一、 高数

1. 映射

   两个非空集合中元素的对应关系

2. 函数

   设有两个变量 x 和 y，如果变量 x 在变化范围 D 内任取一个数值时，按照一定的对应法则 f，变量 y∈M 有确定的数值与之对应，则称变量 y 是 x 的函数(function)。记为 y=f(x)

   自变量，因变量。映射关系、定义域

3. 数列

   如果按照某一法则，对每一 n∈N+，对应着一个确定的实数 xn，则得到一个序列。这一序列叫做数列，记为{xn}，其中第 n 项{xn}叫做数列的一般项.

4. 数列极限、函数极限

   数列极限：当 n 无限增大时，如果数列{xn}的一般项 xn 无限接近于常数 a，则常数 a 称为数列{xn}的极限，或称数列{xn}收敛于 a，记为

   若圆的内接正多边形的面积数列 {Pn} 稳定于某个数 a（当 n 无限增大时），则称 a 是该圆的面积。

   函数极限：自变量在变化的过程中，若对应的函数值无限接近一个确定的常数，那么该常数叫做这一变化过程中函数的极限。分趋于常数（去心邻域）和无穷大（单边）

   数列极限问题可以看做函数极限的特例，函数 f(x)如果限定定义域为自然数集，那么 x 趋于正无穷时 函数的极限也就是数列的极限了

5. 无穷小和无穷大

   无穷小：自变量在变化的过程中，以 0 为极限的函数

   无穷大：无穷小的倒数，自变量在变化的过程中，函数值不断增大的函数

   定义

   如果 $\lim \frac{\beta}{\alpha}=0$，那么就说 β 是比 α 高阶的无穷小，记作 $\beta=o(\alpha)$；

   如果 $\lim \frac{\beta}{\alpha}=\infty$，那么就说 β 是比 α 低阶的无穷小；

   如果 $\lim \frac{\beta}{\alpha}=c\neq 0$，那么就说 β 与 α 是同阶无穷小；

   如果 $\lim \frac{\beta}{\alpha^k}=c\neq 0, k>0$，那么就说 β 是关于 α 的 k 阶无穷小；

   如果 $\lim \frac{\beta}{\alpha}=1$，那么就说 β 与 α 是等价无穷小，记作 $\alpha\sim\beta$.

6. 柯西极限存在准则：

   两个极限准则

   准则一：如果数列{x_n}，{y_n}，{z_n}满足下列条件

   A. 从某项起，即存在 $n\in N$，当 $n \rangle n_0$ 时，有 $y_n \le x_n \le z_n$

   B. $\lim\limits_{n\to\infty} y_n = a, \lim\limits_{n\to\infty} z_n = a$

夹逼准则

准则二：单调有界的函数必有极限.

柯西极限存在准则 数列 $\{x_n\}$ 收敛的充分必要条件是：对于任意给定的正数 ε，存在正整数 N，使得当 $m>N, n>N$ 时，有

$$|x_n - x_m| < \varepsilon.$$

给出了数列收敛的充分必要条件

用于判断某个式子是否收敛的冲要条件

数列收敛的充分必要条件是，该数列中的元素随着序数的增加而愈发靠近，即足够靠后的任意两项都无限接近。

7. 函数一重要性质；连续性，一致连续性

现在我们可对连续性的概念这样描述：如果当 Δx 趋向于零时，函数 y 对应的增量 Δy 也趋向于零，即：$\lim\limits_{\Delta x \to 0} \Delta y = 0$，那么就称函数 $y = f(x)$ 在点 x_0 处连续。

持续变化，当自变量变化很小时，函数值的的变化也很微小。

只要自变量的两个数值接近到一定程度，就可使对应的函数值达到所指定的接近程度。在闭区间上连续那么就可以一致连续

8. 间断点

定义：我们把不满足函数连续性的点称之为间断点.

第一类间断点：如果 x_0 是函数 $f(x)$ 的间断点，且其左、右极限都存在

可去间断点：左、右极限虽然都存在，且相等

跳跃间断点：这我们可以看出函数左、右极限虽然都存在，但不相等，故函数在点 $x=0$ 是不存在极限

其余为第二类（无穷和振荡）。

左右相等为可去间断点，不等为跳跃间断点

无穷间断点：$y = \tan x$ 在 $x = \dfrac{\pi}{2}$ 处没有定义。$\lim\limits_{x \to \frac{\pi}{2}} \tan x = \infty$，

振荡间断点：函数 $y = \sin\dfrac{1}{x}$ 在点 $x=0$ 处没有定义；故当 $x \to 0$ 时，函数值在-1 与+1 之间变动无限多次

初等函数的连续性：基本初等函数在它们的定义域内都是连续的（基本初等函数包括幂函数 指数函数，对数函数，三角函数，反三角函数）；

9. 最值/零点/介值定理

定理 1（最值定理）：若函数 $f(x)$ 在 $[a,b]$ 上连续，则它在 $[a,b]$ 上必有最大值和最小值。

定理 2（零点定理）：若函数 $f(x)$ 在 $[a,b]$ 上连续，且 $f(a)$ 与 $f(b)$ 异号，那么在开区间 (a,b) 内至少有一点 ξ，使 $f(\xi)=0$

