\limits_{x\to x_0}\dfrac{f(x)-f(x_0)}{x-x_0}$

微分：自变量有微小变化时，因变量的变化，即 $dy=f'(x_0)dx$

一元函数可导$\Leftrightarrow$可微。

多元函数偏导数连续才可微，也就是可导不一定可微，但可微一定可导

微分方程:表示未知函数，未知函数的导数与自变量之间的关系的方程

齐次方程：可转化为 $\dfrac{\mathrm{dy}}{dx}=\varphi(\dfrac{y}{x})$ 的形式

一阶线性微分方程：

$$\frac{dy}{dx}+P(x)\mathrm{y}=Q(x)\Leftrightarrow y=e^{-\int P(x)dx}(\int Q(x)e^{\int P(x)dx}dx+C)$$

伯努利方程: $\dfrac{dy}{dx}+P(x)\mathrm{y}=Q(x)y^n$ 两边除以 y^n,令 $\mathrm{z}=\mathrm{y}^{1-n}$ 化为一阶线性微分方程

二阶常系数微分方程：$y''+py'+qy=f(x)$，包括齐次通解（由特征根决定）和非齐次通解（由 f(x)的形式决定）

$$\begin{cases} r_1\neq r_2\Rightarrow y=C_1e^{r_1x}+C_2e^{r_2x}\\ r_1=r\Rightarrow y=(C_1+C_2x)e^{rx}\\ r_{1,2}=\alpha+\beta i\Rightarrow y=e^{\alpha x}(C_1\cos\beta x+C_2\sin\beta x)\end{cases}$$

1.2.2 微分中值定理

费马引理：如果函数 $f(x)$ 在点 x_0 处取得极值，且在 x_0 的某个邻域内可导，那么 $f'(x_0)=0$。罗尔中值

定理是费马引理的延伸。闭区间端点处的函数值相等，那么区间内一定存在极值，也就至少存在一点ξ，$f'(\xi)=0$

罗尔中值定理：$f(x)$在闭区间$[a,b]$上连续；开区间(a,b)内可导；闭区间端点处的函数值相等$\Leftrightarrow$(a,b)区间内至少存在一点ξ，$f'(\xi)=0$

拉格朗日中值定理：$f(x)$在闭区间$[a,b]$上连续；开区间(a,b)内可导；$\Leftrightarrow$(a,b)区间内至少存在一点ξ，

$$\frac{f(b)-f(a)}{b-a}=f'(\xi)$$

柯西中值定理：$f(x)$，$g(x)$在闭区间$[a,b]$上连续；开区间(a,b)内可导$\Leftrightarrow$(a,b)区间内至少存在一点ξ，

$$\frac{f(b)-f(a)}{g(b)-g(a)}=\frac{f'(\xi)}{g'(\zeta)}$$

罗尔，拉格朗日都是柯西的特例：柯西$g(x)=x$就是拉格朗日；$f(b)=f(a)$就是罗尔

1.2.3 导数的应用

求极值：极值点存在于导数为 0 的地方

洛必达法则：当x趋近a时，$f(x),g(x)$都趋近0或∞，$f(x)$和$g(x)$的导数都存在且$g(x)$的导数不等于 0，那么 $\lim\limits_{x\to a}\frac{f(x)}{g(x)}=\frac{f'(x)}{g'(x)}$

求拐点：经过某个点使函数凹凸性发生变化；一个函数在某一区间上是凹函数，当且仅当其二阶导数在该区间上恒非负；是凸函数，当且仅当其二阶导数在该区间上恒非正。

多元函数极值：以二元函数为例，可以使用偏导数的方法来求解。对于一个二元函数$f(x,y)$，其偏导数分别为：

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

对于一般的多元函数，可以使用类似的方式求解各个自变量的偏导数。在求解极值的过程中，我们需要求解函数的所有一阶偏导数，并令其等于 0，得到一组方程组，然后解方程组得到可能的极值点。接下来，我们需要使用二阶偏导数来确定这些极值点是极大值还是极小值。对于二元函数，可以使用二阶偏导数判定法来判断极值的类型。具体来说，假设$f(x,y)$的一阶偏导数都存在，那么有以下结论：

如果$\frac{\partial^2 f}{\partial x^2}>0$且$\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2}-(\frac{\partial^2 f}{\partial x\partial y})^2>0$，则$(x,y)$是一个极小值点；

如果$\frac{\partial^2 f}{\partial x^2}<0$且$\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2}-(\frac{\partial^2 f}{\partial x\partial y})^2>0$，则$(x,y)$是一个极大值点；

如果$\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2}-(\frac{\partial^2 f}{\partial x\partial y})^2<0$，则$(x,y)$是$f(x,y)$的一个鞍点（这个点不是极大值点也不是极小值点，而是沿着某

些方向上是极大值点而在另外一些方向上是极小值点)。

条件极值:上述求极值的过程没对自变量有限制条件,实际上大部分情况都是给定自变量的限制条件。
拉格朗日乘数法:一种用于求解带约束条件的最优化问题的方法。具体来说,假设我们要在一定的约束条件下最大化(或最小化)一个函数,那么这个问题可以表示为

$$\max_{\mathbf{x}} f(\mathbf{x})$$

$$\text{s.t.} g_i(\mathbf{x}) = 0, i = 1, 2, \ldots, m$$

其中,$\mathbf{x}$ 是一个 n 维向量,$f(\mathbf{x})$ 是我们要优化的函数,$g_i(\mathbf{x})$ 是约束条件,m 是约束条件的个数。拉格朗日乘数法的基本思想是,在原问题中加入一个关于拉格朗日系数 λ 的新函数:

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

然后,我们求解该函数的偏导数,并令其为零,得到:

$$\frac{\partial \mathcal{L}}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0, j = 1, 2, \ldots, n$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = g_i(\mathbf{x}) = 0, i = 1, 2, \ldots, m$$

这就是拉格朗日乘数法的关键方程组。我们可以通过求解这个方程组来确定最优解 $\mathbf{x}$ 和拉格朗日乘数 λ 的值。需要注意的是,拉格朗日乘数法只能求解约束条件为等式的最优化问题,对于约束条件为不等式的问题,我们需要使用其他方法,如 KKT 条件等。

