第二章 课程准备知识

殷东生
yindongsheng@tsinghua.edu.cn

清华大学数学科学系

2023年秋季学期

谱

定义 (谱)

对任一$A \in \mathbb{C}^{n\times n}$，其有$n$个特征值（可以相同）。全体特征值的集合称为$A$的 谱(*spectrum*)，记作 $\sigma(A)$，即

$$\sigma(A) \overset{def}{=} \{\lambda_1, \cdots, \lambda_n\}$$

$$\rho(A) \overset{def}{=} \max_{\lambda\in\sigma(A)} |\lambda|$$

称为A的谱半径(*spectrum radius*)。

$$\begin{aligned}\mathrm{trace}(A) &= \sum_{1\le i\le n} \lambda_i,\\ \det A &= \lambda_1\lambda_2\cdots\lambda_n\end{aligned}$$

Matlab 中求特征值和特征向量的命令为 eig(A)，若要得到特征值和其对应的特征向量可用

»[V D]=eig(A)

D 是由特征值组成的对角阵，V 的列向量是特征值对应的特征向量。如求 B 的特征对。

$$B = \begin{pmatrix} 1 & 6 & 3 \\ -1 & 4 & 9 \\ 12 & 35 & 1 \end{pmatrix}$$

```
>> [V E] = eig(B)

V =
    0.0118    0.9119    0.2500
    0.4211   -0.3220    0.4278
   -0.9069    0.2545    0.8686

E =
  -15.4092         0         0
         0   -0.2812         0
         0         0   21.6905
>> eig(B)

ans =
  -15.4092
   -0.2812
   21.6905
```

定义 (内积(inner product))

设 V 是数域 $\mathbb{P}$ 上的线性空间，内积 $(\cdot,\cdot): V\times V\to\mathbb{P}$，对于$\forall u,v\in V,\exists!(u,v)\in\mathbb{P}$与之对应，且有：

1. $(u+v,w)=(u,w)+(v,w)$，$\forall u,v,w\in V$;
2. $(\alpha u,v)=\alpha(u,v)$，$\forall u,v\in V,\alpha\in\mathbb{P}$;
3. $(u,v)=\overline{(v,u)}$，$\forall u,v\in V$;
4. $(u,u)\geq 0$，$\forall u\in V$，且 $(u,u)=0\Leftrightarrow u=\theta$.

则称 (u,v) 是 u 和 v 的内积，定义了内积的空间称为内积空间。

$\mathbb{R}^n$和 $\mathbb{C}^n$的内积

$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$, 则内积 $(\boldsymbol{x},\boldsymbol{y})$定义为

$$(\boldsymbol{x},\boldsymbol{y})=\sum_{i=1}^{n}x_iy_i=\boldsymbol{y}^T\boldsymbol{x}.$$

若 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{C}^n$，则内积定义为

$$(\boldsymbol{x},\boldsymbol{y})=\sum_{i=1}^{n}x_i\bar{y}_i=\bar{\boldsymbol{y}}^T\boldsymbol{x}.$$

如果给定$\omega_i\in\mathbb{R},\omega_i>0,1\le i\le n$(权系数)，可定义一种带权内积：

$$(\boldsymbol{x},\boldsymbol{y})_\omega=\sum_{i=1}^{n}\omega_ix_iy_i$$

注记

超市的结算系统用了内积，内积可以用来新闻分类。

正交向量

定义 (正交(orthogonal))

若向量 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$ 满足$(\boldsymbol{x},\boldsymbol{y})=0$，则称它们是正交的。两个向量的集合 X 和 Y，若每个 $\boldsymbol{x}\in X$ 都和每个 $\boldsymbol{y}\in Y$ 正交，则称 X 与 Y 正交。

定义 (正交集合)

S 是非零向量的集合，若 $\forall \boldsymbol{x},\boldsymbol{y}\in S, \boldsymbol{x}\neq\boldsymbol{y}\Rightarrow(\boldsymbol{x},\boldsymbol{y})=0$，则称其为正交集合。

定理

正交集合 S 中的向量是线性无关的。

推论

若正交集合 $S \subseteq \mathbb{R}^n$ 有 n 个向量，则它是$\mathbb{R}^n$ 的一组基。

定义 (正交矩阵(orthogonal matrix))

设$Q \in \mathbb{R}^{n\times n}$，若满足

$$Q^TQ = I,$$

则称Q为正交矩阵。

正交矩阵有如下的性质：

1. Q不同的列向量相互正交，且各列向量的2-范数为1。
2. $Q^{-1} = Q^T$，且Q^T也是正交矩阵。
3. $|\det Q| = 1$。
4. 若A和B是同阶的正交矩阵，则AB和BA都是正交矩阵。

定义 (L^2 内积)

设$f(x), g(x) \in C[a,b]$，则它们的L^2 内积为

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

定义 (权函数(weight function))

若定在 $[a,b]$ 上的可积函数 $\rho(x)$ 满足

1. $\rho(x) \geq 0, \forall x \in [a,b]$;
2. 在 $[a,b]$ 的任一子区间上$\rho(x)$ 不恒为零。

称 $\rho(x)$ 为 $[a,b]$ 上的一个权函数。

利用权函数可定义带权L^2 内积：

$$(f,g)_\rho = \int_a^b \rho(x)f(x)g(x)dx$$

定理 (Cauchy-Schwarz不等式)

设V 是一个内积空间，则对任一的 u,v有

$$|(u,v)^2| \leq (u,u)(v,v),$$

等号当且仅当 u,v 线性相关时成立。

数学家简介(柯西(Cauchy))