定理 3（介值定理）：若函数 $f(x)$ 在 $[a,b]$ 上连续，且在这区间的端点取不同的函数值 $f(a)=A$，$f(b)=B$，那么，对于 A 与 B 之间的任意一个数 C，在开区间 (a,b) 内至少有一点 ξ，使得 $f(\xi)=C$

闭区间上的连续函数则是在其连续区间的左端点右连续，右端点左连续. 对于闭区间上的连续函数有几条重要的性质，下面我们来学习一下：在闭区间连续的函数必取得介于最大值 M 与最小值 m 之间的任何值。

函数在某区间内连续，端点处函数值异号，区间内必有一个零点，必有在之间的数。

10. 极限、可导与连续的关系

可导必连续（该点左右极限相等，且左导数等于右导数）

连续必极限存在

极限存在不一定连续（可去间断点）

连续不一定可导（y=|x|）

可导——连续——极限　正一定逆不一定

11. 隐函数

显函数：若函数 y 可以用含自变量 x 的算式表示

隐函数：自变量和因变量没有那么明显（分居两边）$F(x,y)=0$

12. 导数与微分

导数：$f^{'}(x_0)=\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\dfrac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$ 存在即可导，值为导数，反应 y 在 x0 处的变化率

函数 $y=f(x)$ 在 x_0 处的左右导数存在且相等是函数 $y=f(x)$ 在 x_0 处的可导的充分必要条件。

导数的几何意义：表示曲线 $y=f(x)$ 在点 M $(x_0,f(x_0))$ 处的切线的斜率，即 $f^{'}(x_0)=\tan\alpha$，其中 α 是切线的倾角。

切线：通过割线和无穷小量定义了切线。

导数：通过切线和无穷小量定义了导数，导数是曲线在某点处切线的斜率，导数的值等于微商。

微分：$A\Delta x$ 叫做函数 $y=f(x)$ 在点 x_0 相应于自变量增量 Δx 的微分，记作 dy，即：

$dy=A\Delta x$。微分 dy 是自变量改变量 Δx 的线性函数，dy 与 Δy 的差 $o(\Delta x)$ 是关于 Δx 的高阶无穷小量

可微可导二者充分必要

微分的几何意义：

可微函数 $y=f(x)$ 在点 x_0 点的微分是当自变量 x 取得增量 Δx 时，曲线 $y=f(x)$ 在点 x_0

的切线的纵坐标的增量。

多元函数：所有方向偏导才叫全导，全微必然全导。

13. 微分中值定理

反映导数局部性和函数整体性的关系，在闭区间连续，开区间可导

罗尔定理：若端点的函数值相同，那么区间内有一导数为 0。曲线弧为一条连续的曲线弧，除端点外处处有不垂直 x 轴的切线，且两端点的纵坐标相等，那么弧上至少有一点，曲线在该点的切线是水平的。

罗尔定理 如果函数 $f(x)$ 满足

(1) 在闭区间 $[a,b]$ 上连续；

(2) 在开区间 (a,b) 内可导；

(3) 在区间端点处的函数值相等，即 $f(a)=f(b)$，

那么在 (a,b) 内至少有一点 ξ $(a<\xi<b)$，使得 $f'(\xi)=0$.

拉格朗日中值定理：端点函数值不同，那么区间内必有一点导数与端点连线的导数相同。曲线上至少存在一点，过该点的切线的斜率和连接曲线的割线的斜率相同，或者说，必然存在至少一点可以做割线的平行线。

(1) 在闭区间 $[a,b]$ 上连续；

(2) 在开区间 (a,b) 内可导，

那么在 (a,b) 内至少有一点 ξ $(a<\xi<b)$，使等式

$$f(b)-f(a)=f'(\xi)(b-a)$$

0

一维运动，必然存在一点与平均速度相等。

柯西中值定理：那么在(a,b) 内至少有一点ξ，使等式[f(b)-f(a)]/[F(b)-F(a)]=f'(ξ)/F'(ξ)，参数方程形式，二维曲线运动，刚开 a 与 v 方向相反，绕圈，总有一刻二者同向。

柯西中值定理：

柯西中值定理 如果函数 $f(x)$ 及 $F(x)$ 满足

(1) 在闭区间 $[a,b]$ 上连续；

(2) 在开区间 (a,b) 内可导；

(3) 对任一 $x\in(a,b)$，$F'(x)\neq 0$，

那么在 (a,b) 内至少有一点 ξ，使等式

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

14. 洛必达法则

未定式：分子分母都为 0 或者都为无穷大

通过分子分母分别求导再求极限来确定未定式的方法称为洛必达法则

15. 泰勒公式（泰勒中值定理）

泰勒(Taylor)中值定理 2 如果函数 $f(x)$ 在 x_0 的某个邻域 $U(x_0)$ 内具有 $(n+1)$ 阶导数,那么对任一 $x\in U(x_0)$,有

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n(x), \tag{3-5}$$

其中

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}, \tag{3-6}$$

这里 ξ 是 x_0 与 x 之间的某个值.

