

第三讲 线性最小二乘问题

线性最小二乘问题的求解方法



目录

3.1 问题介绍

3.2 Householder 变换和 Givens 变换

3.3 QR 分解

3.4 奇异值分解

3.5 线性最小二乘问题的求解方法

3-5

线性最小二乘问题的求解方法

3.5 线性最小二乘问题的求解方法

3.5.1 正规方程

3.5.2 Cholesky 分解法

3.5.3 QR 分解法

3.5.4 SVD 分解法

3-5-1 | 正规方程

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

定理 设 $A \in \mathbb{R}^{m \times n}$ ($m \geq n$), 则 $x_* \in \mathbb{R}^n$ 是线性最小二乘问题的解当且仅当残量 $r = b - Ax_*$ 与 $\text{Ran}(A)$ (值域) 正交, 即 x_* 是下面的**正规方程** (或**法方程**) 的解

$$A^\top(b - Ax) = 0 \quad \text{或} \quad A^\top Ax = A^\top b. \quad (3.1)$$

(板书)

解的存在性与唯一性

由于

$$A^T b \in \text{Ran}(A^T) = \text{Ran}(A^T A),$$

因此正规方程 $A^T A x = A^T b$ 是相容 (consistent) 的, 即 **最小二乘解总是存在的**.

解的存在性与唯一性

由于

$$A^T b \in \text{Ran}(A^T) = \text{Ran}(A^T A),$$

因此正规方程 $A^T A x = A^T b$ 是相容 (consistent) 的, 即 **最小二乘解总是存在的**.

定理 设 $A \in \mathbb{R}^{m \times n}$ ($m \geq n$), 则 $A^T A$ 对称正定当且仅当 A 是列满秩的, 即 $\text{rank}(A) = n$. 此时, 线性最小二乘问题的解是唯一的, 其表达式为

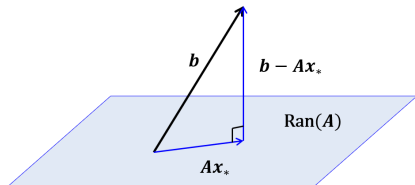
$$x_* = (A^T A)^{-1} A^T b.$$

最小二乘解的几何含义

根据最小二乘问题与正规方程的关系, 我们可以把 b 写成

$$b = Ax_* + r, \quad \text{其中} \quad r \perp \text{Ran}(A).$$

所以 Ax_* 就是 b 在 $\text{Ran}(A)$ 上的正交投影, 见右图.



需要指出的是, 最小二乘解可能并不唯一, 但上述分解是唯一的.

3-5-2

Cholesky 分解法

 当 A 列满秩时, 我们就可以使用 Cholesky 分解来求解正规方程.

3-5-2

Cholesky 分解法

 当 A 列满秩时, 我们就可以使用 Cholesky 分解来求解正规方程.

计算复杂度

- ▶ 计算 $A^T A$: $\rightarrow mn^2$ % 只需计算下三角或上三角部分
- ▶ 计算 Cholesky 分解: $\rightarrow \frac{1}{3}n^3$
- ▶ 回代求解: $\rightarrow \mathcal{O}(n^2)$

总的运算量大约为 $mn^2 + \frac{1}{3}n^3 + \mathcal{O}(n^2)$

Cholesky 分解求解正规方程的优缺点

- 优点: 运算量最小, 简单直观.
- 缺点: $A^T A$ 的条件数是 A 的平方, 对于病态问题, 不建议使用.

Cholesky 分解求解正规方程的优缺点

- 优点: 运算量最小, 简单直观.
- 缺点: $A^T A$ 的条件数是 A 的平方, 对于病态问题, 不建议使用.

例 设 $A = \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon & & \\ & \varepsilon & \\ & & \varepsilon \end{bmatrix}$, 则 $A^T A = \begin{bmatrix} 1 + \varepsilon^2 & 1 & 1 \\ 1 & 1 + \varepsilon^2 & 1 \\ 1 & 1 & 1 + \varepsilon^2 \end{bmatrix}$.

🌸 记 ε_u 为机器精度, 则当 $\varepsilon_u < \varepsilon < \sqrt{\varepsilon_u}$ 时有 $\varepsilon^2 < \varepsilon_u$, 由于舍入误差的原因, 通过浮点运算计算得到的 $A^T A$ 是奇异的. 但我们注意到 A 是满秩的.

3-5-3 | QR 分解法

设 $A \in \mathbb{R}^{m \times n}$ ($m \geq n$) 是满秩的, 其 QR 分解为 $A = QR$, 则

$$x_* = (A^T A)^{-1} A^T b = (R^T Q^T Q R)^{-1} R^T Q^T b = (R^T R)^{-1} R^T Q^T b = R^{-1} Q^T b$$

3-5-3

QR 分解法

设 $A \in \mathbb{R}^{m \times n}$ ($m \geq n$) 是满秩的, 其 QR 分解为 $A = QR$, 则

$$x_* = (A^T A)^{-1} A^T b = (R^T Q^T Q R)^{-1} R^T Q^T b = (R^T R)^{-1} R^T Q^T b = R^{-1} Q^T b$$

- ✍ QR 分解法的运算量大约为 $2mn^2$ (如果采用 Householder 变换, 则运算量大约为 $4mn^2 - 4n^3/3$). 当 $m \gg n$ 时, 大约为正规方程的两倍.
- ✍ QR 分解法比较稳定, 是当前求解最小二乘问题的 **首选方法**, 特别是当 A 条件数较大 (病态) 时.