柯西*(Cauchy, 1789—1857)*是法国数学家、物理学家、天文学家。柯西在数学上的最大贡献是在微积分中引进了极限概念，并以极限为基础建立了逻辑清晰的分析体系。复变函数的微积分理论就是由他创立的。*1821*年柯西提出极限定义的方法，把极限过程用不等式来刻画，后经魏尔斯特拉斯改进，成为现在所说的柯西极限定义或叫$\epsilon-\delta$定义。当今所有微积分的教科书都还沿用着柯西等人关于极限、连续、导数、收敛等概念的定义。柯西对定积分作了最系统的开创性工作，他把定积分定义为和的“极限”。在定积分运算之前，强调必须确立积分的存在性。他利用中值定理首先严格证明了微积分基本定理。使数学分析的基本概念得到严格的论述。把微积分及其推广从对几何概念、运动和直观了解的完全依赖中解放出来，并使微积分发展成现代数学最基础最庞大的数学学科。

Faá di Bruno $\mapsto$ Peano $\mapsto$ Russell

Gram-Schmidt正交化方法

若$\{u_1, u_2, \cdots, u_n\}$ 是内积空间 V 中的一个线性无关元素系列，则

$$\begin{cases} v_1 = u_1, \\ v_i = u_i - \displaystyle\sum_{k=1}^{i-1} \frac{(u_i, v_k)}{(v_k, v_k)} v_k, \ i = 2, 3 \cdots \end{cases}$$

生成V 中的一个正交序列 $\{v_1, v_2, \cdots, v_n\}$ 是 $\mathrm{span}\{u_1, u_2, \cdots, u_n\}$ 的一组基。

注记

Gram-Schmidt 正交化以丹麦数学家 *Gram* 和德国数学家 *Schmidt*的名字命名，但是最先由*Laplace* 和 *Cauchy* 给出。

数学家简介 (Pierre-Simon Laplace)

拉普拉斯(1749—1827)是法国分析学家、概率论学家和物理学家，法国科学院院士。是天体力学的主要奠基人、天体演化学的创立者之一， 还是分析概率论的创始人，是应用数学的先驱。他发表的天文学、数学和物理学的论文有270多篇，专著合计有4006多页。其中最有代表性的专著有《天体力学》、 《宇宙体系论》（中译本1978年版）和《概率分析理论》（1812）。

Lapalce 方程 $-\Delta u = 0;$

Laplace 变换 $F[s] = \mathcal{L}[f(t)] = \int_0^{+\infty} f(t)e^{-st}dt.$

若要求归一化即 $\|\boldsymbol{q}_i\|_2 = 1, 1 \le i \le n$，则由Gram-Schmidt 正交化可得QR 分解。设

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{u}_2 = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \end{pmatrix}$$

1. 归一化：$\boldsymbol{q}_1 = \frac{1}{2}(1, -1, 1, 1)^T$。
2. 计算投影:$(\boldsymbol{u}_2, \boldsymbol{q}_1) = 2$，接着

$$\hat{\boldsymbol{q}} = \boldsymbol{u}_2 - (\boldsymbol{u}_2, \boldsymbol{q}_1)\boldsymbol{q}_1 = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \end{pmatrix} - 2 \times \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}$$

3. 归一化：$\boldsymbol{q}_2 = \frac{1}{2}(1, 1, -1, 1)^T$。

经典的Gram-Schmidt正交化

算法 (经典 Gram-Schmidt 算法)

1. 对$j=1$ 到 n
2. $\boldsymbol{v}_j = \boldsymbol{u}_j$
3. 对 $i=1$ 到 $j-1$
4. $r_{ij} := (\boldsymbol{u}_j, \boldsymbol{q}_i)$
5. $\boldsymbol{v}_j = \boldsymbol{v}_j - r_{ij}\boldsymbol{q}_i$
6. $r_{jj} := \|\boldsymbol{v}_j\|_2$
7. 若 $r_{jj}=0$ 算法终止，或者 $\boldsymbol{q}_j = \boldsymbol{v}_j / r_{jj}$

若$\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ 线性无关，则算法 n 步完成。

定义矩阵R的元素为

$$r_{ij} = \begin{cases} (\boldsymbol{u}_j, \boldsymbol{q}_i), & i < j \\ \|\hat{\boldsymbol{q}}\|_2, & i = j \\ 0, & i > j \end{cases}$$

$$\Rightarrow \boldsymbol{u}_j = r_{1j}\boldsymbol{q}_1 + r_{2j}\boldsymbol{q}_2 + \cdots + r_{jj}\boldsymbol{q}_j = [\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_j] \begin{bmatrix} r_{1j} \\ r_{2j} \\ \vdots \\ r_{jj} \end{bmatrix} \quad 1 \le j \le n.$$

令

$$U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n], \quad Q = [\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n],$$

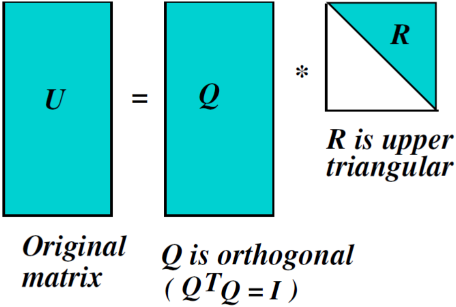
且令 $R \in \mathbb{R}^{n \times n}$ 是上面定义的矩阵的上三角部分

$$R = \{r_{ij}\}_{i,j=1}^n.$$

则有 U 的QR分解(QR factorization)

$$[\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n] = [\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}$$

