

用奇异值分解求解回归方程的迭代加细

吕纯濂¹⁾ 陈舜华²⁾

(1) 南京气象学院基础科学系; 2) 南京气象学院大气科学系, 南京 210044)

摘 要 在解回归方程时, 设计矩阵中的共线性可能产生不精确的参数估计。用奇异值分解的正交矩阵变换可以减少共线性的影响。通过迭代加细过程还可以改进回归系数的估计。本文描述了求解回归方程的奇异值分解的迭代加细过程。结果表明, 在设计矩阵高度共线性时, 用奇异值分解的迭代加细可以改进回归系数的估计。

关键词 回归方程, 共线性, 正交分解, 迭代加细

分类号 O212. 4

本文描述对用奇异值分解所得回归方程的解作迭代加细的过程。该算法用到在迭代加细过程中由奇异值分解得到的一些矩阵。与 SAS 和 GAUSS 的有关过程作了比较和讨论。

回归方程用矩阵符号通常写为

$$Y = XB + E \quad (1)$$

B 是要估计的未知($p \times 1$) 参数向量; X 是自变量观测($n \times p$) 矩阵, 通常称为设计矩阵, X 的第一列皆为 1, 为常数项系数; Y 是因变量观测值($n \times 1$) 向量; E 为未知剩余($n \times 1$) 向量, 假定服从具有零均值和常数方差 σ^2 的正态分布。

估计 B 常用的方法是最小二乘法, 使剩余平方和 $E^T E$ 达到最小, 而产生正规方程组

$$X^T X B = X^T Y \quad (2)$$

众所周知, 回归方程求解的精确程度可能受到数据中共线性的不利影响, 共线性是自变量观测值矩阵 X 中的问题, 到目前为止, 虽还没有共线性的唯一的精确定义^[1], 但一般认为, 共线性与矩阵 X 各列之间的线性相关有关, 这种线性相关不必是“精确”的, 或者说, 对 X 不必是满秩的, 都会出现共线性的问题^[1]。

为减少共线性的影响, 通常用设计矩阵 X 的一种正交分解, 比如 QR 分解, 有许多方法可以对该分解进行所需要的计算^[2, 3]。

1 奇异值分解

另一种减少共线性影响的正交分解是奇异值分解, 即 SVD。在 SVD 中, 设计矩阵 X 分解为

$$X = UDV^T \quad (3)$$

其中 U 是由对应 XX^T 的 p 个最大特征值的 p 个正交特征向量组成的 $(n \times p)$ 矩阵; D 是 X 的按大小排列的奇异值 $(p \times p)$ 对角矩阵; V 是 $X^T X$ 的正交特征向量的 $(p \times p)$ 正交矩阵。

此时, 将 (3) 式代入 (2) 式, 得

$$DDV^T \hat{B} = DU^T Y$$

令 D^+ 为由 D 的所有非零元素的倒数构成的对角矩阵, 有

$$D^+ DDV^T \hat{B} = D^+ DU^T Y$$

则正规方程组 (2) 由以下方程组代替

$$DV^T \hat{B} = U^T Y \quad (4)$$

其解为

$$\hat{B} = VD^+ U^T Y \quad (5)$$

其中乘积 $VD^+ U^T$ 是 X 的 Moore-Penrose 广义逆, 可写为

$$X^+ = VD^+ U^T \quad (6)$$

故有回归方程的解

$$\hat{B} = X^+ Y \quad (5')$$

Moore-Penrose 广义逆 X^+ 对任一矩阵 X 是唯一的^[4]。它满足以下条件^[5]

$$X^+ XX^+ = X^+, \quad XX^+ X = X, \quad (X^+ X)^T = X^+ X, \quad (XX^+)^T = XX^+$$

回归方程的解 (5') 不论对欠定方程组还是对确定方程组皆有效。若 $X^T X$ 奇异, 则对角阵 D 的对角元素至少有一个为零, X^+ 用作广义逆, 若 $X^T X$ 非奇异, 则 $\hat{B} = X^+ Y$ 提供了 $\hat{B}$ 中所有系数的估计。在人们感兴趣的共线性方程组中, 方程组是确定的, 但 D 中某些奇异值却接近于零。

如何用奇异值分解获得回归估计, 已有一些讨论^[1, 3, 6~8], 但用得并不普遍, 原因是, 求解正规方程组所花费的时间, 用奇异值分解比用 QR 分解, 要多 5 到 10 倍; 另一个缺点是, 当 X 矩阵增加新行元素时 (相当于增加新的观测值) 时, 缺少更新参数估计的有效方法^[6]。目前, 还没有一个统计软件包是用奇异值分解提供回归程序的, 只是在 Genstat、SAS 和 GAUSS 中提供了奇异值分解的子程序用来作矩阵计算^[5, 9], 当然, 这些矩阵奇异值分解子程序也可以用来计算回归系数。

