统计计算

2016 年 3 月 22 日

目录

目录

目录

目录

目录

相容矩阵方程的通解

定理

设 $\mathbf{A}\in\mathbb{C}^{m\times n}, \mathbf{B}\in\mathbb{C}^{p\times q}, \mathbf{D}\in\mathbb{C}^{m\times q}$. 矩阵方程

$$\mathbf{AXB}=\mathbf{D} \tag{1}$$

有解的充要条件是

$$\mathbf{AA}^{-}\mathbf{DB}^{-}\mathbf{B}=\mathbf{D}, \tag{2}$$

且通解为

$$\mathbf{X}=\mathbf{A}^{-}\mathbf{DB}^{-}+\mathbf{Y}-\mathbf{A}^{-}\mathbf{AYBB}^{-}, \tag{3}$$

其中 $\mathbf{Y}\in\mathbb{C}^{n\times p}$ 为任意矩阵.

证明　若 $\mathbf{X}$ 为矩阵方程 (1) 的任一解, 则

$$\mathbf{A}\mathbf{A}^{-}\mathbf{D}\mathbf{B}^{-}\mathbf{B}=\mathbf{A}\mathbf{A}^{-}\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{B}^{-}\mathbf{B}=\mathbf{A}\mathbf{X}\mathbf{B}=\mathbf{D},$$

即式 (2) 成立. 反之, 若式 (2) 成立, 显然 $\mathbf{X}=\mathbf{A}^{-}\mathbf{D}\mathbf{B}^{-}$ 为方程 (1) 的一个解.

容易看到: 任意具有形式 (3) 的矩阵 $\mathbf{X}$ 一定满足方程 (1). 另外, 方程 (1) 的任意解 $\mathbf{X}$ 可表成

$$\mathbf{X}=\mathbf{A}^{-}\mathbf{D}\mathbf{B}^{-}+\mathbf{X}-\mathbf{A}^{-}\mathbf{A}\mathbf{X}\mathbf{B}\mathbf{B}^{-},$$

所以 $\mathbf{X}$ 具有式 (3) 的形式. $\square$

相容线性方程组的通解

推论

线性方程组 $\mathbf{Ax}=\mathbf{b}$ 相容的充要条件是

$$\mathbf{AA}^{-}\mathbf{b}=\mathbf{b},$$

且通解为

$$\mathbf{x}=\mathbf{A}^{-}\mathbf{b}+(\mathbf{I}-\mathbf{A}^{-}\mathbf{A})\mathbf{y},$$

其中 $\mathbf{y}$ 为维数相容的任意向量.

广义逆与相容方程的解

设方程 $\mathbf{Ax} = \mathbf{b}$ 相容.

1. 其通解可以表为

$$\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{y}$$

2. 最小 Euclid 范数解为 $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b}$;
3. 如果 $\mathbf{A}$ 满列秩, 则方程组的解唯一, 且

$$\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}.$$

广义逆与相容方程的解

设方程 $\mathbf{Ax}=\mathbf{b}$ 相容.

1. 其通解可以表为
$$\mathbf{x}=\mathbf{A}^{\dagger}\mathbf{b}+(\mathbf{I}-\mathbf{A}^{\dagger}\mathbf{A})\mathbf{y}$$
2. 最小 Euclid 范数解为 $\mathbf{x}=\mathbf{A}^{\dagger}\mathbf{b}$;
3. 如果 $\mathbf{A}$ 满列秩, 则方程组的解唯一, 且
$$\mathbf{x}=\mathbf{A}^{\dagger}\mathbf{b}=(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}.$$

广义逆与相容方程的解

设方程 $\mathbf{Ax}=\mathbf{b}$ 相容.