更一般的，若 $U \in \mathbb{R}^{m\times n}$ 列满秩，则有



Matlab 的 Classical Gram-Schmidt 算法的程序和计算结果：

```
function [Q R]=cgs(A)
[m,n]=size(A);Q=zeros(m,n);R=zeros(n,m);%以使矩阵具有正确的大小。
for j=1:n %Gram-Schmidt正交化
v=A(:,j); %v初始化为A的第j列
for i=1:j-1
R(i,j)=Q(:,i)'*A(:,j); %为提高精确度用A(:,j)代替v
v=v-R(i,j)*Q(:,i); %减去投影
end %v现在和q1,…,qj 正交
R(j,j)=norm(v);
Q(:,j)=v/R(j,j); %将v规范化使之成为单位向量qj
end
```

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{bmatrix}, \quad \epsilon = 10^{-8}$$

```
>> epsilon = 1.0e-8;
A = [1 1 1;epsilon 0 0;0 epsilon 0;0 0 epsilon];
>> [Q R]=cgs(A)

Q =

    1.0000         0         0
    0.0000   -0.7071   -0.7071
         0    0.7071         0
         0         0    0.7071


R =

    1.0000    1.0000    1.0000         0
         0    0.0000         0         0
         0         0    0.0000         0

>> Q'*Q

ans =

    1.0000   -0.0000   -0.0000
   -0.0000    1.0000    0.5000
   -0.0000    0.5000    1.0000
```

经典的Gram-Schmidt正交化方法数值上是不稳定的。

考虑Gram-Schmidt正交化的第j步：

$$\boldsymbol{v}_j = \boldsymbol{u}_j - (\boldsymbol{u}_j, \boldsymbol{q}_1)\boldsymbol{q}_1 - \cdots - (\boldsymbol{u}_j, \boldsymbol{q}_{j-1})\boldsymbol{q}_{j-1}, \quad \boldsymbol{q}_j = \boldsymbol{v}_j / \|\boldsymbol{v}_j\|_2.$$

则

$$(\boldsymbol{u}_j, \boldsymbol{q}_i)\boldsymbol{q}_i = (\boldsymbol{q}_i^T \boldsymbol{u}_j)\boldsymbol{q}_i = \boldsymbol{q}_i(\boldsymbol{q}_i^T \boldsymbol{u}_j) = (\boldsymbol{q}_i \boldsymbol{q}_i^T)\boldsymbol{u}_j = \hat{P}_i \boldsymbol{u}_j.$$

其中$\hat{P}_i = \boldsymbol{q}_i \boldsymbol{q}_i^T$是到$\mathrm{span}\{\boldsymbol{q}_i\}$的秩为1的投影算子。则上式等价于

$$\boldsymbol{v}_j = (I - \hat{P}_1 - \cdots - \hat{P}_{j-1})\boldsymbol{u}_j = (I - P_{j-1})\boldsymbol{u}_j := Q_{j-1}\boldsymbol{u}_j$$

这里正交投影算子：

$$P_{j-1} = \sum_{i=1}^{j-1} \hat{P}_i = \boldsymbol{q}_1 \boldsymbol{q}_1^T + \cdots + \boldsymbol{q}_{j-1} \boldsymbol{q}_{j-1}^T$$

P_j是到

$$\mathrm{span}\{\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_j\}=\mathcal{K}_j$$

的正交投影算子。每次到一维空间的投影 $\hat{P}_j$ 易受误差的影响，导致数值上的不稳定性。而

$$Q_j=I-P_j\longmapsto\mathcal{K}_j^{\perp}.$$

MGS 的想法是把投影算子按另一种方式排列，令

$$\hat{Q}_i=I-\hat{P}_i=I-\boldsymbol{q}_i\boldsymbol{q}_i^T,\quad i=1,\cdots,j$$

其中$\mathrm{rank}(\hat{Q}_i)=n-1$是到$\mathrm{span}\{\boldsymbol{q}_i\}^{\perp}$正交投影算。
由$\boldsymbol{q}_i$的正交性可以验证

$$\begin{aligned}\hat{Q}_j\hat{Q}_{j-1}\cdots\hat{Q}_1 &= (I-\boldsymbol{q}_j\boldsymbol{q}_j^T)(I-\boldsymbol{q}_{j-1}\boldsymbol{q}_{j-1}^T)\cdots(I-\boldsymbol{q}_1\boldsymbol{q}_1^T)\\ &= I-\boldsymbol{q}_j\boldsymbol{q}_j^T-\cdots-\boldsymbol{q}_1\boldsymbol{q}_1^T\\ &= Q_j\end{aligned}$$

改进的G-S 算法的投影解释

则Arnoldi过程中用$\hat{Q}_j\hat{Q}_{j-1}\cdots\hat{Q}_1$代替$Q_j$有

$$
\begin{aligned}
\boldsymbol{q}_1 &= \boldsymbol{u}_1/\|\boldsymbol{u}_1\|_2 \\
\boldsymbol{v}_2 &= (I-\boldsymbol{q}_1\boldsymbol{q}_1^T)\boldsymbol{u}_1,\quad \boldsymbol{q}_2=\boldsymbol{v}_2/\|\boldsymbol{v}_2\|_2 \\
\boldsymbol{v}_3 &= (I-\boldsymbol{q}_2\boldsymbol{q}_2^T)(I-\boldsymbol{q}_1\boldsymbol{q}_1^T)\boldsymbol{u}_2,\quad \boldsymbol{q}_3=\boldsymbol{v}_3/\|\boldsymbol{v}_3\|_2 \\
&\vdots \\
\boldsymbol{v}_n &= (I-\boldsymbol{q}_{n-1}\boldsymbol{q}_{n-1}^T)\cdots(I-\boldsymbol{q}_1\boldsymbol{q}_1^T)\boldsymbol{u}_n,\quad \boldsymbol{q}_n=\boldsymbol{v}_n/\|\boldsymbol{v}_n\|_2.
\end{aligned}
$$