尽管如此, 许多学者 (包括对奇异值分解求解回归方程提出以上缺陷的学者) 仍然认为, 如果其他方法不适用, 如当遇到不常见的高度病态条件时, 则奇异值分解求解回归方程还是适宜的。而且, 奇异值分解除了解提供参数估计外, 还提供了有关整个方程组的其他信息, 如给出了相应的特征值和特征向量。用奇异值分解提供的这些信息, 就可以对设计矩阵 X 的各列之间的线性相关性作全面的研究^[7]。

2 迭代加细

给定回归方程的解, 可通过所谓的迭代加细过程^[7, 9, 10]来改进其精确度。若在回归方程组中不存在共线性, 则所得到的初始解是高度精确的, 因而没有必要通过迭代加细改进这个解。若存在共线性, 迭代加细就可以产生一系列较精确的解。但迭代加细的收敛速度线性地依赖于 X 的共线性程度^[10], 若共线性特别严重, 则可能迭代加细不收敛, 甚至使初始解更加恶化^[11]。可能得到的不同结论, 如迭代加细收敛但对初始解却无改进、收敛到一个新解或不收敛等等, 可以提供该矩阵 X 共线性程度的一个指标。如果对 X 用 QR 分解求解回归方程, 迭代加细后, 解 $\hat{B}$ 没有明显改进, 则应该想到对 X 用奇异值分解^[7]。

迭代加细改进解向量 $\hat{\mathbf{B}}$, 使之更好地满足用于模型的条件

$$\mathbf{X}^T \mathbf{E} = 0 \quad (7)$$

该条件表明, 剩余向量 $\mathbf{E}$ 正交于观测矩阵 $\mathbf{X}$ 的各列。考虑 $\mathbf{u}$ 和 $\mathbf{z}$

$$\mathbf{u} = \mathbf{Y} - \mathbf{E} - \mathbf{X}\mathbf{B} \quad (8)$$

$$\mathbf{z} = -\mathbf{X}^T \mathbf{E} \quad (9)$$

其中 $\mathbf{E}$ 不是由 (1) 式计算的, 即不用

$$\mathbf{E} = \mathbf{Y} - \mathbf{X}\mathbf{B} \quad (1')$$

显然, 若 $\mathbf{E}$ 由上式计算, 则 $\mathbf{u}$ 将为零, 故用另一种方法得到 $\mathbf{E}$ 是很重要的。若 $\mathbf{B}$ 和 $\mathbf{E}$ 是它们的精确值, 则 $\mathbf{u}$ 和 $\mathbf{z}$ 皆等于零。若 $\mathbf{u}$ 和 $\mathbf{z}$ 皆不为零, 则该事实就可用来迭代加细 $\mathbf{B}$ 和 $\mathbf{E}$ 的估计。

将 $\Delta\mathbf{E}$ 加到 $\mathbf{E}$ 上, $\Delta\mathbf{B}$ 加到 $\mathbf{B}$ 上, 代入 (1') 式和 (7) 式, 分别得

$$\mathbf{Y} - (\mathbf{E} + \Delta\mathbf{E}) - \mathbf{X}(\mathbf{B} + \Delta\mathbf{B}) = (\mathbf{Y} - \mathbf{E} - \mathbf{X}\mathbf{B}) - \Delta\mathbf{E} - \mathbf{X}\Delta\mathbf{B} = \mathbf{u} - \Delta\mathbf{E} - \mathbf{X}\Delta\mathbf{B} = 0$$

即

$$\mathbf{u} = \Delta\mathbf{E} + \mathbf{X}\Delta\mathbf{B} \quad (8')$$

和

$$-\mathbf{X}^T(\mathbf{E} + \Delta\mathbf{E}) = -\mathbf{X}^T \mathbf{E} - \mathbf{X}^T \Delta\mathbf{E} = 0$$

即

$$\mathbf{z} = \mathbf{X}^T \Delta\mathbf{E} \quad (9')$$

因而, 给定 $\mathbf{B}$ 和 $\mathbf{E}$ 的估计, 计算 $\mathbf{u}$ 和 $\mathbf{z}$, 对 $\Delta\mathbf{E}$ 和 $\Delta\mathbf{B}$ 求解, 则 $(\hat{\mathbf{B}} + \Delta\mathbf{B})$ 应是 $\mathbf{B}$ 的更精确的解, 而 $(\hat{\mathbf{E}} + \Delta\mathbf{E})$ 应是 $\mathbf{E}$ 的更精确的解。