另外,拉格朗日乘数法的求解需要满足一些充分条件,例如 $f(\mathbf{x})$ 和 $g_i(\mathbf{x})$ 需要是可微的、可导的等等。

1.3 积分

高斯定理
解释莱布尼兹公式
梯度(可以引申出 BP 神经网络的训练参数方法:梯度下降法, 包括批量梯度下降和随机梯度下降)
旋度、散度的物理含义
梯度和散度
解释积分中值定理
连续和积分的关系等

1.3.1 积分的基本概念

定积分:函数 $f(x)$ 在区间 $[a,b]$ 上的积分和 $\int_a^b f(x)\mathrm{d}x$。(由 $x=a$, $x=b$, $f(x)$ 围成的面积)

不定积分:设 $F(x)$ 是函数 $f(x)$ 的一个原函数,我们把函数 $f(x)$ 的所有原函数 $F(x)+C$ 叫做函数 $f(x)$ 的不定积分(其中,C 为任意常数)。

定积分与不定积分之间的关系:若定积分存在,则它是一个具体的数值,而不定积分是一个函数表达式,它们仅仅在数学上有一个计算关系。

微积分基本定理(牛顿-莱布尼茨公式):

第一基本定理：假设 x 定义在[a,b]内，则 $F(x)=\int_a^x f(t)dt$

第二基本定理：连续函数在[a,b]上的定积分等于它的任一原函数在区间上的增量。

$$\int_a^b f(t)dt = F(b)-F(a)$$

反常积分： 1. 积分上限或下限为无穷大； 2. 上限或下限处的函数值不存在。

积分第一中值定理： 若函数 $f(x)$ 在[a,b]上连续，则在积分区间$[a,b]$上至少存在一个点 ε，使下式成立

$$\int_a^b f(x)dx = f(\varepsilon)(b-a) \quad a\le\varepsilon\le b$$

对于二重积分：$\iint_D f(x,y)d\sigma = f(\varepsilon,\mu)\bullet\sigma_0$ σ_0 是 D 的面积，(ε,μ) 是 D 内的一点

积分第二中值定理

（1）$g(x)$ 在[a,b]上单调递减且在 $x\in[a,b]$ 时，$g(x)\ge 0$，那么存在 $\xi\in[a,b]$ 使得

$$\int_a^b f(x)g(x)dx = g(a)\int_a^\xi f(x)dx$$

（2）$g(x)$ 在[a,b]上单调递增且在 $x\in[a,b]$ 时，$g(x)\ge 0$，那么存在 $\xi\in[a,b]$ 使得

$$\int_a^b f(x)g(x)dx = g(b)\int_\xi^{\mathrm{b}} f(x)dx$$

积分中值定理在应用中所起到的重要作用是可以使积分号去掉，或者使复杂的被积函数化为相对简单的被积函数，从而使问题简化。

1.3.2 重积分

二重积分： 求体积
三重积分： 求密度
格林公式： 平面闭区域 D 上的二重积分可以通过沿 D 的边界 L 上的曲线积分来表示。

$$\iint_D\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dxdy = \oint_L Pdx+Qdy$$

高斯定理（公式）： 空间闭区域Ω上的三重积分可用Ω边界上的曲面积分表示

$$\iiint_\Omega\left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)\mathrm{d}V = \oiint_{\partial\Omega} P\mathrm{d}ydz+Q\mathrm{d}zdx+R\mathrm{d}x\mathrm{d}y$$

斯托克斯公式： 平面闭区域上的曲面积分可用边界曲线上的曲线积分来表示。

$$\oint_\Gamma Pdx+Qdy+Rdz = \iint_\Sigma\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)dydz+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)dzdx+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dxdy$$

1.3.3 梯度、散度、旋度

场： 如果空间（或者它的某一部分）的每一点都对应某个物理量的确定值，便叫此空间为该物理量的场。
标量场和矢量场： 如果该物理量仅是数量性质的，便叫相应的场为标量场；如果该物理量是矢量性质的，便叫相应

的场为矢量场

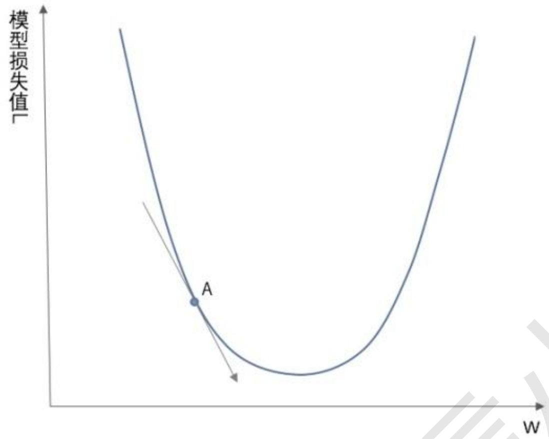
1. 梯度

方向导数：标量场在某点处沿某一方向对距离的变化率

梯度：表示了标量场在某个场点上**变化率最快的方向**（方向导数的方向）和该点的**最大变化率**。梯度是一个矢量。矢量的方向：标量场在某点处所有的方向中，对距离的变化率最大的方向。矢量的大小：等于最大的变化率（方向导数的大小）

矢量场也是有梯度的，具体可以去知乎上搜一搜（超纲了）

梯度的应用--梯度下降法/梯度上升发求极值



对于一个二元函数 $L(w)$,对于任意 w,只要其不断地朝着当前点梯度的方向变换，即 $w \to w-\alpha\dfrac{dL}{w}$ (α 是变化的幅度)，w 最终会慢慢地接近极值点；

对于多元函数 $L(\mathbf{w})$,只要所有自变量都朝着梯度方向移动，最终会找到全局的极值点。**寻找全局最小值的过程就叫做梯度下降；寻找全局最大值的过程叫做梯度上升**；神经网络的优化过程就是最小化一个多元损失函数 $L(\theta)$,θ 是神经元参数，通过梯度下降法寻找最优参数以最小化损失函数。

为什么不能直接令导数为 0？：求出各个维度导数都为 0 的点可能非常复杂，或者求不出（实际中的函数很复杂，例如最近的神经网络的数据元数目可能上亿上千亿，而且还没有一个具体的表达式)，利用梯度下降算法可以通过迭代逼近最优解。

2. 散度

通量：矢量场中矢量在曲面 S 上的曲面积分。$\phi=\int_S \vec{F}d\vec{S}$

散度：单位体积内的矢量通量 $div\vec{F}=\lim\limits_{\Delta V>0}\dfrac{\oint \vec{F}d\vec{S}}{\Delta V}$ (散度>0:源向外发散；散度=0:无源)