又

$$(I-\boldsymbol{q}_i\boldsymbol{q}_i^T)\boldsymbol{v}=\boldsymbol{v}-(\boldsymbol{v},\boldsymbol{q}_i)\boldsymbol{q}_i$$

就可以得到MGS。

改进的(modified Gram-Schmidt MGS)算法

算法 (MGS)

1. 对 $i = 1$ 到 n
2. $\boldsymbol{v}_i = \boldsymbol{u}_i$
3. 对 $i = 1$ 到 n
4. $r_{ii} = \|\boldsymbol{v}_i\|_2$
5. $\boldsymbol{q}_i = \boldsymbol{v}_i / r_{ii}$
6. 对 $j = i + 1$ 到 n
7. $r_{ij} = (\boldsymbol{v}_j, \boldsymbol{q}_i)$
8. $\boldsymbol{v}_j = \boldsymbol{v}_j - r_{ij}\boldsymbol{q}_{ij}$
9. 若 $r_{jj} = 0$ 算法终止。

Matlab 的修正的Gram-Schmidt 算法的程序和计算结果：

```
% Modified Gram-Schmidt method
function [Q,R] = mgs(A)
[m,n]=size(A);
Q=zeros(m,n); %矩阵的行和列
Q(1:m,1) = A(1:m,1);
R=zeros(n); R(1,1)=1;
for k = 1:n
R(k,k) = norm (A(1:m,k));
Q(1:m,k) = A(1:m,k)/R(k,k);%对本次得到的正交向量进行归一化
for j=k+1:n
R(k,j) = Q(1:m,k)'* A(1:m,j);
A(1:m,j) = A(1:m,j) - Q(1:m,k)*R(k,j);%对剩余向量进行修正
end
end
```

$$A=\begin{bmatrix}1&1&1\\ \epsilon&0&0\\ 0&\epsilon&0\\ 0&0&\epsilon\end{bmatrix},\quad \epsilon=10^{-8}$$

```
>> epsilon = 1.0e-8;
A = [1 1 1;epsilon 0 0;0 epsilon 0;0 0 epsilon];
>> [Q R]=mgs(A)

Q =

    1.0000         0         0
    0.0000   -0.7071   -0.4082
         0    0.7071   -0.4082
         0         0    0.8165


R =

    1.0000    1.0000    1.0000
         0    0.0000    0.0000
         0         0    0.0000

>> Q'*Q

ans =

    1.0000   -0.0000   -0.0000
   -0.0000    1.0000    0.0000
   -0.0000    0.0000    1.0000
```

定义 (范数与线性赋范空间)

设V是数域$\mathbb{P}$上的线性空间，定义$\|\cdot\|: u \to \|u\|$是V到$\mathbb{R}$的一个线性映射，满足

1. 正定性*(positivity)*：$\|u\| \geq 0, \forall u \in V$；且$\|u\| = 0 \Leftrightarrow u = \theta$。
2. 齐次性*(scaling)*：$\|\alpha u\| = |\alpha|\|u\|, \forall u \in V, \alpha \in \mathbb{P}$。
3. 三角不等式*(triangle inequality)*：$\|u + v\| \leq \|u\| + \|v\|, \forall u, v \in V$。

称$\|\cdot\|$为V的范数*(模)(norm)*，定义了范数的线性空间称为赋范线性空间。

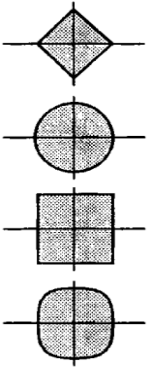
$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$可定义范数结构

$$\|\boldsymbol{x}\|_\infty = \max_{1\le i\le n} |x_i|, \quad \infty - 范数,$$

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|, \quad 1 - 范数,$$

$$\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}, \quad 2 - 范数,$$

$$\|\boldsymbol{x}\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}, \quad p - 范数.$$



内积与2-范数

通过内积可定义 2-范数

$$\|x\|_2 = \sqrt{(x,x)}.$$

设 $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$，其夹角为 α，则

$$\cos\alpha = \frac{(\boldsymbol{x},\boldsymbol{y})}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2}.$$

由上面的公式亦可得出Cauchy-Schwarz 不等式，且

$$\begin{aligned}
(\boldsymbol{u},\boldsymbol{v}) &= \|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2, \quad \alpha = 0,\\
(\boldsymbol{u},\boldsymbol{v}) &= -\|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2, \quad \alpha = \pi.
\end{aligned}$$

说明二者平行时，等号成立。

距离

定义 (距离)

