5-1 线性最小二乘问题

最小二乘问题

无约束最小二乘问题可定义如下：

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{m}[r_i(\boldsymbol{x})]^2 = \frac{1}{2}\boldsymbol{r}(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}), \quad m \geq n, \tag{5.1.1}$$

其中：$\boldsymbol{x} \in \mathbb{R}^n$为待定参数，$\boldsymbol{r}(\boldsymbol{x}) = (r_1(\boldsymbol{x}), r_2(\boldsymbol{x}), \ldots, r_m(\boldsymbol{x}))^T$为**剩余函数**(或**残量函数**)。在点$\boldsymbol{x}$处的值称为**剩余量**。

根据剩余函数，最小二乘问题可分为：

- **线性最小二乘问题**
  所有$r_i(\boldsymbol{x})$都是$\boldsymbol{x}$的线性函数;
- **非线性最小二乘问题**
  至少存在一个$r_i(\boldsymbol{x})$是$\boldsymbol{x}$的非线性函数。

线性最小二乘问题:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{m}[\boldsymbol{a}_i^T\boldsymbol{x} - b_i]^2,$$

这里$r_i(\boldsymbol{x}) = \boldsymbol{a}_i^T\boldsymbol{x} - b_i$,, 其中$\boldsymbol{a}_i = (a_{i1}, a_{i2}, \cdots, a_{in})^T$, $b_i \in \mathbb{R}$。

最小二乘矩阵向量表示:

$$\min \quad f(\boldsymbol{x}) = \frac{1}{2}\|A\boldsymbol{x} - \boldsymbol{b}\|^2, \tag{5.1.2}$$

其中:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

应用1：数据拟合

设已知m组观测值:

$$\begin{array}{ccccc} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}$$

这里:

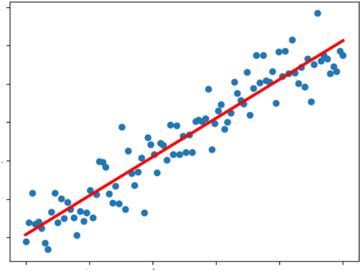
- $\boldsymbol{a}_i = (a_{i1}, a_{i2}, \cdots, a_{in})^T$为输入数据;
- b_i为对应观测结果;

二者之间估计函数关系为：

$$\boldsymbol{b}_i \approx \boldsymbol{a}_i^T \boldsymbol{x}, \quad i = 1, 2, \cdots, m$$

为确定n维待定变量$\boldsymbol{x}$，在ℓ_2范数意义下优化下述问题：

$$\min \sum_{i=1}^{m} [\boldsymbol{a}_i^T \boldsymbol{x} - b_i]^2$$



应用2：方程组求解

科学计算中常见的方程组求解：

$$\boldsymbol{r}(\boldsymbol{x}) = \begin{pmatrix} r_1(\boldsymbol{x}) \\ r_2(\boldsymbol{x}) \\ \vdots \\ r_m(\boldsymbol{x}) \end{pmatrix} = 0.$$

方程组类型：

- 当$m > n$时，超定方程组
- 当$m = n$时，适定方程组
- 当$m < n$时，欠定方程组。
- 当$r_i(\boldsymbol{x})$都是x的线性函数时，线性方程组。

最小二乘方法是常用的方程组求解方法，尤其对于超定方程组的求解。

线性最小二乘问题解法

线性最小二乘问题需优化以下问题：

$$\min \quad q(\boldsymbol{x}) = \frac{1}{2}\|A\boldsymbol{x} - \boldsymbol{b}\|^2, \tag{5.1.3}$$

其中: $A = (a_{ij})$为$m \times n$阶矩阵，$\boldsymbol{b} \in \mathbb{R}^m$为$m$维向量，$\|\cdot\|$为向量的$\ell_2$范数。

将$\|A\boldsymbol{x} - \boldsymbol{b}\|^2$按范数展开得二次问题：

$$\min \quad q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A^T A\boldsymbol{x} - \boldsymbol{b}^T A\boldsymbol{x} + \frac{1}{2}\boldsymbol{b}^T\boldsymbol{b}. \tag{5.1.4}$$