3. 旋度

环流量：矢量场中的一条闭合路径的曲线积分 $\Gamma=\oint_C \vec{F}d\vec{l}=\nabla F$

旋度：矢量场在某点处的旋度的方向是沿着使环流面密度取最大值的方向，大小等于环流面密度的最大值。

$$rot\vec{F}=\vec{n}\lim_{\Delta S\to 0}\frac{1}{\Delta S}\oint_c \vec{F}d\vec{l}\,|_{\max}=\nabla\times\vec{\Gamma}$$

1.3 级数

说一下泰勒公式展开？
欧拉公式（反复被考的）
用复变函数表示弹簧的震动
泰勒公式是为了解决什么问题的？

数列：数列是一组有序的数，$\{a_n\}=a_1,a_2,...,a_n,...$ 数列中的每一个数都叫做这个数列的项

级数：指将数列的项依次用加号连接起来的函数。典型的级数有正项级数、交错级数、幂级数、傅里叶级数等。
函数项级数：数列的每一项都是关于 x 的函数
泰勒级数：级数是无穷级数的简称，谈论到级数一定是无穷项的

如果函数 $f(x)$ 在 x_0 处存在 n 阶导，则幂级数

$$\sum_{n=0}^{\infty}\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+\cdots$$

称为在 $f(x)$ 在 x_0 处的泰勒级数。

泰勒公式(泰勒中值定理)

如果函数 $f(x)$ 在 x_0 处存在 n 阶导，那么存在一个 x_0 的邻域内，域内 $f(x)$ 都能展开 成 泰 勒 展 开 式 ：

$f(x)=f(x_0)+f'(x_0)(x-x_0)+...+\dfrac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n(x)$ $R_n(x)$ 是泰勒公式的余项

泰勒公式的意义： 用一个多项式函数去逼近一个给定的函数，同时可以将所有的可导函数以统一的形势表示出来

欧拉公式的证明（由泰勒公式可证）： $e^{jx}=\cos(x)+j\cdot\sin(x)$

$$\cos(x)\ =\sum_{n=0,2,4,\cdots}^{\infty}(-1)^{\frac{n}{2}}\frac{1}{n!}x^n$$

$$j\cdot\sin(x)\ =\sum_{n=1,3,5,\cdots}^{\infty}j(-1)^{\frac{n-1}{2}}\frac{1}{n!}x^n$$

$$e^{jx}=e^0+je^0x+\frac{(-1)e^0}{2!}x^2+\frac{(-1j)e^0}{3!}x^3+\cdots$$
$$=\left(1+\left(-\frac{1}{2!}\right)x^2+\frac{1}{4!}x^4+\cdots\right)+j\cdot\left(x+\left(-\frac{1}{3!}\right)x^3+\frac{1}{5!}x^5+\cdots\right)$$
$$=\sum_{n=0,2,4,\cdots}^{\infty}(-1)^{\frac{n}{2}}\frac{1}{n!}x^n+\sum_{n=1,3,5,\cdots}^{\infty}j(-1)^{\frac{n-1}{2}}\frac{1}{n!}x^n=\cos(x)+j\cdot\sin(x)$$

傅里叶级数

三角级数：由三角函数组成的函数项级数。

在数学上可以证明：任意一个周期为 T 的连续周期函数在一些限定条件下都可以表示为以下形式：

$$x(t)=\frac{a_0}{2}+\sum_{k=1}^{\infty}\left[a_k\cos\left(k\omega t\right)+b_k\sin\left(k\omega t\right)\right]\begin{cases}a_k=\frac{2}{T}\int_0^T x(t)\cos\left(k\omega t\right)dt\\ b_k=\frac{2}{T}\int_0^T x(t)\sin\left(k\omega t\right)dt\end{cases}$$

三角函数通过欧拉公式 $e^{j\theta}=cos\theta+isin\theta$ 变换为复数形式：

$$x(t)=\sum_{-\infty}^{\infty}a_k e^{jk\omega_0 t},\ a_k=\frac{1}{T}\int_0^T x(t)e^{-jk\omega_0 t}$$

连续周期傅里叶级数的收敛条件，即狄利赫里条件（无穷级数，∴存在收敛问题）

1. 任何周期内，信号必须绝对可积
2. 任意有限区间内，信号具有有限个起伏变化
3. 任意有限区间内，只有有限个不连续点，在这些不连续点上，函数是有限的。

离散时间周期信号的傅里叶级数是有限项级数，∴不存在收敛问题

$$\begin{cases}x[n]=\sum_{k=<N>}a_k e^{jk\omega_0 n}\\ \mathrm{a}_k=\frac{1}{N}\sum_{n=<N>}x[n]e^{-jkw_0 n}\end{cases}\rightarrow y[n]=\sum a_k H(e^{jkw_0})e^{jkw_0 n}\quad \sum_{n=<N>}e^{jk(\frac{2\pi}{N})n}=\begin{cases}N,k=0,\pm N,\pm 2N,...\\ 0,\end{cases}$$

∴傅里叶级数可用来构造任何离散时间周期信号，以及在实践中具有重要意义的几乎所有连续时间周期信号。

2 概率论

2.1 随机变量

问题

全概率公式…
贝叶斯公式
概率为 0 的事件可 能发生吗？
什么是贝叶斯定理？
一个家庭第一胎是女儿，第二胎男孩概率（来年可能是三胎了）？

什么是 0・1 分布？

泊松分布？
二项分布？
均匀分布？
概率密度函数？
多个高斯变量的和服从什么分布
大数定律及其研究意义
中心极限定理及其研究意义

数学期望的性质？
期望和方差关系
独立和相关性的区别与联系？相关系数？
判断随机事件独立的方法？
协方差的物理意义
一阶矩、二阶矩，方差是几阶矩？

2.1.1 概率论基本概念

事件的频率：在若干次试验中某个事件出现的次数占试验总次数的比例；
概率:概率是某个数值，该数值是该非确知系统的输出事件频率在不同时间与地点所遵循的绝对规律。概率是频率所遵循的绝对规律，频率是在概率这个绝对规律之上，叠加一些小扰动之后的具体表现；所以说概率是频率的理想情况，频率是概率叠加扰动后的具体表现
概率的理解：