函数在开区间有 N 阶导数，那么函数可以展开为关于 n 阶（x-x0）的多项式和一个余项的和

N=0 时，变成拉格朗日中值公式

16. 凹凸函数

定义 设 $f(x)$ 在区间 I 上连续,如果对 I 上任意两点 x_1,x_2 恒有

$$f\left(\frac{x_1+x_2}{2}\right)<\frac{f(x_1)+f(x_2)}{2},$$

那么称 $f(x)$ 在 I 上的图形是(向上)凹的(或凹弧);如果恒有

$$f\left(\frac{x_1+x_2}{2}\right)>\frac{f(x_1)+f(x_2)}{2},$$

那么称 $f(x)$ 在 I 上的图形是(向上)凸的(或凸弧).

从连续函数上任取两点，如果之间的函数值总是在两点连线的下面，那么就是凹函数；若在上面就是凸函数。二阶导数大于零则凹，否则凸

极大极小值：一阶导数看是否有驻点，二阶导数看极大极小，大于 0 极小值，大于 0 极大值

驻点：一阶导数的值为 0，不关注函数的单调性变化，如 $y=x^3$，在 $x=0$ 处导数为 0，是驻点，但没有极值，故不是极值点

17. 定积分、不定积分、变限积分、反常积分

定积分是函数图形在一个闭区间上和 x 坐标轴围成的面积

不定积分是用来求原函数，是一个表达式，导数的逆运算，连续函数一定有原函数

分部积分：$\int u\,dv=uv-\int v\,du.$ $\int x\cos x\,dx=x\sin x-\int \sin x\,dx,$

变限积分将定积分的上下限换成了变量 x，取不同的 x ，就会有不同的面积效果。

反常积分是定积分的推广，上下限是无穷或者间断点

18. 牛顿莱布尼兹公式（微积分基本定理）

将定积分和不定积分联系在一起。 $$\int_a^b f(x)dx=F(b)-F(a)$$

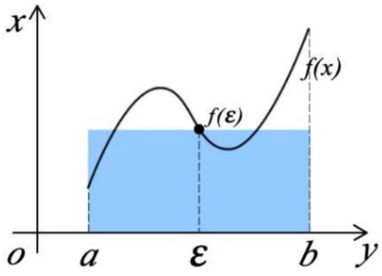
19. 积分中值公式

函数在闭区间连续，则必然存在一点，使得在闭区间上的积分值等于该点函数值乘区间长度。

即曲线所围面积等于矩形的面积

如果函数 f(x) 在积分区间[a, b]上连续，则在 [a, b]上至少存在一个点 ξ，使下式成立

$$\int_a^b f(x)dx = f(\xi)(b-a)$$ 其中(a≤ξ≤b)。



20. 数量积：力做功

向量积：做功的力矩

混合积： $[\boldsymbol{a}\,\boldsymbol{b}\,\boldsymbol{c}] = (\boldsymbol{a}\times\boldsymbol{b})\cdot\boldsymbol{c}$ 六面体体积

21. 两直线夹角：

$\cos\varphi = |\cos(\widehat{\boldsymbol{s}_1,\boldsymbol{s}_2})|$. 按两向量的夹角的余弦公式，直线 L_1 和直线 L_2 的夹角 φ 可由

$$\cos\varphi = \frac{|m_1m_2+n_1n_2+p_1p_2|}{\sqrt{m_1^2+n_1^2+p_1^2}\sqrt{m_2^2+n_2^2+p_2^2}} \qquad (4-5)$$

直线与平面夹角：

$$\sin\varphi = \frac{|Am+Bn+Cp|}{\sqrt{A^2+B^2+C^2}\sqrt{m^2+n^2+p^2}}.$$

22. 多元函数极限、连续性

性质 1（有界性与最大值最小值定理） 在有界闭区域 D 上的多元连续函数，必定在 D 上有界，且能取得它的最大值和最小值.

性质 1 就是说，若 $f(P)$ 在有界闭区域 D 上连续，则必定存在常数 $M>0$，使得对一切 $P\in D$，有 $|f(P)|\leqslant M$；且存在 P_1、$P_2\in D$，使得

$$f(P_1)=\max\{f(P)\mid P\in D\},\quad f(P_2)=\min\{f(P)\mid P\in D\}.$$

性质 2（介值定理） 在有界闭区域 D 上的多元连续函数必取得介于最大值和最小值之间的任何值.

性质 3（一致连续性定理） 在有界闭区域 D 上的多元连续函数必定在 D 上一致连续.