1. 其通解可以表为

$$\mathbf{x}=\mathbf{A}^{\dagger}\mathbf{b}+(\mathbf{I}-\mathbf{A}^{\dagger}\mathbf{A})\mathbf{y}$$

2. 最小 Euclid 范数解为 $\mathbf{x}=\mathbf{A}^{\dagger}\mathbf{b}$;
3. 如果 $\mathbf{A}$ 满列秩, 则方程组的解唯一, 且

$$\mathbf{x}=\mathbf{A}^{\dagger}\mathbf{b}=(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}.$$

广义逆与不相容方程的解

若不存在 $\mathbf{x}$ 使得 $\mathbf{A}\mathbf{x} = \mathbf{b}$, 则称线性方程组不相容, 通常考虑其最小二乘解.

对满足

$$\mathbf{x}_0 = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2,$$

的 $\mathbf{x}_0$ 称为方程 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的最小二乘解.

设方程 $\mathbf{Ax}=\mathbf{b}$ 不相容.

称形如 $\mathbf{A}^{\mathrm{T}}\mathbf{Ax}=\mathbf{A}^{\mathrm{T}}\mathbf{b}$ 方程为 $\mathbf{Ax}=\mathbf{b}$ 的正规方程.

1. 正规方程必有解;

证明　由于

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}(\mathbf{A}^{\dagger}\mathbf{b})=\mathbf{A}^{\mathrm{T}}(\mathbf{AA}^{\dagger})^{\mathrm{T}}\mathbf{b}=(\mathbf{AA}^{\dagger}\mathbf{A})^{\mathrm{T}}\mathbf{b}=\mathbf{A}^{\mathrm{T}}\mathbf{b},$$

所以 $\mathbf{A}^{\dagger}\mathbf{b}$ 是正规方程的一个解.

2. $\mathbf{x}$ 为正规方程的解等价于 $\mathbf{x}$ 是原方程的最小二乘解;

证明　设 $\mathbf{x}$ 是正规方程的解, 即 $\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$, 于是对任意 $\mathbf{x}_1$,

$$
\begin{aligned}
\|\mathbf{b}-\mathbf{A}\mathbf{x}_1\|^2 &= \|(\mathbf{b}-\mathbf{A}\mathbf{x}) + \mathbf{A}(\mathbf{x}-\mathbf{x}_1)\|^2 \\
&= \|\mathbf{b}-\mathbf{A}\mathbf{x}\|^2 + \|\mathbf{A}(\mathbf{x}-\mathbf{x}_1)\|^2 \\
&\qquad + 2\big(\mathbf{A}(\mathbf{x}-\mathbf{x}_1)\big)^{\mathrm{T}}(\mathbf{b}-\mathbf{A}\mathbf{x}) \\
&= \|\mathbf{b}-\mathbf{A}\mathbf{x}\|^2 + \|\mathbf{A}(\mathbf{x}-\mathbf{x}_1)\|^2 \\
&\geq \|\mathbf{b}-\mathbf{A}\mathbf{x}\|^2,
\end{aligned}
$$

即 $\mathbf{x}$ 是原方程的最小二乘解.

反之, 设 $\mathbf{x}$ 是原方程的任一最小二乘解. 令 $\mathbf{x}_0 = \mathbf{A}^{\dagger}\mathbf{b}$, 则 $\mathbf{x}_0$ 是正规方程的一个解. 类似于前述推导可得

$$\begin{aligned}\|\mathbf{b}-\mathbf{A}\mathbf{x}\|^2 &= \|\mathbf{b}-\mathbf{A}\mathbf{x}_0\|^2 + \|\mathbf{A}(\mathbf{x}_0-\mathbf{x})\|^2 \\ &\geq \|\mathbf{b}-\mathbf{A}\mathbf{x}_0\|^2.\end{aligned}$$

由 $\mathbf{x}$ 为最小二乘解知 $\|\mathbf{b}-\mathbf{A}\mathbf{x}\|^2 = \|\mathbf{b}-\mathbf{A}\mathbf{x}_0\|^2$, 于是 $\|\mathbf{A}(\mathbf{x}_0-\mathbf{x})\|^2 = 0$, 即 $\mathbf{A}(\mathbf{x}_0-\mathbf{x}) = 0$, 因此

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x}_0 = \mathbf{A}^{\mathrm{T}}\mathbf{b},$$

即 $\mathbf{x}$ 是正规方程的解.