概率不能用事件发生的可能性大小来定义，因为事件发生可能性大小是用概率来定义（度量）的，如果用事件发生可能性大小来定义概率，就犯了循环定义的错误。

必然事件的概率为 1，不可能事件的概率为 0。而反过来的说法却是不成立的。这是很容易误解的地方。当考虑的概型为古典概型时，概率为零的事件一定是不可能事件；当考虑的概型是几何概型时，概率为零的事件未必是一个不可能事件。

全概率公式： $P(Y)=\sum_{i=1}^{n}P(X_i)P(Y\mid X_i)$：已知各种原因的概率，各种原因导致某个结果的概率，求结果的概率。

贝叶斯公式： $P(X_{\mathrm{i}}\mid Y)=\dfrac{P(X_iY)}{P(Y)}=\dfrac{P(X_i)P(Y\mid X_i)}{\sum_{j=1}^{n}P(X_j)P(Y\mid X_j)}$：已知原因的概率，各种原因导致结果的概率，求结果发生条件下，某个原因的概率。

2.1.2 随机变量及其分布

1. **分布的基本概念**

分布律：表示离散型随机变量 x 取各个可能值的概率
概率密度：用于描述概率分布，大多数随机变量的概率分布函数无法写出来，∴用概率密度来表示。

分布函数： $F(x)=P\left\{X\le x\right\},-\infty<x<\infty$

常见分布

0-1 分布（又称伯努利分布）：是概率论中一种二元随机变量分布，即随机变量 x 只取 0，1 两个值

$$P\{X=k\}=p^k(1-p)^{1-k},k=0,1$$

二项分布：n 次独立重复伯努利实验得到的分布；伯努利实验是指每次试验得到结果 A 的概率为 p，结果 $\overline{A}$ 的概率概率为 $1-p$，这两个结果为互斥事件。在这样的条件下，试验成功 k 次的概率符合二项分布

$$P\{X=k\}=C_n^k p^k q^{n-k}$$

泊松分布：随机变量 X 所有可能取的值为 0，1，2，…，而取各个值的概率为 $P(X=\mathrm{k})=\dfrac{\lambda^k e^{-\lambda}}{k!}$

均匀分布： $f(x)=\begin{cases}\dfrac{1}{b-a}, a<x<b \\ 0,\ \text{Æä Ëû}\end{cases}$ 每个数值的概率密度相等的概率分布。

指数分布： $f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, \mathrm{x}>0 \\ 0,\ \text{Æä Ëû}\end{cases}$

统计学四大分布

正态分布（高斯分布）： $f(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{(x-E(X))^2}{2(D(X))^2}}$ 正态随机变量的线性组合仍为正态随机变量。

χ^2 **分布：** $X_1,\ldots X_{\mathrm{n}}$ 是来自总体 $N(0,1)$ 的样本，$\chi^2=X_1^2+X_2^2\ldots+X^2{}_{\mathrm{n}}$

t **分布：** $\mathrm{X}\sim\mathrm{N}(0,1)$， $Y\sim\chi^2(\mathrm{n})$ $t=\dfrac{X}{\sqrt{Y/n}}$

F **分布：** $\mathrm{u}\sim\chi^2(\mathrm{n}_1),\mathrm{v}\sim\chi^2(\mathrm{n}_2),u,v$ 相互独立 $F=\dfrac{\mathrm{u}/n_1}{u/n_2}$

多维随机变量及其分布

边缘概率密度：x,y 各自的概率密度。$f_X(\mathrm{x})\int_{-\infty}^{\infty}f(x,y)dy$ x 的概率密度就是联合概率密度 $f(x,y)$ 在 y 上的积分。$f_Y(y)\int_{-\infty}^{\infty}f(x,y)dx$ ，y 也同理。

条件概率密度： $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$

2.随机变量的数字特征

均值（数学期望）：描述随机变量 x 取值的平均大小 $E(X)=\int_{-\infty}^{\infty}xf(x)dx$

假设 c 为常数，则 $E(c)=c$，$E(cX)=cE(X)$

$E(X+Y)=E(X)+E(Y)$，\$X,Y\$相互独立时 $E(XY)=E(X)E(Y)$

方差：用来度量随机变量 x 与均值\$E（x）\$的偏离程度 $D(X)=E\left\{[X-E(X)]^2\right\}$

协方差： $\mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}=E(XY)-E(X)E(Y)$

单个随机变量的方差表示它分布的分散程度，两个随机变量的协方差就可以理解成它们一致的分散程度有多大，即 度量各个维度偏离其均值的程度。

相关系数： 表征 X,Y 之间线性关系的紧密程度 $\rho_{XY}=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

$|\rho_{XY}|$ 大：X,Y 线性相关程度好

$|\rho_{XY}|$ 小：X,Y 线性相关程度差

$|\rho_{XY}|=0$：X,Y 不相关

相关性是就线性关系而言，独立性是根据一般关系而言。独立一定不相关，不相关不一定独立。另外对于正态分布，不相关等价于相互独立。

独立： P(A ,B)=P(A) P(B) 同时发生的概率等于各自发生的概率的乘积;或者 A 发生的概率不受 B 是否发生的影响，即 $P(A|\ B)=P(A)$

矩： $\mu_n=E(X-c)^n$

离散型： $\mu_n=\sum_{i=1}^{\infty}\left(x_i-c\right)^n P\left(x_i\right)$　　连续型：$\mu_n=\int_{-\infty}^{\infty}(x-c)^n f(x)dx$

特别地，当 $c=0$ 时称之为原点矩，当 $c=E(X)$ 时称之为中心矩。

每一阶矩都告诉我们这个分布的一些信息：一阶原点矩就是均值；二阶中心矩就是方差

2.1.3 大数定律及中心极限定理

大数定律（定理）

大数定律（定理）：**在试验不变的条件下，重复试验多次，随机事件的频率近似于它的概率**。有的教材对“定律”和“定理”两个词进行了区分，大数定律并不是经验规律，它是一种自然规律因而通常不叫定理而是大数“定律”。而我们说的大数定理通常是经数学家证明并以数学家名字命名的大数定理。

切比雪夫大数定理： $x_1,x_2,\cdots,x_n,\ldots$ 是一列相互独立的随机变量(或者两两不相关)，他们分别存在期望 $E\left(x_k\right)$ 和方差 $D\left(x_k\right)$。若存在常数 c 使得：$D\left(x_k\right)\le C(k,=1,2\cdots n)$