23. 偏导数

偏导数在求解时可以将另外一个变量看做常数，利用普通的求导数方法求解

几何意义：样,偏导数 fy(xo,yo)的几何意义是曲面被平面 x=xo 所截得的曲线在点的切线对 y 轴的斜率.

24. 方向导数、梯度

函数沿某一方向的变化率，梯度是方向导数的最大值。
梯度下降，由于梯度的方向是与方向导数的最大值方向（向上增长的方向）一致的，所以梯度下降实际是沿着梯度的反方向进行的，采用梯度下降的意义在于使模型始终沿着使函数值最快减小的方向去优化，而这个函数便是损失函数 loss。而且梯度在优化过程中并不是一成不变的，随着函数上点的移动，也就是优化过程的进行，方向导数时刻在变化，梯度也就时刻在变化，需要不断的调整梯度来完善模型。

25. 全微分

可微必要条件：偏导数存在，全微分为 $dz = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y.$

充要条件：偏导数连续

26. 雅可比行列式：

27. 多元函数极值

定理 2（充分条件） 设函数 $z = f(x,y)$ 在点 (x_0,y_0) 的某邻域内连续且有一阶及二阶连续偏导数，又 $f_x(x_0,y_0)=0, f_y(x_0,y_0)=0$，令

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

则 $f(x,y)$ 在 (x_0,y_0) 处是否取得极值的条件如下：

（1）$AC-B^2>0$ 时具有极值，且当 $A<0$ 时有极大值，当 $A>0$ 时有极小值；

（2）$AC-B^2<0$ 时没有极值；

（3）$AC-B^2=0$ 时可能有极值，也可能没有极值，还需另作讨论.

28. 积分中值定理

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

第一定理

如果函数 $f(x)$、$g(x)$ 在闭区间 $[a,b]$ 上连续，且 $g(x)$ 在 $[a,b]$ 上不变号，则在积分区间 $[a,b]$ 上至少存在一个点 ε，使下式成立：

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

第二定理

一、如果函数 $f(x)$、$g(x)$ 在闭区间 $[a,b]$ 上可积，且 $f(x)$ 为单调函数，则在积分区间 $[a,b]$ 上至少存在一个点 ε，使下式成立：

$$\int_a^b f(x)g(x)dx = f(a)\int_a^\varepsilon g(x)dx + f(b)\int_\varepsilon^b g(x)dx$$

二重积分中值定理：

二重积分的中值定理

设f(x,y)在有界闭区域D上连续，σ_0 是D的面积，则在D内至少存在一点 (ε,μ)，使得

$$\iint_D f(x,y)\,d\sigma = f(\varepsilon,\mu)\bullet\sigma_0$$

29. 格林公式（复连通区域，有洞）

将第二类闭曲线积分与面积积分通过被积函数矢量的旋度联系

对于一个水流流速图，区域内所有出水口入水口的流入或流出的水的速度和，就是你在区域边界所得到的流入或流出的水的速度和。

条件一封闭且正向，条件 2 是保证函数及其偏导数在域内连续有定义

设闭区域 D 由分段光滑的曲线 L 围成，函数 $P(x,y)$ 及 $Q(x,y)$ 在 D 上具有一阶连续偏导数，则有

$$\iint_D \left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dxdy=\oint_L Pdx+Qdy \quad (1)$$

其中 L 是 D 的取正向的边界曲线。

公式(1)叫做格林(green）公式。[1]

30. 高斯公式

高斯公式将第二类闭曲面积分与体积积分通过被积矢量函数的散度联系。而第二类曲面积分物理上能理解为被积矢量在曲面的通量；高斯公式可以推论出在闭曲面包围空间内若矢量的散度为 0 则矢量在闭曲面上的通量为 0 ； 其实某空间位置矢量的散度一直与该位置是否有矢量源（汇）联系起来，可以想象如果闭曲面包围区域内无源汇（就是说矢量是从区域外穿向区域内，并最终穿出去），则对于一个闭曲面来说，矢量总是一进一出，通量一正一负，整个曲面的通量必然就为 0 。

设空间有界闭合区域 Ω，其边界 $\partial\Omega$ 为分片光滑闭曲面。函数 $P(x,y,z),Q(x,y,z),R(x,y,z)$ 及其一阶偏导数在 Ω 上连续，那么：[1]

