1990年1月 重庆大学学报 Vol.13, No.1
第13卷第1期 JOURNAL OF CHONGQING UNIVERSITY Jan.1990

岭回归估计$\hat{\beta}(k)$的一个特性及其应用

A PROPERTY OF RIDGE REGRESSION ESTIMAPER $\hat{\beta}_{(k)}$
AND ITS' APPLICATION

何 良 材

He Liangcai

（应用数学系）

摘 要 本文就岭回归估计$\hat{\beta}_{(k)}$在$0<K\leqslant\sigma^2/\max r_i^2$条件下，对标准$E\|\hat{\beta}_{(k)}-\beta\|^2$优于$LS$估计问题给出了证明，并结合实例确定较优的岭回归$K$值。

主题词 岭回归估计；最小二乘估计；特征根；回归离差／标准；均方误差
中国图书资料分类法分类号 O 212.1

ABSTRACT in this paper we prove that ridge regression estimater $\hat{\beta}_{(K)}$ under the condition of $0<K\leqslant\sigma^2/\max r_i^2$ is more superier to the LS estimater of standard $E\|\hat{\beta}_{(K)}-\beta\|^2$. We also give examples for showing the superier ridge regression K-Value.

SUBJECT WORDS ride regression estimater; least square estimater; characteristic root; regression deviation / standard; mean square errors

一、引 言

对回归系数β的最小二乘估计（简记为LS估计）已有相当深入的研究，并且也发现LS估计具有若干不够理想的地方，如线性模型

$$Y=X\beta+e$$

中的回归系数β用LS法来估计，得估计量为

$$\hat{\beta}=(X'X)^{-1}X'Y \tag{1}$$

由（1）所确定的$\hat{\beta}$，显然地依赖于相关矩阵$X'X$之逆。如果$X'X$退化，则$\hat{\beta}$不确定，亦即β

收文日期 1988年9月12日

不可估。如果矩阵$X'X$的行列式值十分地小，则$\hat{\beta}$也不十分可靠。鉴于LS法在某些情况下的这些弱点和不足，近十年来一些统计工作者在为提出更好的估计方法作了不少工作。A.E.Horel在六十年代提出一个新的概念——岭回归，又在1970年他同 Kennard 较为系统地介绍了这个方法，这就是所谓的岭回归分析法。

定义　设$K\geqslant 0$，$e\sim N(0, \sigma^2 I)$称

$$\hat{\beta}_{(K)}=(\mathbf{X}'\mathbf{X}+K\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y} \qquad (2)$$

为模型$\mathbf{Y}=\mathbf{X}\beta+\mathbf{e}$中回归系数的岭回归估计，简称岭估计。

其中

$$\mathbf{X}=\begin{pmatrix} x_{11} & x_{12}\cdots x_{1m} \\ x_{21} & x_{22}\cdots x_{2m} \\ \cdots\cdots & \cdots\cdots \\ x_{n1} & x_{n2}\cdots x_{nm} \end{pmatrix}_{n\times m}, \quad \mathbf{Y}=\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}_{n\times 1}$$

$\mathbf{I}$为m阶单位矩阵

显然当$K=0$时，则为通常所说的最小二乘估计，用如（2）式$\hat{\beta}_{(K)}$来估计回归系数β以及一套与之有关的处理方法，称为岭回归分析。在岭回归分析中K值的选择是一个关键的问题，本文就岭回归估计$\hat{\beta}_{(K)}$在$0<K\leqslant\sigma^2/\max r_i^2$条件下，对于标准$E\|\hat{\beta}_{(K)}-\beta\|^2$优于$LS$估计的问题作了一些研究，给出了证明；并结合实例选择较优的K值。

二、岭估计$\hat{\beta}_{(K)}$的一个特性

定理　设$0<K\leqslant\sigma^2/\max r_i^2$成立，且$\mathbf{e}\sim N(0, \sigma^2\mathbf{I})$，则由$\hat{\beta}_{(K)}=(\mathbf{S}+K\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$所定义的岭估计对标准$E\|\hat{\beta}_{(K)}-\beta\|^2$而言优于$LS$估计。其中

$$(r_1\quad r_2\cdots r_m)'=\mathbf{r}=\mathbf{P}\beta,\ \mathbf{S}=\mathbf{X}'\mathbf{X}=\mathbf{P}'\Lambda\mathbf{P}$$

$\mathbf{P}$为正交矩阵，Λ为对角矩阵，其主对角线上元素为$\mathbf{S}$的正值特征根，记为$\lambda_i(i=1, 2, \cdots, m)$，$\mathbf{I}$为$m$阶单位矩阵