则对于任意小的正数满足 $\lim\limits_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}x_k-\frac{1}{n}\sum_{k=1}^{n}Ex_k\right|<\varepsilon\right\}=1$

伯努利大数定律：只要重复独立试验次数足够大，事件发生的频率与时间发生的概率就几乎相等（频率的稳定性）。所以试验次数很大时，可用事件发生的频率代替事件的概率。设 μ_n 是 n 次独立试验中事件 A 发生的次数，且事件 A 在每次试验中发生的概率为 p，则对任意正数 ε，有：$\lim\limits_{n\to\infty}P\left(\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right)=1$

辛钦大数定律： 前 n 个独立同分布变量的算数平均值 $\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ 依概率收敛于单个随机变量的均值

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} a_i - \mu\right| < \varepsilon\right) = 1$$

中心极限定理：

大数定理示了大量随机变量的平均结果，但没有涉及到随机变量的分布的问题，中心极限定理在此基础上做了补充。设随机变量 $X_1, X_2, \ldots\ldots X_n, \ldots\ldots$ 独立同分布，并且具有有限的数学期望和方差：$E(X_i) = \mu$, $D(X_i) = \sigma^2$，则对任意分布函数满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

在一般情况下，当独立随机变量的个数不断增加时，其和的分布逼近于正态分布。因此在实际中，大量独立随机变量综合在一起才会造成影响，个别随机变量的影响很小。

2.2 随机过程

问题

什么是自相关函数和 互相关函数？
平稳随机信号
平稳随机过程的定义
广义平稳随机过程
狭义平稳随机过程
广义平稳和严格平稳的区别
严平稳、宽平稳？
功率谱密度和什么时间函数成什么关系
随机信号这门课讲了啥
马尔可夫过程最大特点
学过什么独立增量过程？（泊松过程、维纳过程，以及具体过程的一些定义和特点）

2.2.1 随机过程的基本概念

随机过程：定义域具有时间意义的随机变函 $X(t)$，随机过程在任意时刻的值都是一个随机变量。因为有无穷个时刻，所以随机变函 $X(t)$ 又叫做无穷维随机变量。虽然对于每一个固定的时刻 t，随机过程 $X(t)$ 都是一个随机变量，但是随机过程绝对不是这些随机变量在时间上的简单堆砌。不同时刻的随机变量之间往往都是强相关的，也即一者确定，另一者也已经确定。

随机变量样本：

对于随机变量 X，观察到足够多的样本 $x_1,x_2,...,x_n$，可以得到 X 的概率分布；

对于随机过程 $X(t)$ 也是类似，观察做够多的样本函数 $x(t)_1,x(t)_2,...,x(t)_n$，可以得到随机过程的概率分布

随机信号：信号一般都是随机的，有以下几个关键原因。

客观世界中干扰和噪声存在的必然性（加性噪声：如空杂波、地杂波、海杂波等；乘性噪声：多径效应干扰等）；

传输信息的不可预知性（如语音、通信、敌方目标行踪等）；

对客观世界认识的局限性（如对水声、地声的传播机理掌握不够等）

随机过程是数学概念，随机信号可以看做是一个随机过程，当然随机过程不只有随机信号。

随机过程的数字特征

均值函数：$\mu_X(t)=E[X(t)]$ 所有样本函数在时刻 t 的函数值的平均值

方差函数：$\sigma_X^2(t)=D_X(t)=\mathrm{Var}[X(t)]=E\left\{\left[X(t)-\mu_X(t)\right]^2\right\}$

自相关函数：$R_{XX}\left(t_1,t_2\right)=E\left[X\left(t_1\right)X\left(t_2\right)\right]$

互相关函数：$R_{XY}\left(t_1,t_2\right)=E\left[X\left(t_1\right)Y\left(t_2\right)\right]$

自协方差函数：$C_{XX}\left(t_1,t_2\right)=\mathrm{Cov}\left[X\left(t_1\right),X\left(t_2\right)\right]=E\left\{\left[X\left(t_1\right)-\mu_X\left(t_1\right)\right]\left[X\left(t_2\right)-\mu_X\left(t_2\right)\right]\right\}$

互协方差函数：$C_{XY}\left(t_1,t_2\right)=E\left\{\left[X\left(t_1\right)-\mu_X\left(t_1\right)\right]\left[Y\left(t_2\right)-\mu_Y\left(t_2\right)\right]\right\}=R_{XY}\left(t_1,t_2\right)-\mu_X\left(t_1\right)\mu_Y\left(t_2\right)$

2.2.2 随机过程的性质

平稳性：

宽平稳：信号的主要统计特性不随时间的推移而变化

严格平稳（强平稳）：信号的全部统计特性不随时间的推移而变化

广义平稳：信号的均值，相关系数不随时间的推移而变化。

狭义平稳：任意 n 维分布不随时间的推移而变化。如一维分布与 t 无关，二维分布只与时间间隔有关。

严格平稳一定是广义平稳，广义平稳不一定是严格平稳。广义平稳的高斯信号必定是严格平稳。

循环平稳：随机过程不平稳，但它有某种周期性，时间推移某个周期的整数倍时，统计特性保持不变。

维纳-辛钦定理：平稳随机过程的自相关函数和功率谱密度是一对傅里叶变换对。意义：给出了用自相关函数表示功率谱密度的方法。

各态历经性（又称埃尔哥德性，遍历性）： 随机过程的任一个样本函数的时间平均，等于随机过程的统计平均。物理含义：随机过程的任一次实现都经历了随机过程的所有可能状态。所以由一次样本函数的统计特性，就能得到整个随机过程的统计特性。

2.2.3 典型随机过程（信号）

高斯随机过程：

① **若线性系统的输入是高斯过程，那么输出也是高斯过程**

② 高斯过程的 n 维分布只取决于各个随机变量的均值，方差，协方差（只需要研究数字特征）

随机正弦信号；常用于电路

伯努利随机序列：最简单的随机序列，常用于数字通信传输二进制比特流

马尔可夫过程：“下一时刻的状态只与当前状态有关，与上一时刻状态无关”的性质，称为无后效性或者马尔可夫性。而具有这种性质的过程就称为马尔可夫过程。

马尔科夫链：信源发出一个符号，信源所处的状态即发生改变，这些状态的变化组成了马氏链

马尔科夫链的遍历性：对固定状态 i，不管链在某一时刻从什么状态出发，通过长时间的转移，到达状态 i 的概率都趋近于一个值→系统经过长时间转移后会达到平衡状态。

独立增量过程： 互不重叠的区间上，状态的增量相互独立。例如：$\mathrm{x}(t_1)-\mathrm{x}(t_0),\mathrm{x}(t_2)-\mathrm{x}(t_1)\ldots$ 相互独立，则 $x(t)$ 为独立增量过程。独立增量过程是一种特殊的马尔可夫过程。