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

31. 三重积分

质量，被积函数表示密度。

32. 高等数学中最重要的两种集合

实数集、复数集。

二、 线性代数

1. 矩阵

矩阵（Matrix）是一个按照长方阵列排列的复数或实数集合

2. 为什么引入代数余子式

低阶好计算，为了计算高阶。如果第 i 行除了（i，j）之外的元素都为零，那么行列式的值就等于该元素与代数余子式的乘机

行列式=任一行的各元于其代数余子式乘积之和

3. 行列式按行展开法则

行列式等于他任一行的各元素与其对应的代数余子式乘机之和。

4. 范德蒙德行列式

一个 e 阶的范德蒙行列式由 e 个数 c_1，c_2，…，c_e决定，它的第 1 行全部都是 1，也可以认为是 c_1，c_2，…，c_e各个数的 0 次幂，它的第 2 行就是 c_1，c_2，…，c_e（的一次幂），它的第 3 行是 c_1，c_2，…，c_e的二次幂，它的第 4 行是 c_1，c_2，…，c_e的三次幂，…，直到第 e 行是 c_1，c_2，…，c_e的 e-1 次幂。

$$D_n=\prod_{1\le j<i\le n}\left(x_i-x_j\right),$$

数学归纳法证明。

5. 伴随矩阵

行列式各个元素的代数余子式所构成的矩阵

A* A=AA*=|A|E

类似于逆矩阵的概念

逆矩阵和伴随矩阵只差一个系数

性质 与原矩阵形成映射

6. 逆矩阵

与原矩阵相乘等于单位矩阵，可用于求位置矩阵的值

可逆-A 可经一系列初等行变换化为单位矩阵

求法：编程/初等变换

初等变换后两矩阵等价，n 阶单位矩阵 E 经过一次初等变换所得的矩阵为 n 阶初等矩阵

7. 奇异、满秩、可逆

当|A|=0 时，A 为奇异矩阵，非满秩，AX=0 只有零解

否则为非奇异矩阵，可逆矩阵，满秩，AX=0 有非零解

8. 克拉默法则

用于解决方程个数与未知数个数相当并且系数行列式不等于 0 的线性方程组，复杂度很高，时间复杂度依赖于矩阵行列式的算法复杂度 O(f(n))，其复杂度为 O(n·f(n))

9. 线性相关与线性无关

若存在有不全为 0 的 k1,k2,...,ks 使得 k1α1+k2α2+...+ksαs=0，则向量组α1,α2,...,αs (s≥≥1)称为是线性相关的，

线性有关就是在一组数据中有一个或者多个量可以被其余量表示。线性无关，就是在一组数据中没有一个量可以被其余量表示。互相垂直的三个坐标轴就是一组最简单的现行无关的向量

10. 极大无关组

αi1、αi2……αir 是部分组，部分组线性无关，全组可由部分组线性表出，部分组为全组的一个极大无关组。

α1、α2...αs（I）,秩（I）<s,向量组（I）线性相关；秩（I）=s，（I）线性无关

11. 什么叫矩阵的秩

不在同一条直线上的向量就是不相关的向量。

矩阵中的最大的不相关的向量的个数，就叫秩，可以理解为有秩序的程度。最高非零子式的阶数，可逆矩阵的秩等于矩阵的阶数，不可逆矩阵的秩小于矩阵的阶数。

从社会学的角度在考虑一下，政府机关是讲人际关系的地方，可谓是关系错综复杂，通常都是近亲繁殖。显然，这些部门，用矩阵来说，就不满秩，秩非常小。可以想象这些地方的工作肯定是搞不好的，因为没有秩序。所以想找个好单位，满秩可以作为一项评价指标哦。

矩阵的秩越低表示冗余度越大，用很少的几个基就可以表达所有数据，用 rank（A）个特征向量基本还原图片 A 的信息。容易建模容易学习。

12. 线性变换与非线性变换

从几何意义上，线性变换 y=kx+b 表示的是直线的特性，符合两个性质：变换前后零点不变，变换前后直线还是直线。

线性变换意味着可以将空间中的向量围绕零点进行旋转伸缩，但不能将其弯曲，否则则是非线性变化。

非线性变换将空间进行了扭曲，比如把 SVM 中的核函数看做描述低维空间到高维空间的映射，把原始低维空间中线性不可分的数据变成高维空间中线性可分的数据，优雅解决了问题。

从数值意义上，变换即函数，线性变换就是一阶导数为常数的函数，譬如 y=kx，把 y=kx 拓展为 n 维空间的映射，x、y 看做 n 维向量，当 k 为常数时，易得满足同质性 f(ka)=kf(a)，当 k 为一个矩阵时，易得满足可加性 f(a+b)=f(a)+f(b)。

同质性和可加性又称为线性条件，满足该条件则为线性变换，反之则为非线性变换。
线性映射（ linear mapping）是从一个向量空间 V 到另一个向量空间 W 的映射且保持加法运算和数量乘法运算，而线性变换（linear transformation）是线性空间 V 到其自身的线性映射。

13. 奇异值分解（SVD）与特征值分解　特征值和特征向量

不同特征值的特征向量线性无关
只有方阵才进行特征值分解，特征值的本质：
特征值分解：把方阵分解为缩放矩阵+特征向量矩阵，没有旋转或旋转角度为 0
特征值分解总结：特征值分解可以得到：
特征值：特征值表示的是这个特征到底有多重要
特征向量：而特征向量表示这个特征是什么，可以将每一个特征向量理解为一个线性的子空间，我们可以利用这些线性的子空间干很多的事情。