3-5-4

奇异值分解法

设 A 列满秩, 其奇异值分解为 $A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T$. 令 U_n 为 U 的前 n 列, 即 $U = [U_n, \tilde{U}]$, 则

$$\begin{aligned} \|Ax - b\|_2^2 &= \left\| U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T x - b \right\|_2^2 = \left\| \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T x - [U_n, \tilde{U}]^T b \right\|_2^2 \\ &= \left\| \begin{bmatrix} \Sigma V^T x - U_n^T b \\ -\tilde{U}^T b \end{bmatrix} \right\|_2^2 = \|\Sigma V^T x - U_n^T b\|_2^2 + \|\tilde{U}^T b\|_2^2 \geq \|\tilde{U}^T b\|_2^2, \end{aligned}$$

等号当且仅当 $\Sigma V^T x - U_n^T b = 0$ 时成立, 即

$$x_* = (\Sigma V^T)^{-1} U_n^T b = V \Sigma^{-1} U_n^T b$$

关于 SVD 法的几点说明

- 相比于 Cholesky 分解法和 QR 分解法, SVD 法具有更高的健壮性
- 计算 SVD 的运算量 **远超** Cholesky 分解法和 QR 分解法.
- 只有当系数矩阵秩亏或者接近秩亏时才使用 (此时 QR 分解法可能会失效).

例 分别用三种方法求解最小二乘问题, 比较运算时间.

(LS_3methods.m)

三种方法的运算时间如下 ($A \in \mathbb{R}^{2n \times n}$, $b \in \mathbb{R}^n$, 以秒为单位):

n	正规方程法	QR 分解法	奇异值分解法
500	0.0050	0.0220	0.0370
1000	0.0160	0.0340	0.1440
1500	0.0490	0.1330	0.5530
2000	0.0870	0.2070	1.4840
2500	0.1910	0.4430	3.1160
3000	0.2500	0.6950	5.9600
3500	0.4640	1.2130	10.0500
4000	0.4940	1.5700	14.8750
4500	0.6690	2.1680	20.6410
5000	1.0720	2.9350	28.6360

补充: 最小二乘问题的应用

最小二乘问题的应用

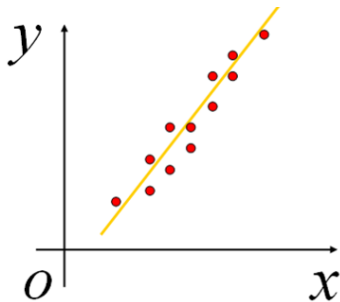
- 多项式数据拟合
- 线性预测
- 信号恢复
- 图像恢复

应用: 多项式数据拟合

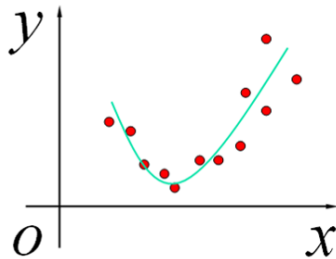
已知 m 个数据样本 $\{(x_i, f_i)\}_{i=1}^m$, 寻找一个低次多项式来拟合这些数据.

应用: 多项式数据拟合

已知 m 个数据样本 $\{(x_i, f_i)\}_{i=1}^m$, 寻找一个低次多项式来拟合这些数据.



$$y = ax + b$$



$$y = ax^2 + bx + c$$

设拟合多项式为

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n, \quad (m \gg n)$$

设拟合多项式为

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n, \quad (m \gg n)$$

➤ 将上述 m 个点 $\{(x_i, f_i)\}_{i=1}^m$ 代入可得

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}_{m \times (n+1)} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} \quad \text{或} \quad Ax = f,$$

➤ 超定方程组, 解通常不存在, 需转化为求解最小二乘问题

$$\min_{x \in \mathbb{R}^{n+1}} \|f - Ax\|_2^2$$

应用: 线性预测

预测一个时间序列的未来走向, 其中一个常用方法就是 线性预测.

应用: 线性预测

预测一个时间序列的未来走向, 其中一个常用方法就是 **线性预测**.

➤ 假定时间序列在 t_k 时刻的值 f_k 线性依赖于前 n 个时刻的值, 即

$$f_k = a_1 f_{k-1} + a_2 f_{k-2} + \cdots + a_n f_{k-n} \quad (k \geq n)$$

❓ 已测得前 m 个值: $f_i, i = 0, 1, \dots, m-1$, 如何预测未来的取值? ($m \gg n$)

应用: 线性预测

预测一个时间序列的未来走向, 其中一个常用方法就是 **线性预测**.