泊松过程： 事件在时间轴上是离散的；事件之间的时间间隔是独立的；事件到达的时间间隔服从指数分布；服从这三个条件的事件发生的概率满足泊松分布。数据包的到达的数目常建模为泊松过程。泊松过程的增量服从泊松分布

维纳过程： 如果一个过程是独立增量过程过程，并且在单位时间变量变化的期望值服从期望为 0，方差为 1 的正态分布，则这个过程为维纳过程。

3 线代

3.1 线性代数基本概念算术：一般就是指四则运算

代数： 代数的英语为 algebra ，源于阿拉伯语单字“al-jabr”后来清代学者华蘅芳和英国人傅兰雅合译英国瓦里斯的《代数学》，卷首有“代数之法，无论何数，皆可以任何记号代之”，说明了所谓“代数”，就是用符号来代表数的一种方法。

初等代数

算术中只有数字与其运算（如：加、减、乘、除），使用字母符号诸如 x、y 或 a、b 等表示数字，习惯上用前者表示未知数与变量，用后者表示任意的已知数。初等代数使得算术等式（或不等式）可以被描述成命题或定理（如：$\forall$ 实数 a 和 $b, a+b=b+a$），因此这是系统化学习实数性质的第一步。

它允许探究数量之间的数学关系的可能（如“若你卖了 x 张票，你的收益将有 $3x+10$ 元”）。

它允许涉及未知的数字。在一个问题的内容里，变量或许代表某一还不确定，但可以通过解方程解出来的值。

高等代数

初等代数学向两个方向进一步发展：求解未知数更多的一次方程组（线性代数），未知数次数更高的高次方程（多项式代数），因此线性代数是高等代数的分支之一。

线性代数

线性代数的研究最初出现于对行列式的研究上。行列式当时被用来求解线性方程组。莱布尼茨在 1693 年使用行列式。随后，加布里尔·克拉默在 1750 年推导出求解线性方程组的克莱姆法则。然后，高斯利用高斯消元法发展出求解线性系统的理论。

线性方程组的求解

线性方程组是线性代数的核心。在求解的过程中引入了矩阵、向量和向量空间，线性变换等概念。

线性方程组的两个基本问题

1：是否至少有一个解？

2：有解的话，有几个解？

3.2 矩阵

问题

克莱姆法则，适用解决哪类问题（解方程组）？

矩阵特征值、特征向量的概念

秩的概念

矩阵求秩

矩阵的秩、矩阵的行向量组的秩、矩阵的列向量组的秩的关系？（相等）如何证明？

矩阵的秩的概念，研究其的意义

特征向量分解？

n 阶矩阵 A 可对角化的充要条件？

判断矩阵是否可逆的方法？（四种）

有没有学过线性空间？线代学过哪些内容？

矩阵的奇异值分解的三个矩阵分别是什么矩阵？

矩阵的秩物理意义？

线性方程组的解，Ax=b，Am×n 分别为长矩阵（m＞n）和扁矩阵（n＞m）？怎么确定哪个解是最优解？

线性方程组什么时候无解

正定矩阵的概念。

3.2.1 矩阵的基本概念

矩阵：$A^{m\times n}$ 就是由 $m\times n$ 个数排成 m 行，n 列的数表，矩阵代数就是用 A 这个符号去代替数表运算。

常见矩阵：

可逆矩阵：$|A|\neq 0 \Leftrightarrow$ 可逆； n 阶矩阵的秩为 $n \Leftrightarrow |A|\neq 0$； 齐次方程组只有零解 $\Leftrightarrow |A|\neq 0$

行向量或列向量线性无关 $\Leftrightarrow |A|\neq 0$ ； 矩阵 A 的特征值不全为 0 $\Leftrightarrow |A|\neq 0$

单位矩阵：对角线上元素都是 1；

奇异矩阵（针对方阵）：|A|= 0

对称矩阵 $A^T = A$：伴随矩阵：方阵 A 的行列式中各个元素的代数余子式构成的矩阵 $A^* = A^{-1}|A|$

矩阵的初等变换：对换两行（列）；以数 k 乘以某行（列）中的所有元素；把某一行（列）所有元素乘以 k 倍加到另一行

3.2.2 求解方程组

1. **对于系数矩阵为方阵的情况**

矩阵的行列式：**只有方阵才有行列式！行列式是一个数值**，具体来说矩阵的行列式就是一个平行多面体的（定向的）体积，这个多面体的每条边对应着对应矩阵的列。

$$\det\left(\begin{vmatrix} a & b \\ c & d \end{vmatrix}\right) = ad - bc$$

1. 对换行列式的两行（列）行列式变号
2. 行列式的某一行（列）所有元素都乘以同一数 k，等于用 k 乘以此行列式
3. 行列式的某一行（列）所有元素都乘以同一数加到另一行（列）对应的元素上，行列式不变
4. n 阶行列式的定义

三角行列式：主对角线以上（以下）全为 0

对角行列式：主对角线上下均为 0

余子式：元素 a_{ij} 所在的第 i 行，j 列去掉后的行列式

代数余子式：在余子式的基础上乘以 $(-1)^{i+j}$

行列式的展开法则：行列式等于任一行（列）的各元素与其对应的代数余子式乘积的和。

克拉默法则（克莱姆法则）：对于一个线性方程组 $Ax = c$

若系数矩阵 A 的行列式不等于 0，即 $\det A \neq 0$，线性方程组有唯一解： $x_i = \dfrac{\det(A_i)}{\det(A)}$