奇异值往往对应着矩阵中隐含的重要信息，且重要性和奇异值大小正相关。每个矩阵[公式]都可以表示为一系列秩为 1 的“小矩阵”之和，而奇异值则衡量了这些“小矩阵”对于[公式]的权重。
只有非方阵才进行奇异值分解
SVD 分解：把矩阵分解为缩放矩阵+旋转矩阵+特征向量矩阵
A 的非 0 奇异值的个数等于它的秩
奇异值σ与特征值类似，矩阵Σ中也是从大到小排列，而且σ的减少特别的快，在很多情况下，前 10%甚至 1%的奇异值的和就占了全部的奇异值之和的 99%以上了
右边的三个矩阵相乘的结果将会是一个接近于 A 的矩阵。而这三个矩阵的“面积”之和要远小于原始的矩阵 A，可以压缩空间来表示原矩阵 A，只存储三个矩阵 U、Σ、V 即可。
分别左乘与右乘得到左右奇异矩阵。

14. 向量空间/线性空间

集合 V 非空，对于向量的加法及数乘两种运算封闭，V 即为向量空间。八大运算，交换律结合律

15. 基

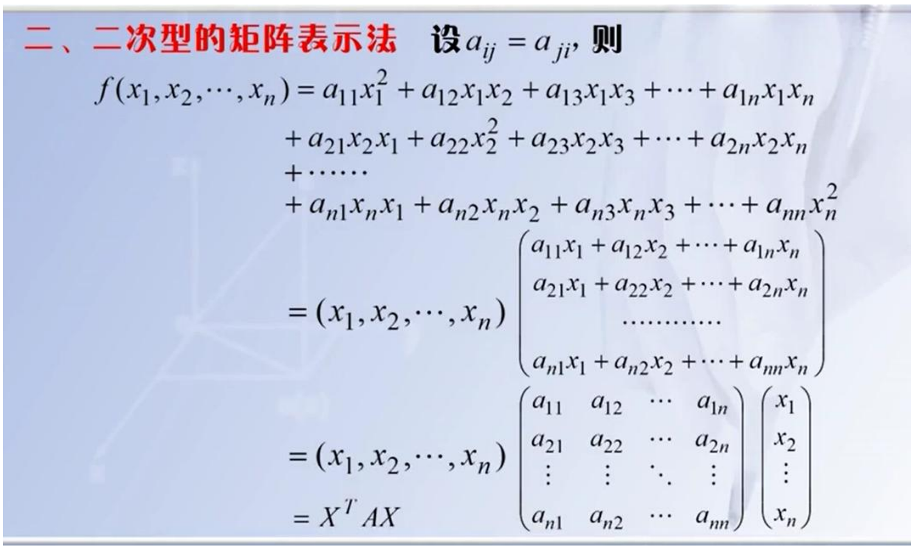
向量空间中间的几个向量，线性无关，其他向量都可以用这几个表示，则为 V 的一个基

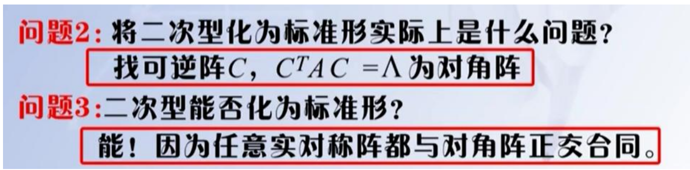
向量空间的基：我们想找一组向量是线性无关，且所有空间的向量可以由这组向量唯一表示，即向量数目不多不少。

定义：空间的每一个向量都可以由基向量唯一线性表示。

16. 二次型

二次型的矩阵一定是对称阵，才能保证多项式用矩阵表示的唯一性





17. 正交

两个向量的内积等于 0，若 n 维向量是向量空间的一个基，其中的向量两两正交且都为单位向量，那么为标准正交基。

18. 施密特正交化

我们更喜欢正交基（比如机器学习里面，第一步往往都是正交化，这样可以简化计算）

施密特正交化就是把非正交基变为正交基的。

线性无关向量组未必是正交向量组，从线性无关组导出正交向量组的过程为施密特正交化。可以找到这个线性空间的标准正交基。此时变换前与变换后的两组向量等价（可通过初等变换得到）

第二个向量减去它在第一个方向上的分量、依次。

第二个得到是原长度减去投影长度。

19. 正交变换

正交矩阵：矩阵的转置等于矩阵的逆。

对正交变换的线性变换即为正交变换，正交变换线段长度保持不变

20. 正定是什么意思

对于任意的 x 不等于，$f=x^TAx$ 都大于 0，则 f 为正定二次型，A 为正定矩阵，充分必要条件是他的标准型 n 个系数全为正，即规范型全为 1，正惯性系数等于 n

把二次型 f 所化得的标准二次型的平方项的系数中,正的个数和负的个数分别称为 f 的正惯性指数和负惯性指数．

正负惯性指数之和=f 的秩．

设 A 为实对称矩阵，$f=x^TAx$，若对于每个非零实向量 X，都有 X'AX≥0，则称 A 为半正定矩阵，称 X'AX 为半正定二次型。(其中，X'表示 X 的转置。)