设X是任一非空集合，X中的任意两点x,y有$d(x,y)\in\mathbb{R}$与之对应且满足：

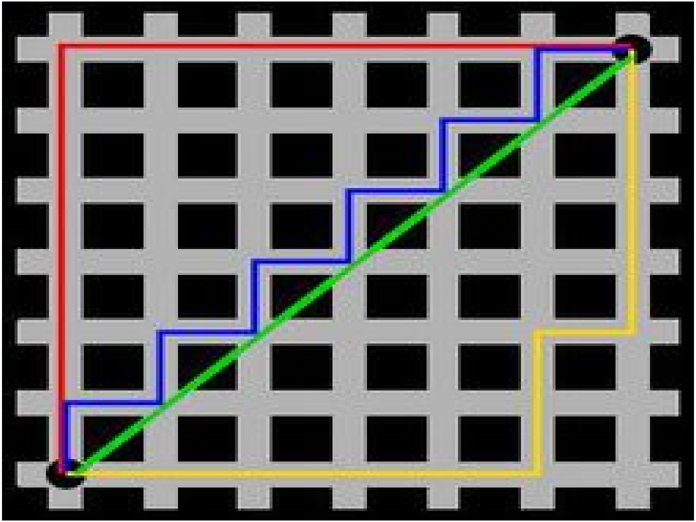
1. 非负性：$d(x,y)\geq 0$，且$d(x,y)=0$ 当且仅当 $x=y$;
2. 对称性：$d(x,y)=d(y,x)$;
3. 三角不等式：$d(x,y)\leq d(x,z)+d(z,y)$。

称$d(x,y)$为X上的一个距离，定义了距离d的集合 X 称为一个距离空间(X,d)，简记为 X。

范数可以诱导出距离，设V是赋范空间，则$u, v \in V$的距离可定义为

$$d(u, v) = \|u - v\|.$$

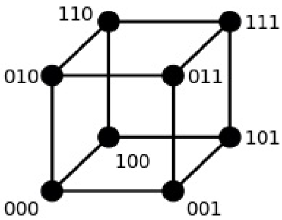
曼哈顿距离–1范数

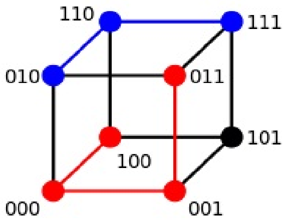


定义 (Hamming 距离)

*Hamming 距离*表示两个等长字符串在对应位置上不同字符的数目，海明距离度量了通过替换字符的方式将字符串x变成y所需要的最小的替换次数。

Hamming 距离主要应用在通信编码领域上，用于制定可纠错的编码体系。在机器学习中，海明距离也常常被用于作为一种距离的度量方式。





常用距离

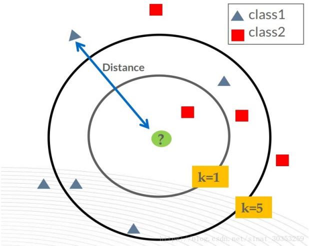
- 欧氏距离：$\|\boldsymbol{x}-\boldsymbol{y}\|_2=\sqrt{\sum_{i=1}^{n}|x_i-y_i|^2}$；
- 曼哈顿距离：$\|\boldsymbol{x}-\boldsymbol{y}\|_1=\sum_{i=1}^{n}|x_i-y_i|$；
- 切比雪夫距离：$\|\boldsymbol{x}-\boldsymbol{y}\|_\infty=\max_{1\le i\le n}|x_i-y_i|$；
- 余弦距离：$\cos\theta=\frac{(\boldsymbol{x},\boldsymbol{y})}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2}$；
- Jaccard相似系数：$J(A,B)=\frac{|A\cap B|}{|A\cup B|}$；
- 相关系数：$\rho_{XY}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=\frac{E((X-E(X)(Y-E(Y)))}{\sigma_X\sigma_Y}$

K近邻（K-Nearest Neighbour，KNN）算法

K近邻法是一种基本分类与回归方法，是最简单的机器学习算法之一。其基本做法是：给定测试实例，基于某种距离度量找出训练集中与其最靠近的K个实例点，然后基于这K个最近邻的信息来进行预测。

距离度量、K值的选择及分类决策规则是K近邻法的三个基本要素。其计算过程如下：

1. 计算测试数据与各个训练数据之间的距离；
2. 按照距离的递增关系进行排序；
3. 选取距离最小的K个点；
4. 确定前K个点所在类别的出现频率；
5. 返回前K个点中出现频率最高的类别作为测试数据的预测分类。



k-means 聚类(k-means clustering algorithm)

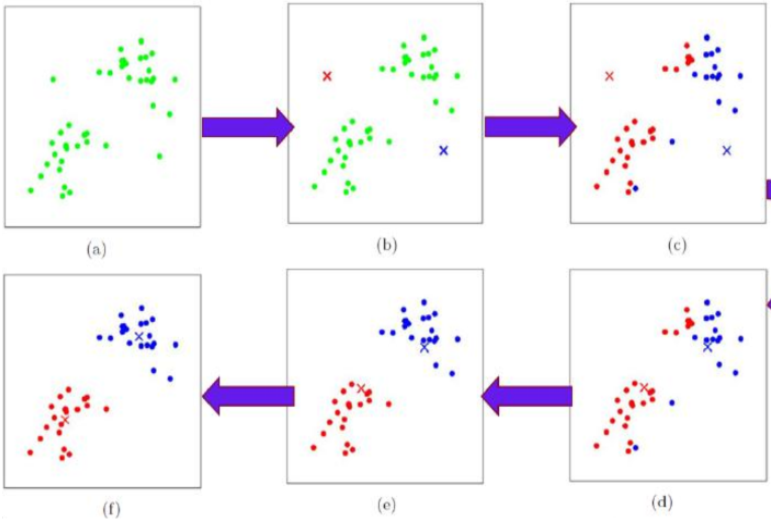
k-means算法是一种聚类算法，所谓聚类，即根据相似性原则，将具有较高相似度的数据对象划分至同一类簇，将具有较高相异度的数据对象划分至不同类簇。聚类与分类最大的区别在于，聚类过程为无监督过程，即待处理数据对象没有任何先验知识，而分类过程为有监督过程，即存在有先验知识的训练数据集。