其中：A_i 是列向量 c 取代了 A 的第 i 列后得到的矩阵

若 $\det A = 0$，对于非齐次方程组（c≠0）无解或无穷多解；对于齐次方程组有零解或无穷多解

注：克莱姆法则适用于 n 个未知数和 n 个方程组的情况。

3.2.2 矩阵的秩

矩阵的秩的定义

行空间：矩阵行向量组成的所有线性组合的集合；假设每一行有 n 个元素，则$rowA$是 R^n 的子空间

行空间：矩阵列向量组成的所有线性组合的集合；

秩：线性无关的行向量的最大个数（行空间的维数）；线性无关的列向量的最大个数（列空间的维数）

矩阵的秩的研究意义

可以判断方程组解的个数

无解的充要条件： $rank(A)<rank(A,b)$

有唯一解的充要条件：$rank(A)=rank(A,b)=n$

有有限多解的充要条件： $r(A)=r(A,b)<n$

3.2.3 矩阵 vs 线性变换

满足可加性 $f(\mathbf{v}_1+\mathbf{v}_2)=f(\mathbf{v}_1)+f(\mathbf{v}_2)$ 和齐次性 $f(a\cdot\mathbf{v})=a\cdot f(v)$ 的变换叫做线性变换。

拉伸（改变向量长度但是不改变方向）和旋转（改变向量方向但是不改变长度）、投影等都是线性变换（可以在草稿纸上画画图）。

矩阵是对向量做线性变换($A(\mathbf{x}+\mathbf{y})=A\mathbf{x}+A\mathbf{y}$, $A(a\mathbf{x})=aA\mathbf{x}$)，矩阵的逆就是进行复原。当然变换前后要能一一对应才能进行复原，所以有些矩阵没有逆，这些矩阵所作的线性变换使得前后不能一一对应。

对于平面中的任一向量 $\begin{bmatrix}x\\y\end{bmatrix}$， $\begin{bmatrix}a&c\\b&d\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=x\begin{bmatrix}a\\b\end{bmatrix}+y\begin{bmatrix}c\\d\end{bmatrix}=\begin{bmatrix}ax+cy\\bx+dy\end{bmatrix}$

$\begin{bmatrix}a&0\\0&a\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}$:将向量拉伸 a 倍;对角矩阵的几何意义就是单纯坐标值的拉伸; $a=1$ 就是对向量不做变化; $a=0$ 就是将所有点都压缩到了原点。

$\begin{bmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}$： 逆时针旋转 θ 度;

$\begin{bmatrix}0&0\\0&1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}$： 向 y 轴投影（满足 $P^2=P$ 的矩阵 P 都可以看做是投影矩阵） ;

3.2.4 特征值与特征向量

只有方阵才能计算特征值和特征向量（不是方阵的话可以补 0）;方阵总有特征值，因为总有特征多项式（特征方程），但不是所有方阵都有实数特征解;

几何意义：矩阵 A 是对向量 $\mathbf{x}$ 做了一个线性变换，**但对于某些向量，矩阵 A 不能改变向量的方向，仅仅是做了拉伸,这些向量就叫做特征向量,拉伸的倍数就叫做特征值**。即 特征向量在线性变化下能保持方向不变 $A\mathbf{x}=\lambda\mathbf{x}$。

特征方程： $det(A-\lambda I)=0$： λ 是方阵 A 的特征值的充要条件是： λ 是特征方程的根

特征多项式： $n\times n$ 矩阵 A 的特征方程是 n 次多项式，称为 A 的特征多项式。

特征值的结论：

矩阵 的秩等于其非零特征值的个数，反过来， n 阶矩阵特征值为 0 的个数等于 n 减去秩的大小

特征值之和等于矩阵的迹： $\lambda_1+\lambda_2+\ldots+\lambda_n=$ trace $=a_{11}+a_{22}+\ldots+a_{nn}$

特征值之积等于矩阵的行列式：$\lambda_1\lambda_1\ldots\lambda_n=\det(A)$

相似矩阵： 存在 B 和可逆矩阵 P 使得 $P^{-1}AP=B$,则称 A 和 B 相似；相似方阵具有相同的特征多项式，进而拥有相同的特征值。

对角化：若方阵 A 相似于对角阵，则称 A 可对角化；$n\times n$ 矩阵 A 可对角化的充要条件是 A 有 n 个线性无关的特征向量，此时对角矩阵上的值是矩阵 A 相对于特征向量的特征值。因此可以将方阵 A 分解为对角矩阵 $A=PDP^{-1}$

不同特征值对应的特征向量线性无关，很显然，若 n 个特征值都不相同，那么一定有 n 个线性无关的特征向量，A 就一定可对角化。

特征值分解（谱分解）： $A=Q\sum Q^{-1}$,其中 Q 是矩阵 A 的特征向量组成的矩阵；Σ 是对角矩阵，对角线上的值是 A 的特征值

奇异值分解： 上述对角化适用于方阵，对于一般的 $m\times n$ 矩阵可将 A 分解为 $A=U\Sigma V^{\mathrm{T}}$，其中 U 是 $m\times m$ 的正交矩阵，V 是 $n\times n$ 的正交矩阵；$\Sigma=\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ D 是 $r\times r$ 的对角矩阵，对角线元素是前 r 个奇异值，$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$，$r$ 为矩阵 A 的秩；

A 的奇异值是 A^TA 的特征值的平方根，即 $\sigma_i=\sqrt{\lambda_i}$

3.2.5 对称矩阵和二次型

对称矩阵： 满足 $A^T=A$ 的矩阵叫做对称矩阵，主对角线任意，两边的数成对出现；对称矩阵一定是方阵。

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}$$

一个 $n\times n$ 矩阵 A 可正交对角化的充要条件是 A 是对称矩阵

二次型：方阵二次型定义为：$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}=\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_ja_{ij}=\sum_{i=1}^{n}a_{ii}x_i^2+\sum_{i=1,i\neq j}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j=\sum_{i=1}^{n}a_{ii}x_i^2+\sum_{j=i+1}^{n-1}\left(a_{ij}+a_{ji}\right)x_ix_j$

对于任意一个二次型函数，存在许多矩阵 A，二次型相同，但是只有一个唯一的对称矩阵,所以为了讨论的唯一性，一般只考虑实对称矩阵 A 或复共轭对称矩阵（Hermitian）

正定矩阵： $x^{\mathrm{H}}Ax>0$，$\det(\mathrm{A})>0$,所有特征值取正实数

半正定矩阵： $x^{\mathrm{H}}Ax\geqslant0$ $\det(\mathrm{A})\geq\mathbf{0}$，所有特征值取非负数

3.2.6 合同、相似、等价（了解就行）

一、矩阵等价、相似和合同之间的区别:

1、等价、相似、合同三者都是等价关系。

2、矩阵相似或合同必等价，反之不一定成立。

3、矩阵**等价**，只需满足两矩阵之间可以通过一系列可逆变换，也即若干可逆矩阵相乘得到。

$\boldsymbol{PAQ}=\boldsymbol{B}$，其中$\boldsymbol{P}$、$\boldsymbol{Q}$可逆，则$\boldsymbol{A}$、$\boldsymbol{B}$等价，表示为$\boldsymbol{A}\cong\boldsymbol{B}$.