21. 合同、惯性定理

    存在一个满秩 n 阶方阵 P,使得 P’AP = B.，合同关系也是等价关系

    惯性定理：对于 n 阶实对称矩阵，都一定与实对角矩阵合同，但不只一个，但在这些实对角矩阵的对角元中，正数的个数是一定的，叫做 A 的正惯性指数，负的个数也是一定的叫 A 的负惯性指数。

22. 矩阵的迹

    在这里可以看到就是对角的元素表示缩放，那么表示物体具有相似性。

    在机器学习里，主要为了获取数据的特征值，那么就是说，在任何一个矩阵计算出来之后，都可以简单化，只要获取矩阵的迹，就可以表示这一块数据的最重要的特征了，这样就可以把很多无关紧要的数据删除掉，达到简化数据，提高处理速度。

23. 范数
24. 厄米特矩阵

    厄米特矩阵（Hermitian Matrix，又译作“埃尔米特矩阵”或“厄米矩阵”），指的是自共轭矩阵。矩阵中每一个第 i 行第 j 列的元素都与第 j 行第 i 列的元素的共轭相等。

25. Jardon 标准型

    用于判断两个矩阵是否相似。

判定两个矩阵是否相似不是件容易的事，即使他们有相同的特征多项式、迹和行列式，他们仍然可能不相似。

那么我们的想法是：如果能将给定的两个矩阵 A,B 通过相似性转换成同一类矩阵，那么他们必然是相似的。

而由于种种原因直接寻找对角矩阵或者上三角矩阵在实际操作中都或多或少有一些问题，

因此我们的想法是：寻找介于对角矩阵和上三角矩阵之间的一类矩阵，这就是 Jordan 矩阵，Jordan 矩阵有如下性质，使得我们判定矩阵相似变得容易起来：

两个 Jordan 矩阵相似，当且仅当他们有相同的对角分块（不计排列次序)，那么这两个 Jordan 矩阵相似。

因此，我们想要判定两个矩阵是否相似就等价于做这件事：

如果两个矩阵 A,B 可以相似于两个 Jordan 矩阵，且这两个 Jordan 矩阵相似，那么矩阵 A,B 也相似。

如果把矩阵化成对角矩阵，关于矩阵的函数计算问题就会大大简化。但一般的矩阵未必与对角矩阵相似。

矩阵的标准型有多重，Jordan （约当）标准型是最接近对角矩阵的形式

26. 雅各比行列式

    它是以 n 个 n 元函数的偏导数为元素的行列式 。事实上，在函数都连续可微（即偏导数都连续）的前提之下，它就是函数组的微分形式下的系数矩阵（即雅可比矩阵）的行列式。

27. 相似矩阵

三、 概率论

1. 古典概型和几何概型

古典：扔色子，样本个数有限，发生的可能性相同

几何：实验结果无限多个，可能的结果几何可以用一个有度量的几何区域表示，发生的可能性相同。

2. 全概率和贝叶斯公式

全概率公式是计算由若干复杂“原因”引起的复杂事件概率的一个有效公式。贝叶斯公式是用来计算在复杂事件已经发生的情况下，求某一种”原因“引起的条件概率。

全概率公式

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

例如，三门炮同时射击飞机，飞机坠毁的概率

贝叶斯公式

已知分别在每一个条件B_j下，事件发生的概率是$P(A|B_j)$，现在是事件结果A已经发生了，求它是由于B_j 引起的概率

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

在全概率公式中，如果将 A 看成是“结果”，Bj 看成是导致结果发生的诸多“原因”之一，那么全概率公式就是一个“原因推结果” 的过程。

但贝叶斯公式却恰恰相反。贝叶斯公式中，我们是知道结果 A 已经发生了，所要做的是反过来研究造成结果发生的原因，是 XX 原因造成的可能性有多大，即“结果推原因”。

面对复杂而笼统的问题，人们往往走捷径，依据可能性而非根据概率来决策。而是给予最近发生的事件和最新的经验以更多的权值

贝叶斯公式就是当已知结果，问导致这个结果的第 i 原因的可能性是多少？执果索因！

3. 独立性

直观意义上来说，如果 A 与 B 独立，A-B 就等于从 A 中去除 AB 相交的部分

注意：独立不等于互斥，是两个概念，互斥是 A 与 B 没有交集。

独立：P(AB)=P(A)P(B)　互斥：P(AB)=0;