k-means算法中的k代表类簇个数，means代表类簇内数据对象的均值（这种均值是一种对类簇中心的描述），因此，k-means算法又称为k-均值算法。k-means算法是一种基于划分的聚类算法，以距离作为数据对象间相似性度量的标准，即数据对象间的距离越小，则它们的相似性越高，则它们越有可能在同一个类簇。数据对象间距离的计算有很多种，k-means算法通常采用欧氏距离来计算数据对象间的距离。

k-means 算法

算法流程：

1. 首先确定一个k值，即希望将数据集经过聚类得到k个集合。
2. 从数据集中随机选择k个数据点作为质心。
3. 对数据集中每一个点，计算其与每一个质心的距离，离哪个质心近，就划分到那个质心所属的集合。
4. 把所有数据归好集合后，一共有k个集合。然后重新计算每个集合的质心。
5. 若新计算出来的质心和原来的质心之间的距离小于某一个设置的阈值（表示重新计算的质心的位置变化不大，趋于稳定，或者说收敛），可认为聚类已经达到期望的结果，算法终止。
6. 如果新质心和原质心距离变化很大，返回3∼5步骤。


(a)
(b)
(c)
(f)
(e)
(d)

定义 (范数等价)

设$\|\cdot\|_\alpha$和$\|\cdot\|_\beta$是线性空间V的两个范数。若存在正的常数C_1和C_2，使得

$$C_1\|u\|_\alpha \le \|u\|_\beta \le C_2\|u\|_\alpha,\ \forall u \in V$$

则称范数$\|\cdot\|_\alpha$和范数$\|\cdot\|_\beta$等价。

定理

有限维空间中的任意两个范数等价。

$$\boldsymbol{x} \in \mathbb{R}^n \Longrightarrow \begin{array}{rcl} \|\boldsymbol{x}\|_\infty & \le & \|\boldsymbol{x}\|_2 \le \sqrt{n}\|\boldsymbol{x}\|_\infty, \\ \|\boldsymbol{x}\|_\infty & \le & \|\boldsymbol{x}\|_1 \le n\|\boldsymbol{x}\|_\infty, \\ \|\boldsymbol{x}\|_2 & \le & \|\boldsymbol{x}\|_1 \le \sqrt{n}\|\boldsymbol{x}\|_2. \end{array}$$

对称阵

定义

如果 $A \in \mathbb{R}^{n\times n}$，$A^T = A$，则称 A 为 对称阵。

对称矩阵有下列性质：

1. A的特征值均为实数，且有n个线性无关的特征向量。
2. A对应于不同特征值的特征向量必正交。
3. 存在正交阵Q使得$Q^{-1}AQ$为对角阵。

正定阵

定义

若对称矩阵A满足

$$(A\boldsymbol{x}, \boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x} \geq 0, \ \forall \boldsymbol{x} \in \mathbb{R}^n,$$

则称A为半正定阵(*semidefinite matrix*)。若上式等号只当$\boldsymbol{x} = \boldsymbol{0}$时成立，则称$A$为正定阵(*definite matrix*)。

定理

实对称矩阵为正定阵的充要条件是它的所有特征值都是正数，或者所有顺序主子式都是正定的。

矩阵范数

定义 (矩阵范数)

$\mathbb{P}^{n\times n}$中的范数$\|\cdot\|$：是$\mathbb{P}^{n\times n}\to\mathbb{R}: A\to\|A\|$的一个映射满足

- 正定性 $\|A\|\geq 0$, $\|A\|=0\Leftrightarrow A=\mathbf{O}$;
- 齐次性 $\|\alpha A\|=|\alpha|\|A\|$;
- 三角不等式 $\|A+B\|\leq\|A\|+\|B\|$;
- 次可乘性*(submultiplicative)* $\|AB\|\leq\|A\|\|B\|$。

Frobenius范数

$$A \in \mathbb{R}^{n\times n}, \quad \|A\|_F \stackrel{def}{=} \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} = \sqrt{\text{trace}(A^TA)}$$

$$\begin{aligned}
\|AB\|_F^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}|\sum_{k=1}^{n}a_{ik}b_{kj}|^2 \\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{n}|a_{ik}|^2\sum_{k=1}^{n}|b_{kj}|^2\right) \\
&= \left(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|^2\right)\left(\sum_{j=1}^{n}\sum_{k=1}^{n}|b_{kj}|^2\right) \\
&= \|A\|_F^2\|B\|_F^2
\end{aligned}$$

定义 (相容性)

称矩阵范数与向量范数相容：若

$$\|A\boldsymbol{x}\| \le \|A\| \cdot \|\boldsymbol{x}\|, \ \forall A \in \mathbb{R}^{n\times n}, \ \forall \boldsymbol{x} \in \mathbb{R}^{n}.$$

矩阵的F-范数与向量的2-范数相容：

$$\begin{aligned} \|A\boldsymbol{x}\|_2^2 &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}x_j\right)^2 \\ &\le \sum_{i=1}^{n}\left(\sum_{j=1}^{n} |a_{ij}|^2 \sum_{j=1}^{n} x_j^2\right) \\ &= \|A\|_F^2 \cdot \|\boldsymbol{x}\|_2^2 \end{aligned}$$