➤ 假定时间序列在 t_k 时刻的值 f_k 线性依赖于前 n 个时刻的值, 即

$$f_k = a_1 f_{k-1} + a_2 f_{k-2} + \cdots + a_n f_{k-n} \quad (k \geq n)$$

❓ 已测得前 m 个值: $f_i, i = 0, 1, \dots, m-1$, 如何预测未来的取值? ($m \gg n$)

➤ 将现有的数据代入关系式可得

$$\begin{bmatrix} f_{n-1} & f_{n-2} & f_{n-3} & \cdots & f_0 \\ \vdots & \vdots & \vdots & & \vdots \\ f_{m-2} & f_{m-3} & f_{m-4} & \cdots & f_{m-n-1} \end{bmatrix}_{(m-n) \times n} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f_n \\ \vdots \\ f_{m-1} \end{bmatrix} \quad \text{或} \quad Ax = f,$$

这也是一个超定问题, 其解也需通过求解最小二乘问题来获得.

应用: 信号恢复

在获取数字信号时, 由于各种各样的原因, 最后得到的信号总会带有一定的噪声, 即

$$b = x + \eta$$

其中 x 是真实的信号, b 是观察到的信号, η 是噪声.

应用: 信号恢复

在获取数字信号时, 由于各种各样的原因, 最后得到的信号总会带有一定的噪声, 即

$$b = x + \eta$$

其中 x 是真实的信号, b 是观察到的信号, η 是噪声.

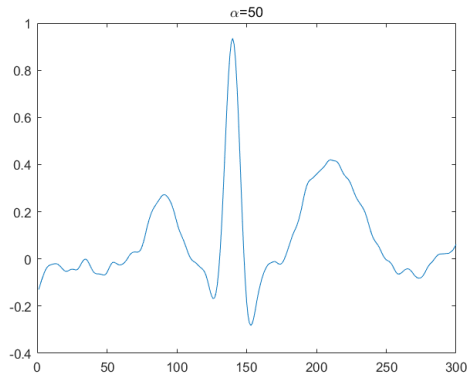
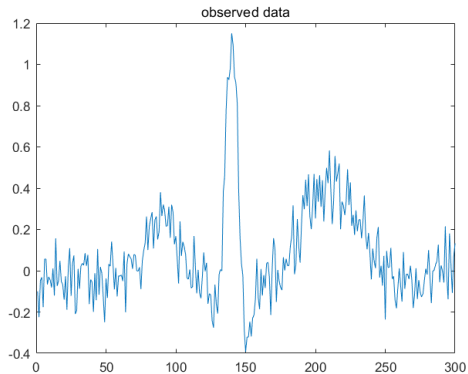


去噪是数字信号和图像处理的基本问题, 有效方法之一是加权最小二乘法

$$\min_x \frac{1}{2} \|x - b\|_2^2 + \frac{1}{2} \alpha \|Dx\|_2^2$$

其中 D 是离散的二阶导算子或 TV 算子.

例 去噪举例: LS_denoise.m



应用: 图像恢复

数字图像的获取模型

$$f = \mathcal{B}(x) + \eta$$

其中 x 是真实图像, f 是观察到的图像, $\mathcal{B}(\cdot)$ 是卷积算子 (模糊机制), η 是噪声.

应用: 图像恢复

数字图像的获取模型

$$f = \mathcal{B}(x) + \eta$$

其中 x 是真实图像, f 是观察到的图像, $\mathcal{B}(\cdot)$ 是卷积算子 (模糊机制), η 是噪声.



Tikhonov regularization model

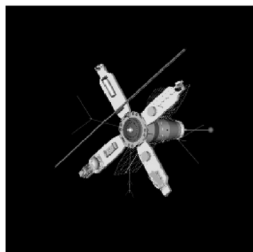
$$\min_x \|\mathcal{B}(x) - f\|_2^2 + \mu^2 \|x\|_2^2$$



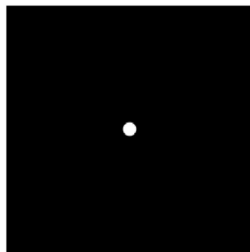
Weighted regularization model

$$\min_x \|\mathcal{B}(x) - f\|_2^2 + \mu^2 \|Dx\|_2^2$$

例 由于对焦不准而造成图像模糊 (此时 B 是一个线性算子), 噪声为高斯白噪声.

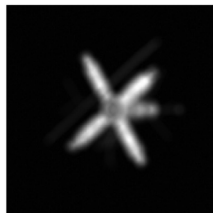


(a)



(b)

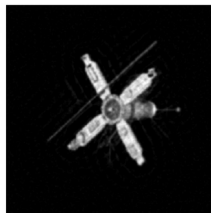
(a) The original satellite image; (b) the out-of-focus blurring function.



(a)



(b)



(c)

(a) The blurred and noisy image (b) the restored image using the spatial invariant model;
(c) the restored image using the spatial variant model

谢谢
THANK YOU