二阶Hesse矩阵$A^T A$为对称且至少半正定，问题(5.1.4)的任何最优解也是全局最优解。

解线性最小二乘问题

设$\boldsymbol{x}^*$为问题(5.1.4)的最优解，

- 如果$m \leq n$，则必有$\|A\boldsymbol{x}^* - \boldsymbol{b}\| = 0$，
- 如果$m > n$，则$\|A\boldsymbol{x}^* - \boldsymbol{b}\|$可能非零。当方程组$A\boldsymbol{x} = \boldsymbol{b}$有不兼容方程时，问题的最优值不为零。

根据无约束最优一阶必要条件，有：

$$\nabla q(\boldsymbol{x}) = A^T A\boldsymbol{x} - A^T\boldsymbol{b} = 0$$

继而有**正规方程组(法方程组)**：

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b} \tag{5.1.5}$$

情形1：若矩阵A列满秩，即A的列线性无关，此时矩阵A^TA正定，正规方程组(5.1.5)的**解唯一**且可表示为：

$$\boldsymbol{x}^* = (A^TA)^{-1}A^T\boldsymbol{b} = A^+\boldsymbol{b}, \tag{5.1.6}$$

其中：记A的广义逆矩阵为：

$$A^+_{n\times m} = (A^TA)^{-1}A^T.$$

例题

例5.1.1 给定方程组：

$$\begin{cases} x_1 + x_2 & = 3, \\ 2x_1 - 3x_2 & = 2, \\ -x_1 + 4x_2 & = 4. \end{cases}$$

利用最小二乘法解该方程组。

解：记系数矩阵为A, 右端向量为$\boldsymbol{b}$：

$$A = \begin{bmatrix} 1 & 1 \\ 2 & -3 \\ -1 & 4 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}$$

求函数：$f(\boldsymbol{x}) = (A\boldsymbol{x} - \boldsymbol{b})^T(A\boldsymbol{x} - \boldsymbol{b})$的极小点。

例题

解续:

$$A^TA=\begin{bmatrix}6 & -9\\ -9 & 26\end{bmatrix}, (A^TA)^{-1}=\frac{1}{75}\begin{bmatrix}26 & 9\\ 9 & 6\end{bmatrix}$$

由正规方程组:

$$\boldsymbol{x}=(A^TA)^{-1}A^T\boldsymbol{b}=\frac{1}{75}\begin{bmatrix}26 & 9\\ 9 & 6\end{bmatrix}\begin{bmatrix}1 & 2 & -1\\ 1 & -3 & 4\end{bmatrix}\begin{bmatrix}3\\ 2\\ 4\end{bmatrix}=\begin{bmatrix}\frac{13}{5}\\ \frac{7}{5}\end{bmatrix}$$

$\square$

解线性最小二乘问题

方程组(5.1.6)求解也可以通过其它有效方法进行，如

- 可用一般凸二次函数的无约束最优化方法求得最小二乘问题的解，如共轭梯度法。
- 若矩阵A^TA正定，可用分解方法求解：
  - 采用矩阵A^TA的Cholesky分解LL^T，其中L为下三角矩阵
  - 或LDL^T分解，这里的L为对角线元素全为1的单位下三角矩阵

  然后再通过前代和回代求得问题的最优解$\boldsymbol{x}^*$.

计算A^TA后进行分解求解方程组将产生解的误差。

误差分析

假设A^TA存在误差E，其产生解的误差为$\delta \boldsymbol{x}^*$，则由方程组(5.1.5)可得：

$$(A^TA+E)(\boldsymbol{x}+\delta\boldsymbol{x}^*)=\boldsymbol{b}, \tag{5.1.7}$$

解的相对误差与矩阵的相对误差之间有如下关系：

$$\frac{\|\delta\boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|}\le \mathrm{cond}(A^TA)\frac{\|E\|}{\|A^TA\|}, \tag{5.1.8}$$