$F(\boldsymbol{x}) = \|A\boldsymbol{x}\|$ 是$\boldsymbol{x} \in \mathbb{R}^n$的连续函数，因为

$$
\begin{aligned}
|F(\boldsymbol{x}) - F(\boldsymbol{y})| &= \Big| \|A\boldsymbol{x}\| - \|A\boldsymbol{y}\| \Big| \\
&\le \|A(\boldsymbol{x} - \boldsymbol{y})\| \\
&\le C\|A(\boldsymbol{x} - \boldsymbol{y})\|_2 \\
&\le C\|A\|_F \|\boldsymbol{x} - \boldsymbol{y}\|_2 \\
&\le C\|A\|_F C' \|\boldsymbol{x} - \boldsymbol{y}\| \\
&= \tilde{C}\|\boldsymbol{x} - \boldsymbol{y}\|
\end{aligned}
$$

设$\|\cdot\|$ 是 $\mathbb{R}^n$ 中任意给定的一向量范数，集合（$\mathbb{R}^n$中的单位球）

$$D = \{\boldsymbol{x} | \boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n, \|\boldsymbol{x}\| = 1\}$$

是一个有界闭集，又 $\|A\boldsymbol{x}\|$ 是 $\boldsymbol{x}$ 的连续函数，则在 D 上可达极大值，即存在 $\boldsymbol{x}_0 \in D$ 使得

$$\|A\boldsymbol{x}_0\| = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|.$$

对于任意 $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{0}$，有

$$\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \|A\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\|, \quad \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \in D$$

则 $\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$ 在$\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{0}$ 上有最大值，可定义

$$\max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|.$$

定义 (从属范数)

设$\|\cdot\|$是一个向量范数，对$\forall A \in \mathbb{R}^{n\times n}$，

$$\|A\| \overset{def}{=} \max_{x\neq 0} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|$$

定义了$\mathbb{R}^{n\times n}$上的一种范数，称为从属于给定向量范数的矩阵范数，简称从属范数*(subordinate matrix norm)*，亦称算子范数 *(operator norm)*。

- 单位矩阵I的任意从属范数为1。
- 相容的范数未必是从属范数。如$\|\cdot\|_F$与$\|\cdot\|_2$相容，但$\|I\|_F=\sqrt{n}$。

定理

设$A\in\mathbb{P}^{n\times n}$，则

$$\|A\|_\infty=\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|$$

$$\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|$$

$$\|A\|_2=[\rho(A^TA)]^{1/2}$$

证明$\|A\|_2=[\rho(A^TA)]^{1/2}$

若 $A=A^T$，则 $\rho(A)=\|A\|_2$，这是否说明 $\rho(A)$ 是一个矩阵范数？

$$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$$

则 $\rho(A)=0$，但 $A\neq 0$。不满足正定性。
考虑 $B=A^T$，则

$$\begin{aligned}\rho(A+B) &= 1,\\ \rho(A)+\rho(B) &= 0.\end{aligned}$$

则

$$\rho(A)+\rho(B)<\rho(A+B),$$

三角不等式不成立。

Matlabe 求矩阵范数

Matlab 中求矩阵范数的命令为 norm，除了求 $1-, 2-, \infty-$ 范数，还可求 Frobenius 范数。

$$
\begin{aligned}
A &= \begin{bmatrix} -2 & 1 & -8 & 1 \\ 0 & -4 & -21 & 18 \\ -33 & 16 & -6 & 20 \\ 14 & -20 & -18 & 5 \\ 8 & -1 & 12 & 16 \end{bmatrix} \\
\|A\|_\infty &= 33+16+6+20=75 \\
\|A\|_1 &= 8+21+6+18+12=65
\end{aligned}
$$

```
>> norm(A,'inf')
ans =
    75
>> norm(A,1)
ans =
    65
>> norm(A,'fro')
ans =
  63.5767
```

定理 (谱半径和范数)

设$\|\cdot\|$为任一矩阵范数，则$\|A\| \geq \rho(A)$。

若取$\mathbb{R}^{n\times n}$的范数为$\|\cdot\|_\infty$，$\|\cdot\|_1$和$\|\cdot\|_2$之一，则有

$$\rho(A) \leq \|A\|.$$

定理

$\forall A \in \mathbb{R}^{n\times n}$及实数$\epsilon > 0$，存在一种从属的矩阵范数$\|\cdot\|$，使得

$$\|A\| \leq \rho(A) + \epsilon$$

定理

设$\|\cdot\|$为$\mathbb{R}^{n\times n}$一从属范数。若$\|B\|<1$，则$I+B$非奇异，且

$$\|(I+B)^{-1}\| \leq \frac{1}{1-\|B\|}$$

推论 (摄动引理)

设$A, C \in \mathbb{R}^{n\times n}$，若$A^{-1}$存在，$\|A^{-1}\|\|A-C\| < 1$，则

$$\|C^{-1}\| \leq \frac{\|A^{-1}\|}{1-\|A^{-1}\|\|A-C\|}$$

定义 (实初等矩阵)

设$\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n$，$\sigma\in\mathbb{R}$，称

$$E(\boldsymbol{u},\boldsymbol{v};\sigma)\overset{def}{=}\boldsymbol{I}-\sigma\boldsymbol{u}\boldsymbol{v}^T$$

为实初等矩阵。

设$\boldsymbol{u}=(u_1,u_2,\cdots,u_n)^T$，$\boldsymbol{v}=(v_1,v_2,\cdots,v_n)^T$，则

$$\begin{aligned}\boldsymbol{u}\boldsymbol{v}^T &= \begin{pmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{pmatrix}\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \\ &= \begin{pmatrix} u_1v_1 & u_1v_2 & \cdots & u_1v_n \\ u_2v_1 & u_2v_2 & \cdots & u_2v_n \\ \cdots & \cdots & \cdots & \cdots \\ u_nv_1 & u_nv_2 & \cdots & u_nv_n \end{pmatrix}\end{aligned}$$