迭代加细可用在回归方程的解上, 而不论初始解是如何得到的。本文将迭代加细用在通过使用奇异值分解得到的回归方程的解上。

3 迭代加细 SVD 回归解

对矩阵 $\mathbf{X}$ 应用奇异值分解将产生三个矩阵 $\mathbf{U}$ 、 $\mathbf{D}$ 和 $\mathbf{V}$, 有

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \quad \mathbf{X}^T = \mathbf{V}\mathbf{D}\mathbf{U}^T \quad (3')$$

$\mathbf{U}$ 的各列正交, $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, $\mathbf{V}$ 亦为正交矩阵, $\mathbf{V}^T \mathbf{V} = \mathbf{I}$, $\mathbf{V}\mathbf{V}^T = \mathbf{I}$ 。还有 $\mathbf{X}^T \mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{D}\mathbf{V}^T$ 和 $(\mathbf{X}^T \mathbf{X})^+ = \mathbf{V}\mathbf{D}^+ \mathbf{D}^+ \mathbf{V}^T$ 以及 $\mathbf{X}$ 的 Moore-Penrose 广义逆 $\mathbf{X}^+ = \mathbf{V}\mathbf{D}^+ \mathbf{U}^T$ 。

另一个有用的结论是

$$(\mathbf{X}^T \mathbf{X})^+ \mathbf{X}^T = \mathbf{V}\mathbf{D}^+ \mathbf{D}^+ \mathbf{V}^T \mathbf{V}\mathbf{D}\mathbf{U}^T = \mathbf{V}\mathbf{D}^+ \mathbf{U}^T = \mathbf{X}^+ \quad (10)$$

用于迭代加细的方程 (8') 可写为

$$\mathbf{X}\Delta\mathbf{B} = \mathbf{u} - \Delta\mathbf{E}, \quad \mathbf{X}^T \mathbf{X}\Delta\mathbf{B} = \mathbf{X}^T \mathbf{u} - \mathbf{X}^T \Delta\mathbf{E}$$

由 (9') 有

$$\mathbf{X}^T \mathbf{X}\Delta\mathbf{B} = \mathbf{X}^T \mathbf{u} - \mathbf{z}, \quad \Delta\mathbf{B} = (\mathbf{X}^T \mathbf{X})^+ (\mathbf{X}^T \mathbf{u} - \mathbf{z})$$

利用奇异值分解的 (10) 式, 上式可重写为

$$\Delta\mathbf{B} = \mathbf{X}^+ \mathbf{u} - (\mathbf{X}^T \mathbf{X})^+ \mathbf{z} \quad (11)$$

而

$$\Delta\mathbf{E} = \mathbf{u} - \mathbf{X}\Delta\mathbf{B} \quad (12)$$

用作迭代加细的方程可以保留为 (11) 式和 (12) 式的形式, 也可用 $\mathbf{u}$ 和 $\mathbf{z}$ 的 (8) 式和 (9) 式代入, 再由 (10) 式和 (5') 式而有

$$\Delta\mathbf{B} = \mathbf{X}^+ \mathbf{u} - (\mathbf{X}^T \mathbf{X})^+ \mathbf{z} = \mathbf{X}^+ (\mathbf{Y} - \mathbf{E} - \mathbf{X}\mathbf{B}) + (\mathbf{X}^T \mathbf{X})^+ \mathbf{X}^T \mathbf{E}$$

$$= X^+ Y - X^+ E - X^+ XB + X^+ E = \hat{B} - X^+ XB \quad (13)$$

其中 $\hat{B}$ 是 B 的标准奇异值分解估计, 即(5)式。它并不直接涉及到 E 的估计, 尽管(13)式可写为 $\Delta B = X^+ (Y - XB)$, 说明 ΔB 是 X^+ 乘以误差的标准估计。

B 的估计可以用下式来迭代加细

$$B_{k+1} = B_k + \Delta B = B_k + \hat{B} - X^+ XB_k$$

为使在每一次迭代时的计算减少到最小程度, 上式可写为最终的用奇异值分解求解回归方程的迭代加细公式

$$B_{k+1} = \hat{B} - (X^+ X - I) B_k \quad (14)$$

其中 $\hat{B}$ 和 $(X^+ X - I)$ 只需计算一次。然而事实上, 通常需要以某种形式得到 u 和 z , 因为它们给出算法是否收敛的指标, 因而能够确定停止迭代的条件。

