《线性回归》
—线性回归(8)

杨　　瑛

清华大学　数学科学系
Email: yangying@mail.tsinghua.edu.cn
Tel: 62796887

2019.04.04

主要内容：LSE的性质

- 设计矩阵列正交
- 有线性约束的LSE
- 其它专题: 广义最小二乘回归
- 其它专题：问题
- 其它专题: 中心化和尺度化协变量
- 其它专题: Bayes估计
- 其它专题: 稳健估计

设计矩阵列正交

在线性模型

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon$$

中，当设计矩阵$\mathbf{X}$表示为列的形式：

$$\mathbf{X} = (\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(p-1)}),$$

列向量是正交的。

- θ的LSE有如下的表达式：

$$\begin{aligned} \hat{\theta} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y \\ &= \begin{pmatrix} (\mathbf{x}^{(0)\prime}\mathbf{x}^{(0)})^{-1}\mathbf{x}^{(0)\prime}\mathbf{Y} \\ (\mathbf{x}^{(1)\prime}\mathbf{x}^{(1)})^{-1}\mathbf{x}^{(1)\prime}\mathbf{Y} \\ \vdots \\ (\mathbf{x}^{(p-1)\prime}\mathbf{x}^{(p-1)})^{-1}\mathbf{x}^{(p-1)\prime}\mathbf{Y} \end{pmatrix} \end{aligned}$$

- 此时，
$$\hat{\theta}_j = \mathbf{x}^{(j)\prime}\mathbf{Y}/\mathbf{x}^{(j)\prime}\mathbf{x}^{(j)},$$
- 相当于$\mathbf{Y}$对单个变量$\mathbf{x}_j$做通过原点的回归$\mathbf{E}[\mathbf{Y}] = \mathbf{x}^{(j)}\beta_j$.

有线性约束的LSE

- 考虑线性模型

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \epsilon \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n),$$

  其中$\mathbf{X}$是秩为p的$n \times p$的设计矩阵。
- 有时要考虑约束条件：$\mathbf{A}\theta = \mathbf{c}$, 其中$\mathbf{A}$是秩为$q$的$q \times p$矩阵，$\mathbf{c}$是$q \times 1$的列向量。
- 利用Lagrange乘子法可以求解这个问题。上面的约束相当于q个约束：$\mathbf{a}_i^T\theta = c_i, (i = 1, \cdots, q)$, 其中$\mathbf{a}_i^T$是$\mathbf{A}$的行向量。
- 注意：

$$\sum_{i=1}^{q} \lambda_i(\mathbf{a}_i^T\theta - c_i) = \lambda^T(\mathbf{A}\theta - \mathbf{c}) = (\theta^T\mathbf{A}^T - \mathbf{c}^T)\lambda.$$

- 定义
$$L(\theta,\lambda)=\|\mathbf{Y}-\mathbf{X}\theta\|^2+(\theta^T\mathbf{A}^T-\mathbf{c}^T)\lambda$$
- 关于$L(\theta,\lambda)$求最小值，相当于求解：
$$\begin{aligned}\mathbf{A}\theta &= \mathbf{c}, &(1)\\ -2\mathbf{X}^T\mathbf{Y}+2\mathbf{X}^T\mathbf{X}\theta+\mathbf{A}^T\lambda &= \mathbf{0}. &(2)\end{aligned}$$
设$\hat{\theta}_H$和$\widehat{\lambda}_H$表示上面方程组的解。
- 由(2)得
$$\begin{aligned}\hat{\theta}_H &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^TY-\frac{1}{2}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T\hat{\lambda}_H\\ &= \hat{\theta}-\frac{1}{2}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T\hat{\lambda}_H, &(3)\end{aligned}$$

- 从(1),得到:

$$\mathbf{c} = \mathbf{A}\hat{\theta}_H = \mathbf{A}\hat{\theta} - \frac{1}{2}\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T\hat{\lambda}_H.$$

- 由于$(\mathbf{X}^T\mathbf{X})^{-1}$正定，可以证明$\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T$是对称正定的，因而是非奇异的。故有

$$-\frac{1}{2}\hat{\lambda}_H = [\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(\mathbf{c} - \mathbf{A}\hat{\theta})$$

带入(3), 得到

$$\hat{\theta}_H = \hat{\theta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T[\mathbf{A}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{A}^T]^{-1}(\mathbf{c} - \mathbf{A}\hat{\theta}). \tag{4}$$