广义逆与不相容方程的解 (续)

3. 原方程的全部最小二乘解可表为

$$\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{y};$$

证明　由 (1) 知正规方程 $\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$ 相容, 其解 (由 (2) 知也是原方程的最小二乘解) 为

$$\begin{aligned}\mathbf{x} &= (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{\dagger}\mathbf{A}^{\mathrm{T}}\mathbf{b} + (\mathbf{I} - (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{\dagger}\mathbf{A}^{\mathrm{T}}\mathbf{A})\mathbf{y} \\ &= \mathbf{A}^{\dagger}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{y}.\end{aligned}$$

4. 若 $\mathbf{x}_1, \mathbf{x}_2$ 为原方程最小二乘解, 则 $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2$.

证明　由 (3) 可知

$$\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^{\dagger}\mathbf{b} + \mathbf{A}(\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{y} = \mathbf{A}\mathbf{A}^{\dagger}\mathbf{b}$$

与 $\mathbf{y}$ 无关. 因此结论成立.

广义逆与线性方程组的解 (总结)

1. 对于相容方程 $\mathbf{Ax} = \mathbf{b}$
   1. 若 $\mathbf{A}$ 满列秩 (包括可逆), 则方程有唯一解 $\mathbf{x} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}$;
   2. 若 $\mathbf{A}$ 非满列秩, 则方程有一般解 $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{y}$, 有最小范数解 $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b}$;
2. 对于不相容方程 $\mathbf{Ax} = \mathbf{b}$
   1. 方程有最小二乘解 $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A})\mathbf{y}$;
   2. 方程有最小范数最小二乘解 $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b}$.

问　题?

最小二乘解的批处理方法

用上面的公式直接计算最小二乘解称为批处理方法.

1. 优点：精度高;
2. 缺点：计算量大, 存储量大.

在应用中更有效的方法是递推最小二乘算法.

最小二乘解的递推计算

假设 $\mathbf{A}, \mathbf{b}$ 按如下方式更新:

$$\mathbf{A}_n = \begin{bmatrix} \mathbf{A}_{n-1} \\ \mathbf{a}_n \end{bmatrix}, \quad \mathbf{b}_n = \begin{bmatrix} \mathbf{b}_{n-1} \\ b_n \end{bmatrix}.$$

若 $\mathbf{A}_{n-1}$ 满列秩 (以后 $\mathbf{A}_n$ 也就满列秩), 则 $\mathbf{x}_n$ 可按如下递推公式计算

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \mathbf{k}_n(b_n - \mathbf{a}_n\mathbf{x}_{n-1})$$

$$\mathbf{k}_n = \frac{\mathbf{P}_{n-1}\mathbf{a}_n^{\mathrm{T}}}{1 + \mathbf{a}_n\mathbf{P}_{n-1}\mathbf{a}_n^{\mathrm{T}}}$$

$$\mathbf{P}_n = (\mathbf{I} - \mathbf{k}_n\mathbf{a}_n)\mathbf{P}_{n-1}$$

其中初值为

$$\mathbf{P}_{n-1} = (\mathbf{A}_{n-1}^{\mathrm{T}}\mathbf{A}_{n-1})^{-1}, \quad \mathbf{x}_{n-1} = \mathbf{P}_{n-1}\mathbf{A}_{n-1}^{\mathrm{T}}\mathbf{b}_{n-1}.$$

最小二乘解的递推计算 (续)

矩阵求逆引理

$$
\begin{aligned}
&(\mathbf{A}+\mathbf{BCD})^{-1}\\
&\quad=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}
\end{aligned}
$$

其中 $\mathbf{A},\mathbf{C}$ 和 $\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1}$ 可逆.