对 E 的估计, 无需用 ΔE 迭代加细, 而可以用 B 的迭代加细估计 B_k 由 $E_k = Y - XB_k$ 来计算 E 的估计。可以看出, 用这种算法迭代加细的代价是每迭代一次只要作 p^2 次操作(p 为自变量个数。若包括检测迭代停止条件, 则为 $2p^2$ 次操作)。这比直接用奇异值分解计算的代价(每迭代一次约要作 $2np^2 + 4p^3$ 次操作)要少的多, 尤其当观测次数 n 比自变量个数 p 大很多时^[1]。

4 比较和讨论

SAS^[5]中的 *REG* 过程和 *ORTHOREG* 过程都是用来求解回归方程的, *REG* 用 *SWEEP* 法, 而 *ORTHOREG* 用 *QR* 分解; *GAUSS*^[5,9]中的 *OLS* 过程用来计算最小二乘回归, *OLSQR* 过程用 *QR* 分解计算回归系数, *SVD* 过程计算矩阵的奇异值, *SVD1* 过程计算矩阵 X 的奇异值分解 $X = U D V^T$ 。

奇异值分解估计和迭代加细估计都比由 SAS 中的 *REG* 过程产生的有关估计更为精确, 而在大多情况下, SAS 中用 *QR* 分解的 *ORTHOREG* 过程产生比用加细奇异值分解过程更为精确的估计, 但在严重共线性的系统中, SAS 中的 *ORTHOREG* 过程并不能得到所需要的正确结果。

迭代加细可以改进由奇异值分解得到的回归估计的精度。用奇异值分解可以获得所有的奇异值, 它们提供了有关共线性出现类型的更多的信息, 而且此时再用迭代加细就又可以得到共线性的诊断度量。

参 考 文 献

- 1 Belsley D A, Kuh E, Welsch R E. Regression diagnostics. New York: Wiley, 1980
- 2 Smith D M. A comparison of methods for forming the QR decomposition for use in regression. Comput Statist Data Anal, 1992, 14: 229~248
- 3 Seber G A F. Linear regression analysis. New York: Wiley, 1977
- 4 吕纯濂. 广义逆矩阵的背景和计算. 南京气象学院学报. 1981, 1(1): 112~120
- 5 吕纯濂, 陈舜华. 矩阵语言与多元分析. 北京: 气象出版社, 1994. 14~177
- 6 Albert A. Regression and the Moore-Penrose Pseudoinverse. New York: Academic Press, 1972
- 7 Belsley D A. Conditioning diagnostics. New York: Wiley, 1991
- 8 Kahaner D, Moler C, Nash S. Numerical methods and software, Englewood Cliffs, NJ: Prentice-Hall, 1989
- 9 GAUSS Software. GAUSS TM, Version 2.0 System and Graphics Manual, Washington, USA: Aptech Systems Inc, 1988
- 10 Bjorck A. Comment on the iterative refinement of least squares solutions. J Amer Statist Assoc, 1978, 73: 161~166

11 Smith D M. Iterative refinement of parameter estimates and residuals from a regression model fitted using QR decomposition methods. Comput Statist Data Anal, 1995, 19(3): 327~349

12 吕纯濂, 陈舜华. 带有变量中误差的 Probit 回归模型. 系统科学与数学, 1993, 13(4): 338~344

13 吕纯濂, 陈舜华. 带有变量中误差的 Logit 模型及其计算. 数值计算与计算机应用, 1994, 15(2): 130~137

14 Lu Chunlian, Chen Shunhua. The Logit Regression Model with Errors in Variables and its Calculation. Chinese Journal of Numerical Mathematics and Applications, 1994, 16(4): 29~38

15 SAS/STAT. User's Guide, Version 6.4, Cary, NC, USA: SAS Institute Inc, 1990

ITERATIVE REFINEMENT OF SVD SOLUTION TO REGRESSION EQUATIONS

L Chunlian

(Department of Basic Course, NIM, Nanjing 210044)

Chen Shunhua

(Department of Atmospheric Physics, NIM, Nanjing 210044)

Abstract In solving the regression equations, collinearity in the design matrix can result in inaccurate parameter estimates. The use of orthogonal matrix transformations such as the singular-value decomposition (SVD) can reduce the effect of collinearity. Also, estimates of the regression coefficients can sometimes be improved through iterative refinement. The application of iterative refinement to the SVD solution of the regression equations is described. Results show that iterative refinement using the SVD can improve regression coefficient estimates in the cases where the design matrix is highly collinear.

Keywords regression equations, collinearity, orthogonal decomposition, iterative refinement