引理

初等矩阵若可逆，其逆也是（同类型）初等矩阵。

证明： 因为

$$
\begin{aligned}
E(\boldsymbol{u},\boldsymbol{v};\sigma)E(\boldsymbol{u},\boldsymbol{v};\tau) &= (\boldsymbol{I}-\sigma\boldsymbol{u}\boldsymbol{v}^T)(\boldsymbol{I}-\tau\boldsymbol{u}\boldsymbol{v}^T) \\
&= \boldsymbol{I}-\left(\sigma+\tau-\sigma\tau(\boldsymbol{v}^T\boldsymbol{u})\right)\boldsymbol{u}\boldsymbol{v}^T,
\end{aligned}
$$

若令

$$\sigma+\tau-\sigma\tau(\boldsymbol{v}^T\boldsymbol{u})=0,$$

即

$$\tau=\frac{\sigma}{\sigma\boldsymbol{v}^T\boldsymbol{u}-1},$$

则

$$E(\boldsymbol{u},\boldsymbol{v};\sigma)^{-1}=E(\boldsymbol{u},\boldsymbol{v};\tau).$$

- 若$\boldsymbol{v}^T\boldsymbol{u}=0$，即$\boldsymbol{u}$与$\boldsymbol{v}$垂直，则有$\tau=-\sigma$，

$$E(\boldsymbol{u},\boldsymbol{v};\sigma)^{-1}=E(\boldsymbol{u},\boldsymbol{v};-\sigma).$$

- $$\det E(\boldsymbol{u},\boldsymbol{v};\sigma)=1-\sigma\boldsymbol{v}^T\boldsymbol{u}$$

  因为$\boldsymbol{u}\boldsymbol{v}^T$的所有特征值为0和$\boldsymbol{v}^T\boldsymbol{u}$。

若记$\boldsymbol{l}_j = (0, \cdots, 0, l_{j+1,j}, \cdots, l_{n,j})^T$，其前$j$个分量为零。令

$$L_j(\boldsymbol{l}_j) = E(\boldsymbol{l}_j, \mathbf{e}_j; -1) = \boldsymbol{I} + \boldsymbol{l}_j \mathbf{e}_j^T = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & l_{j+1,j} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & l_{n,j} & 0 & \cdots & 1 \end{pmatrix}$$

注记

$L_j(\boldsymbol{l}_j)$左乘一个矩阵表示将j行的$l_{k,j}$倍加到k行上去，k从$j+1$行到n行。

因为 $\mathbf{e}_j^T \boldsymbol{l}_j = 0$，则 $L_j(\boldsymbol{l}_j)^{-1} = L_j(-\boldsymbol{l}_j)$。

当$i \le j$时，

$$\begin{aligned} L_i(\boldsymbol{l}_i)L_j(\boldsymbol{l}_j) &= (I + \boldsymbol{l}_i\boldsymbol{e}_i^T)(I + \boldsymbol{l}_j\boldsymbol{e}_j^T) \\ &= I + \boldsymbol{l}_i\boldsymbol{e}_i^T + \boldsymbol{l}_j\boldsymbol{e}_j^T \end{aligned}$$

因而一个（单位）下三角阵可以写成

$$L = L_1(\boldsymbol{l}_1)L_2(\boldsymbol{l}_2)\cdots L_{n-1}(\boldsymbol{l}_{n-1}).$$

定义

设$A \in \mathbb{R}^{n\times n}$，若存在排列矩阵$P \in \mathbb{R}^{n\times n}$使得

$$P^TAP = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

如果A_{11}, A_{22}为方阵，则称A为可约矩阵*(reducible matrix)*；否则，称A不可约*(irreducible)*。

注记 (判别准则)

矩阵 $A \in \mathbb{R}^{n\times n}$ 是可约的，若它的指标1到n可被分成两个互不相交的非空集合 $S = \{i_1, i_2, \cdots, i_\alpha\}$和 $T = \{j_1, j_2, \cdots, j_\beta\}$，$\alpha + \beta = n$， 使得

$$a_{i_p, j_q} = 0, \;\; 1 \le p \le \alpha, \; 1 \le q \le \beta.$$

例 (可约矩阵)

$$
A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 2 & 3 \\ 8 & 0 & 0 & 0 & 9 \\ 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12 & 0 \end{bmatrix},
$$

$$
S = \{1,3,4,5\}, \quad T = \{2\},
$$

且

$$
a_{12} = a_{32} = a_{42} = a_{52} = 0
$$

因此 A 可约。

若A是可约阵，因为$P^T = P^{-1}$，方程组$A\boldsymbol{x} = \boldsymbol{b}$可写为

$$
\begin{aligned}
P^T A P \boldsymbol{y} &= P^T \boldsymbol{b}, \\
\boldsymbol{y} &= P^T \boldsymbol{x}.
\end{aligned}
$$

方程组有分块形式

$$
\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{b}}_1 \\ \bar{\boldsymbol{b}}_2 \end{bmatrix}
$$

注记

若矩阵可约，则原方程可以化为先解一个低阶方程组$A_{22}\boldsymbol{y}_2 = \bar{\boldsymbol{b}}_2$，再解$A_{11}\boldsymbol{y}_1 = \bar{\boldsymbol{b}}_1 - A_{12}\boldsymbol{y}_2$。