- 还要证明：在条件$\mathbf{A}\theta=\mathbf{c}$之下，$\hat{\theta}_H$确实是$\|\mathbf{Y}-\mathbf{X}\theta\|^2$的最小值点。
- 注意到：

$$
\begin{aligned}
\|\mathbf{X}(\hat{\theta}-\theta)\|^2 &= (\hat{\theta}-\theta)\mathbf{X}^T\mathbf{X}(\hat{\theta}-\theta) \\
&= (\hat{\theta}-\hat{\theta}_H+\hat{\theta}_H-\theta)\mathbf{X}^T\mathbf{X}(\hat{\theta}-\hat{\theta}_H+\hat{\theta}_H-\theta) \\
&= (\hat{\theta}-\hat{\theta}_H)\mathbf{X}^T\mathbf{X}(\hat{\theta}-\hat{\theta}_H) \\
&\quad +(\hat{\theta}_H-\theta)\mathbf{X}^T\mathbf{X}(\hat{\theta}_H-\theta) \qquad (5) \\
&= \|\mathbf{X}(\hat{\theta}-\hat{\theta}_H)\|^2+\|\mathbf{X}(\hat{\theta}_H-\theta)\|^2, \qquad (6)
\end{aligned}
$$

(5)成立是因为由(3),

$$
2(\hat{\theta}-\hat{\theta}_H)\mathbf{X}^T\mathbf{X}(\hat{\theta}_H-\theta)=\hat{\lambda}_H'\mathbf{A}((\hat{\theta}_H-\theta))=\hat{\lambda}_H'(\mathbf{c}-\mathbf{c})=0. \qquad (7)
$$

- 从(6)得：

$$\begin{aligned}\|\mathbf{Y}-\mathbf{X}\theta\|^2 &= \|\mathbf{Y}-\mathbf{X}\hat{\theta}\|^2+\|\mathbf{X}(\hat{\theta}_H-\theta)\|^2\\ &= \|\mathbf{Y}-\mathbf{X}\hat{\theta}\|^2+\|\mathbf{X}(\hat{\theta}-\hat{\theta}_H)\|^2+\|\mathbf{X}(\hat{\theta}_H-\theta)\|^2\end{aligned} \tag{8}$$

在$\|\mathbf{X}(\hat{\theta}_H-\theta)\|^2=0$时取得最小值，即：$\mathbf{X}(\hat{\theta}_H-\theta)=0$，或者$\theta=\hat{\theta}_H$ (因为$\mathbf{X}$的列是线性独立的，列满秩)。

- 令$\theta=\hat{\theta}_H$, 则得到有用的等式：

$$\|\mathbf{Y}-\mathbf{X}\hat{\theta}_H\|^2=\|\mathbf{Y}-\mathbf{X}\hat{\theta}\|^2+\|\mathbf{X}(\hat{\theta}-\hat{\theta}_H)\|^2. \tag{9}$$

或者，$\widehat{\mathbf{Y}}=\mathbf{X}\hat{\theta}$, $\widehat{\mathbf{Y}}_H=\mathbf{X}\hat{\theta}_H$, 直接得到

$$\|\mathbf{Y}-\widehat{\mathbf{Y}}_H\|^2-\|\mathbf{Y}-\widehat{\mathbf{Y}}\|^2=\|\widehat{\mathbf{Y}}-\widehat{\mathbf{Y}}_H)\|^2. \tag{10}$$

其它专题:广义最小二乘回归

- 广义最小二乘回归
  考虑线性模型
  $$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \epsilon \sim N_n(\mathbf{0}, \sigma^2\Sigma), \tag{11}$$
  其中$\mathbf{X}$是秩为p的$n \times p$的设计矩阵, Σ是已知的$n \times n$的正定阵，σ^2未知。
- 对Σ做分解：$\Sigma = \mathbf{K}\mathbf{K}^T$, 可到新的线性模型：
  $$\mathbf{K}^{-1}\mathbf{Y} = \mathbf{K}^{-1}\mathbf{X}\theta + \mathbf{K}^{-1}\epsilon, \mathbf{K}\epsilon \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n).$$
  由此可以利用普通的LS得到θ的估计θ_G.
- 细节详见: Seber and Lee (2003), p. 66-69.
- 模型(11)的参数θ也可以极小化$\|\mathbf{Y} - \mathbf{X}\theta\|^2$得到$\hat{\theta}$。
- 【作业】：试比较$\hat{\theta}_G$和$\hat{\theta}$的性质，试构造σ^2的两种估计。

其它专题：问题

- 考虑线性模型

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon, \epsilon \sim N_n(\mathbf{0}, \Sigma), \tag{12}$$

其中$\mathbf{X}$是秩为p的$n \times p$的设计矩阵, Σ是未知的$n \times n$正定阵。

- θ可以极小化$\|\mathbf{Y} - \mathbf{X}\theta\|^2$的$\theta$的LSE。
- 现在的问题：如何估计矩阵Σ，并对所得到的估计研究其性质。

其它专题: 中心化和尺度化协变量

- 为得到回归系数的合理解释，有时需要对协变量进行中心化或者尺度化。
- 结合前面关于回归系数的解释，阅读Seber and Lee (2003), p. 69–72.

其它专题: Bayes估计

- 需要对线性模型中的参数θ或者σ^2添加先验分布，然后做估计或者推断。
- 想一下：如何给这些参数添加先验分布？
- 阅读Seber and Lee (2003), p. 73-77.

其它专题: 稳健估计

- 后面有ppt简述这个问题。
- 亦可阅读Seber and Lee (2003), p. 77-95.

【第十二讲】