4、矩阵**相似**，则存在可逆矩阵$\boldsymbol{P}$，使得$\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{B}$，表示为$\boldsymbol{A}\sim\boldsymbol{B}$.

5、矩阵**合同**，则存在可逆矩阵$\boldsymbol{P}$，使得$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}=\boldsymbol{B}$，表示为$\boldsymbol{A}\simeq\boldsymbol{B}$.

6、当上述矩阵$\boldsymbol{P}$是**正交矩阵**(即$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}=\boldsymbol{E}$，$\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{P}^{-1}$)时，则有$\boldsymbol{A}$、$\boldsymbol{B}$之间既满足相似，又满足合同关系。

二、矩阵等价、相似、合同之间联系:

1、矩阵等秩是相似、合同、等价的必要条件，相似、合同、等价是等秩的充分条件。

2、矩阵等价是相似、合同的必要条件，相似、合同是等价的充分条件。

3、矩阵相似、合同之间没有充要关系，存在相似但不合同的矩阵，也存在合同但不相似的矩阵。

4、总结起来就是：相似$\Rightarrow$等价，合同$\Rightarrow$等价，等价$\Rightarrow$等秩。

相关性质 / 矩阵间的关系	定义	满足性质	相关结论
矩阵**等价** $\boldsymbol{A}\cong\boldsymbol{B}$	若矩阵$\boldsymbol{A}$经过有限次的初等变换（行、列变换）变到矩阵$\boldsymbol{B}$，则称 A 与 B 等价	是等价关系即满足反身性，对称性和传递性	$\boldsymbol{A}\cong\boldsymbol{B} \Leftrightarrow r(\boldsymbol{A})=r(\boldsymbol{B})$ $\Leftrightarrow$ $\boldsymbol{A}$经过有限次的初等变换得到矩阵$\boldsymbol{B}$. $\Leftrightarrow$ $\exists$可逆矩阵$\boldsymbol{P}$、$\boldsymbol{Q}$，使得$\boldsymbol{PAQ}=\boldsymbol{B}$
矩阵**相似** $\boldsymbol{A}\sim\boldsymbol{B}$	设$\boldsymbol{A}$、$\boldsymbol{B}$是n阶矩阵，若存在可逆矩阵$\boldsymbol{P}$，使得$\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{B}$，则称矩阵$\boldsymbol{A}$、$\boldsymbol{B}$相似	是等价关系即满足反身性，对称性和传递性	若$\boldsymbol{A}$、$\boldsymbol{B}$相似，则 ① $\boldsymbol{A}^{\mathrm{T}}$、$\boldsymbol{B}^{\mathrm{T}}$相似; ② 若$\boldsymbol{A}$、$\boldsymbol{B}$可逆，则$\boldsymbol{A}^{-1}\sim\boldsymbol{B}^{-1}$; ③ $\boldsymbol{A}$、$\boldsymbol{B}$有相同的特征值和特征多项式; ④ $\lvert\boldsymbol{A}\rvert=\lvert\boldsymbol{B}\rvert$，$r(\boldsymbol{A})=r(\boldsymbol{B})$，$tr(\boldsymbol{A})=tr(\boldsymbol{B})$.
矩阵**合同** $\boldsymbol{A}\simeq\boldsymbol{B}$	设$\boldsymbol{A}$、$\boldsymbol{B}$是n阶矩阵，若存在可逆矩阵$\boldsymbol{P}$，使得$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}=\boldsymbol{B}$，则称矩阵$\boldsymbol{A}$、$\boldsymbol{B}$合同	是等价关系即满足反身性，对称性和传递性	① 任一对称矩阵合同于一个对角矩阵; ② 两同阶实对称矩阵相似，则两矩阵必合同; ③ 若$\boldsymbol{A}\simeq\boldsymbol{B}$，则$r(\boldsymbol{A})=r(\boldsymbol{B})$，$\boldsymbol{A}$、$\boldsymbol{B}$的正、负特征值的个数相同。

3.3 向量和向量空间

问题

判断向量线性相关/无关的方法

线性相关的概念和性质

空间的概念（回答三个点：维数对应着生成空间的线性无 关向量个数；定义运算数乘和加法；运算具备完备性）

3.3.1 向量空间的基本概念

n 维向量空间 R^n 的每一个向量都由 n 个成分构成，R^3 空间的每一个向量都有 3 个成分 $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$；向量空间：集合 V 对于其中所包含的向量的加法，数乘两种运算封闭。

3.3.2 线性无关和基

向量组： 同维度的列（行）向量组成的集合叫做向量组 $\left\{\mathbf{v}_1,\cdots,\mathbf{v}_p\right\}$

向量组的线性组合： $c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_p\mathbf{v}_p$

线性无关： $c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_p\mathbf{v}_p=\mathbf{0}$ 只有零解，即 $c_1=0,\cdots,c_p=0$

线性相关： 存在不全为零的系数使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_p\mathbf{v}_p=\mathbf{0}$，此时向量组中的至少有一个向量都能由另一个向量组线性表达。

A 线性相关，在 A 的基础上增加一个向量也线性相关；线性无关则减少一个向量也线性无关
A 线性无关，在 A 的基础上增加一个向量却线性相关，那么增加的那个向量一定能由 A 唯一表示
m 个 n 维向量组成的向量组，n<m 时，一定线性相关。

基和维数： $\alpha_1,\alpha_2,\cdots,\alpha_n\in R^n$ 线性无关且任意 $\eta\in R^n$ 均可由 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性表示，则 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 为 R^n 的一组基；向量空间基的个数就是向量空间的维数；只含一个零向量的空间称为零空间，规定零子空间的维数为 0，零空间不存在基。.