伯努利概型是重复独立试验的一个重要概率模型，其特点是：

1）事件只有发生与不发生两种结果；

2）各次试验中结果 A 发生的概率都相同，各次试验是相互独立的；

3）n 次重复实验。

4. 指数分布 均匀分布 泊松分布 二项分布

离散分布：

超几何分布：它描述了从有限 N 个物件（其中包含 M 个指定种类的物件）中抽出 n 个物件，成功抽出该指定种类的物件的次数（不放回）

二项分布和几何分布：抛硬币，重复试验，每次试验有两种结果，每种结果概率恒定，

当概率很小， $p(x) = C_n^x p^x (1-p)^{n-x}$　　$p(x) = (1-p)^{x-1} p$

重复次数 n 很大的的情况下，近似泊松分布。

泊松分布：离散型分布，一天内邮件丢失的数量，描述大量随机事件中稀有事件出现的次数。

2、此事件发生 K 次的概率

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

连续分布：

均匀分布：离散或者连续都有，每个可能的概率是相同的

指数分布：寿命，表示独立随机事件发生的时间间隔，无记忆性

正态分布：连续，描述群体的某一个指标，有一个均值一个标准差

5. 概率分布的定义

概率分布，是指用于表述随机变量取值的概率规律。事件的概率表示了一次试验中某一个结果发生的可能性大小。若要全面了解试验，则必须知道试验的全部可能结果及各种可能结果发生的概率，即随机试验的概率分布

6. 联合概率分布定义

合概率分布简称联合分布，是两个及以上随机变量组成的随机向量的概率分布。根据随机变量的不同，联合概率分布的表示形式也不同。对于离散型随机变量，联合概率分布可以以列表的形式表示，也可以以函数的形式表示；对于连续型随机变量，联合概率分布通过一非负函数的积分表示。

如果将二维随机变量(X,Y)看成是平面上随机点的坐标，那么分布函数 F(x,y)在(x,y)处的函数值就是随机点(X,Y)落在以点(x,y)为顶点而位于该点左下方的无穷矩形域内的概率。

7. 条件概率分布

相关事件的概率也叫“条件概率”。条件概率是指事件 A 在另外一个事件 B 已经发生条件下的发生概率。

有时，我们要考虑在其中一个随机变量取得(可能的)固定值的条件下，另一随机变量的概率分布。这样得到的 X 或 Y 的概率分布叫做条件概率分布，简称条件分布。

8. 中心极限定理

样本的平均值约等于总体的平均值。

不管总体是什么分布，任意一个总体的样本平均值都会围绕在总体的整体平均值周围，并且呈正态分布。n 个相互独立且同分布的随机变量的和近似为正态分布

用样本来估计总体。根据总体信息判断某个样本是否属于总体：3 标注差准则

无论样本如何分布，无数个样本之和就是正态分布。

中心极限定理的应用场景

1）在没有办法得到总体全部数据的情况下，我们可以用样本来估计总体

2）根据总体的平均值和标准差，判断某个样本是否属于总体

总体均值标准误反比于样本数量的平方根： $S_{\bar{x}} \to \frac{\sigma}{\sqrt{n}}$　（公式 1）

（2）和噪声密切相关，广泛应用在通信理论、信号处理、图像处理中

信号处理最简单的应用——signal averaging除噪：通过在时域上平均多次测量值提高信噪比（signal-noise ratio），好用到不行。

9. 大数定理

贝努利大数定律：事件发生的频率依收敛于事件的概率。明确了频率的稳定性，当 很大时，事件发生的频率与概率有较大偏差的可能性很小。

这个版本定义是从概率的角度，当 N 很大的时候，事件 A 发生的概率等于 A 发生的频率。

保险学应用：

数定律在保险学上的应用包括保费的厘定，以及保险金的赔偿等等。关于保险金的赔偿其实是符合大数定律的，因为现实中每个人的保费是不同的，但是因为投保的基数很大,所以根据大数定律，每个投保户的平均赔偿金额将会稳定在某一数值附近。

说明:

1、理论依据:保险的赔偿遵从大数定律,即如果投保人数充分大,则平均赔偿率几乎恒等于一个常数。利用大数定律与中心极限定理计算相关事件的概率。

2、应用与推广:大数定律的一个重要应用是在保险学方面。基本原理是一系列相互独立随机变量的平均值几乎恒等于一个常数,这个常数就是它的数学期望，或者说一系列相互独立随机变量的平均值依概率收敛于它的数学期望，可以广泛应用于保险精算、资源配置等方面。

银行经营应用：

然而要想利用大数定律来预测坏账出现的概率，银行贷款必须要满足两个条件: (1)每一笔贷款都必须是小额的; (2) 借款的群体要足够大。第一个条件是要保证每一笔贷款不会对总体贷款平均结果产生影响，因为学过概率论的知道,如果总体里面有一项很大，那么这一项将影响总体平均结果的走向;其次,这个条件还能降低因借款人的道德风险给银行带来的损失,因为如果出现--笔大额坏账，那么银行将会严重亏损，对于规模较小的银行可能会直接倒闭。另外一个条件则是大数定律最本质的要求,因为只有在样本量很大的情况下大数定律预测的结果方才准确。这两个条件缺一不可，非国有中小银行只有同时达到这两个条件方才能保证贷款业务的欣荣。

10. 切比雪夫不等式