特别地, 若 $\mathbf{A}$ 可逆, 且 $\mathbf{u},\mathbf{v}$ 均为列向量, 则

$$
(\mathbf{A}+\mathbf{uv}^{\mathrm{T}})^{-1}=\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{uv}^{\mathrm{T}}\mathbf{A}^{-1}}{1+\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u}}.
$$

令

$$\mathbf{R}_n = \mathbf{A}_n^{\mathrm{T}}\mathbf{A}_n = \mathbf{A}_{n-1}^{\mathrm{T}}\mathbf{A}_{n-1} + \mathbf{a}_n^{\mathrm{T}}\mathbf{a}_n = \mathbf{R}_{n-1} + \mathbf{a}_n^{\mathrm{T}}\mathbf{a}_n,$$

于是由矩阵求逆引理可得

$$\mathbf{R}_n^{-1} = \mathbf{R}_{n-1}^{-1} - \frac{\mathbf{R}_{n-1}^{-1}\mathbf{a}_n^{\mathrm{T}}\mathbf{a}_n\mathbf{R}_{n-1}^{-1}}{1+\mathbf{a}_n\mathbf{R}_{n-1}^{-1}\mathbf{a}_n^{\mathrm{T}}}.$$

令

$$\mathbf{P}_n = \mathbf{R}_n^{-1},\ \ \mathbf{k}_n = \frac{\mathbf{P}_{n-1}\mathbf{a}_n^{\mathrm{T}}}{1+\mathbf{a}_n\mathbf{P}_{n-1}\mathbf{a}_n^{\mathrm{T}}}.$$

最小二乘解的递推计算 (续)

于是

$$\mathbf{P}_n = \mathbf{P}_{n-1} - \mathbf{k}_n \mathbf{a}_n \mathbf{P}_{n-1} = (\mathbf{I} - \mathbf{k}_n \mathbf{a}_n)\mathbf{P}_{n-1},$$

$$\begin{aligned}
\mathbf{x}_n &= \mathbf{P}_n \mathbf{A}_n^{\mathrm{T}} \mathbf{b}_n \\
&= (\mathbf{I} - \mathbf{k}_n \mathbf{a}_n)\mathbf{P}_{n-1}(\mathbf{A}_{n-1}^{\mathrm{T}} \mathbf{b}_{n-1} + \mathbf{a}_n^{\mathrm{T}} b_n) \\
&= (\mathbf{I} - \mathbf{k}_n \mathbf{a}_n)\mathbf{x}_{n-1} + (\mathbf{I} - \mathbf{k}_n \mathbf{a}_n)\mathbf{P}_{n-1}\mathbf{a}_n^{\mathrm{T}} b_n \\
&= (\mathbf{I} - \mathbf{k}_n \mathbf{a}_n)\mathbf{x}_{n-1} + \mathbf{k}_n b_n \\
&= \mathbf{x}_{n-1} + \mathbf{k}_n (b_n - \mathbf{a}_n \mathbf{x}_{n-1}).
\end{aligned}$$

问　题?

目录

利用奇异值分解

1. 求 $\mathbf{A}$ 的奇异值分解 $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{H}}$;
2. 计算 $\mathbf{A}^{\dagger} = \mathbf{V}\mathbf{D}^{\dagger}\mathbf{U}^{\mathrm{H}}$.

在 MATLAB 中, 有函数 pinv.

设 $\mathbf{A}$ 的秩为 r.

1. 对 $\mathbf{A}$ 左乘一系列 Householder 变换阵 (记其乘积为 $\mathbf{H}$), 右乘列置换阵 $\mathbf{P}$, 使 $\mathbf{A}$ 变为

$$\mathbf{HAP} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{R} \\ \mathbf{O} & \mathbf{O} \end{bmatrix},$$

其中 $\mathbf{T}_1$ 为 $r \times r$ 非奇异上三角阵.

2. 对 $\mathbf{HAP}$ 右乘一系列 Householder 变换阵 (记其乘积为 $\mathbf{H}_1$), 使其变为

$$\mathbf{HAPH}_1 = \begin{bmatrix} \mathbf{T}_1 & \mathbf{R} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{H}_1 = \begin{bmatrix} \mathbf{T} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix},$$

其中 $\mathbf{T}$ 为 $r \times r$ 非奇异上三角阵.