其中：矩阵A^TA的条件数定义为：

$$\mathrm{cond}(A^TA)=\sigma_1^2/\sigma_n^2=\mathrm{cond}(A)^2$$

这里σ_1和σ_n分别是矩阵A的最大和最小奇异值。

QR正交分解– 降低误差

对增广矩阵$[A\quad \boldsymbol{b}]$作QR正交分解。注意ℓ_2范数在正交变换下具有不变性，即

$$\|Q(A\boldsymbol{x}-\boldsymbol{b})\|_2^2=\|QA\boldsymbol{x}-Q\boldsymbol{b}\|_2^2=\|A\boldsymbol{x}-\boldsymbol{b}\|_2^2.$$

设

$$A=QR=[Q_1\ Q_2]\begin{bmatrix}R_1\\0\end{bmatrix}=Q_1R_1 \tag{5.1.9}$$

和

$$[A\quad \boldsymbol{b}]=Q[R\quad \bar{\boldsymbol{b}}] \tag{5.1.10}$$

其中:

- Q为$m\times m$阶正交矩阵，$Q_1\in R^{m\times n}$，$Q_2\in R^{m\times(m-n)}$
- $\bar{\boldsymbol{b}}=Q^T\boldsymbol{b}$, R_1为$n\times n$上三角矩阵。

QR正交分解– 降低误差

由$A^TA\boldsymbol{x}=A^T\boldsymbol{b}$有：

$$R_1^TR_1\boldsymbol{x}=R_1^TQ_1^T\boldsymbol{b}=R_1^T\bar{\boldsymbol{b}}_n, \tag{5.1.11}$$

最优解$\boldsymbol{x}^*$可由三角方程组

$$R_1\boldsymbol{x}=\bar{\boldsymbol{b}}_n \tag{5.1.12}$$

经回代确定，其中$\bar{\boldsymbol{b}}_n=Q_1^T\boldsymbol{b}$.

矩阵R_1的条件数即为矩阵A的条件数

$$\mathrm{cond}(R_1)=\sigma_1/\sigma_n$$

通过上述方程组确定的最优解的相对误差要远小于通过先计算积矩阵再分解后所求得解的相对误差。

无穷多解情形

情形2：当矩阵A的列线性相关，即A不是列满秩时，在上述矩阵A的QR正交分解式中，R_1不再是一个$n \times n$而是$r \times n$ 阶的上三角矩阵, 确定最优解的方程组成为:

$$R_1 \boldsymbol{x} = \bar{\boldsymbol{b}}_r, \tag{5.1.13}$$

其中$r = \mathrm{rank}(A) < n$表示矩阵A的秩数，$\bar{b}_r$表示由向量$\bar{\boldsymbol{b}}$的前r个分量组成的向量。

方程组(5.1.13)有无穷多组解。由矩阵A的奇异值分解算法, 从中确定使ℓ_2范数最小的解x^*，即

$$\|\boldsymbol{x}^*\| = \min\{\|\boldsymbol{x}\| | \boldsymbol{x} \in X^*\},$$

其中X^*是最小二乘问题(5.1.5)所有解的集合。

奇异值分解处理

设$m\times n(m>n)$阶矩阵A的秩为$r<n$，则存在$m\times m$阶正交矩阵U和$n\times n$阶正交矩阵V使得

$$A=USV^T=U\begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix}V^T,$$

其中：

- S为$m\times n$阶的块对角矩阵;
- $\Sigma=\mathrm{diag}(\sigma_1,\ldots,\sigma_r)$为$r\times r$阶对角矩阵;
- $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>0$为矩阵$A$的奇异值。

满足上述要求使ℓ_2范数取极小的最优解为：

$$\boldsymbol{x}^*=VS^+U^T\boldsymbol{b}=A^+\boldsymbol{b}$$

奇异值分解处理

其中：广义逆$A^+ = VS^+U^T$且$S^+ = \begin{bmatrix} \Sigma^+ & 0 \\ 0 & 0 \end{bmatrix}$。而$\Sigma^+ = \begin{bmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-} \end{bmatrix}$

当$r = \mathrm{rank}(A) = n$时，由奇异值分解所确定的矩阵A^+即为前述定义的广义逆矩阵:

$$A^+ = VS^+U^T = (A^TA)^{-1}A^T.$$