3. 将 $\mathbf{T}^{-1}$ 看成是方程 $\mathbf{TX} = \mathbf{I}$ 的解用三角系数阵的线性方程回代算法求 $\mathbf{T}^{-1}$.
4. $\mathbf{A}^{\dagger} = \mathbf{P}\mathbf{H}_1 \begin{bmatrix} \mathbf{T}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{H}.$

利用满秩分解

1. 求 $\mathbf{A}$ 的满秩分解 $\mathbf{A} = \mathbf{F}\mathbf{G}$;
2. 计算

$$\mathbf{A}^{\dagger} = \mathbf{G}^{\mathrm{H}}(\mathbf{G}\mathbf{G}^{\mathrm{H}})^{-1}(\mathbf{F}^{\mathrm{H}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{H}}.$$

它根据 $\mathbf{A}$ 的前 k 列所构成的子矩阵的广义逆来构造前 $k+1$ 列所构成子矩阵的广义逆矩阵, 直到 $\mathbf{A}$ 的所有列处理完.

这种方法称为阶数递推算法.

设 $\mathbf{A} \in \mathbb{R}^{m\times n}$. 记

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{a}_1, \\ \mathbf{A}_k &= [\mathbf{A}_{k-1}, \mathbf{a}_k], \quad k = 2, \cdots, n. \end{aligned}$$

Greville 方法 (续)

对 $k=2,\cdots,n$ 做

1. 计算 $\mathbf{d}_k = \mathbf{A}_{k-1}^{\dagger}\mathbf{a}_k$;
2. 计算 $\mathbf{c}_k = \mathbf{a}_k - \mathbf{A}_{k-1}\mathbf{d}_k$;
3. 计算
$$\mathbf{b}_k^{\mathrm{T}} = \begin{cases} \mathbf{c}_k^{\dagger} & \mathbf{c}_k \neq \mathbf{0} \\ \dfrac{\mathbf{d}_k^{\mathrm{T}}\mathbf{A}_{k-1}^{\dagger}}{1+\mathbf{d}_k^{\mathrm{T}}\mathbf{d}_k} & \mathbf{c}_k = \mathbf{0} \end{cases}$$
4. 计算
$$\mathbf{A}_k^{\dagger} = \begin{bmatrix} \mathbf{A}_{k-1}^{\dagger} - \mathbf{d}_k\mathbf{b}_k^{\mathrm{T}} \\ \mathbf{b}_k^{\mathrm{T}} \end{bmatrix}$$

Greville 方法的优点

1. 除了向量的广义逆外, 无需计算任何矩阵的逆或广义逆;
2. 只利用矩阵乘法即可求得 Moore-Penrose 逆.

Greville 方法 (例)

设 $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}$, 利用 Greville 方法求 $\mathbf{A}^\dagger$.

$$\mathbf{A}_1^\dagger = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}^\dagger = \frac{1}{5}[1\ \ 2\ \ 0],$$

$$\mathbf{d}_2 = \mathbf{A}_1^\dagger \mathbf{a}_2 = \frac{1}{5}[1\ \ 2\ \ 0] \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \frac{2}{5},$$

$$\mathbf{c}_2 = \mathbf{a}_2 - \mathbf{A}_1 \mathbf{d}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \frac{2}{5} = \begin{bmatrix} -2/5 \\ 1/5 \\ 1 \end{bmatrix},$$

Greville 方法 (例)

$$\mathbf{b}_2^{\mathrm{T}} = \mathbf{c}_2^{\dagger} = \frac{1}{6}[-2\ \ 1\ \ 5],$$

$$\mathbf{A}^{\dagger} = \begin{bmatrix} \frac{1}{5}[1\ \ 2\ \ 0] - \frac{2}{5}\frac{1}{6}[-2\ \ 1\ \ 5] \\ \frac{1}{6}[-2\ \ 1\ \ 5] \end{bmatrix}$$

$$= \frac{1}{6}\begin{bmatrix} 2 & 2 & -2 \\ -2 & 1 & 5 \end{bmatrix}.$$

作业

分别利用上述四种方法求第 296 页 $\mathbf{A}_4$ 的广义逆.