证明：由定义

$$\hat{\beta}_{(K)}=(\mathbf{S}+K\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$$

有
$$\begin{aligned}\hat{\beta}_{(K)}-\beta&=(\mathbf{S}+K\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}-\beta\\&=(\mathbf{S}+K\mathbf{I})^{-1}\mathbf{X}'(\mathbf{X}\beta+\mathbf{e})-\beta\\&=(\mathbf{S}+K\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X}+K\mathbf{I}-K\mathbf{I})\beta+(\mathbf{S}+K\mathbf{I})^{-1}\mathbf{X}'\mathbf{e}-\beta\\&=-(\mathbf{S}+K\mathbf{I})^{-1}K\beta+(\mathbf{S}+K\mathbf{I})^{-1}\mathbf{X}'\mathbf{e}\\&=[(\mathbf{X}'\mathbf{e}-K\beta)'(\mathbf{S}+K\mathbf{I})^{-1}]'\end{aligned}$$

$$\therefore\quad E(\hat{\beta}_{(K)}-\beta)=-K(\mathbf{S}+K\mathbf{I})^{-1}\beta \qquad (3)$$

又
$$\begin{aligned}\|\hat{\beta}_{(K)}-\beta\|^2&=(\hat{\beta}_{(K)}-\beta)'(\hat{\beta}_{(K)}-\beta)\\&=(\mathbf{X}'\mathbf{e}-K\beta)'(\mathbf{S}+K\mathbf{I})^{-1}(\mathbf{S}+K\mathbf{I})^{-1}(\mathbf{X}'\mathbf{e}-K\beta)\\&=\mathbf{e}'\mathbf{X}(\mathbf{S}+K\mathbf{I})^{-2}\mathbf{X}'\mathbf{e}+K^2\beta'(\mathbf{S}+K\mathbf{I})^{-2}\beta\end{aligned}$$

$$-K\beta'(\mathbf{S}+K\mathbf{I})^{-2}\mathbf{X}'\mathbf{e}-K\mathbf{e}'\mathbf{X}(\mathbf{S}+K\mathbf{I})^{-2}\beta$$

令$L(\widehat{\beta}_{(K)})=E\|\widehat{\beta}_{(K)}-\beta\|^2$

$$=E[\mathbf{e}'\mathbf{X}(\mathbf{S}+K\mathbf{I})^{-2}\mathbf{X}'\mathbf{e}+K^2\beta'(\mathbf{S}+K\mathbf{I})^{-2}\beta]$$

$$=K^2E[\beta'(\mathbf{S}+K\mathbf{I})^{-2}\beta]+E[\mathbf{e}'\mathbf{X}(\mathbf{S}+K\mathbf{I})^{-2}\mathbf{X}\mathbf{e}] \qquad (4)$$

对（4）式第一项

$$\because\quad \mathbf{S}=\mathbf{X}'\mathbf{X}=\mathbf{P}'\Lambda\mathbf{P}=\mathbf{P}'\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & x_m\end{pmatrix}\mathbf{P},\ \mathbf{P}\text{正交}$$

$$\text{有}(\mathbf{S}+K\mathbf{I})^{-1}=\mathbf{P}'\begin{pmatrix}\lambda_1+K & & 0\\ & \ddots & \\ 0 & & \lambda_m+K\end{pmatrix}^{-1}\mathbf{P}$$

$$=\mathbf{P}'\begin{pmatrix}\frac{1}{\lambda_1+K} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\lambda_m+K}\end{pmatrix}\mathbf{P}$$

$$(\mathbf{S}+K\mathbf{I})^{-2}=\mathbf{P}'\begin{pmatrix}\left(\frac{1}{\lambda_1+K}\right)^2 & & 0\\ & \ddots & \\ 0 & & \left(\frac{1}{\lambda_m+K}\right)^2\end{pmatrix}\mathbf{P}$$

$$\beta'(\mathbf{S}+K\mathbf{I})^{-2}\beta=\beta'\mathbf{P}'\begin{pmatrix}\left(\frac{1}{\lambda_1+K}\right)^2 & & 0\\ & \ddots & \\ 0 & & \left(\frac{1}{\lambda_m+K}\right)^2\end{pmatrix}\mathbf{P}\beta$$

$$=\mathbf{r}'\begin{pmatrix}\left(\frac{1}{\lambda_1+K}\right)^2 & & 0\\ & \ddots & \\ 0 & & \left(\frac{1}{\lambda_m+K}\right)^2\end{pmatrix}\mathbf{r}$$

$$=\sum_{i=1}^{m}\frac{r_i^2}{(\lambda_i+K)^2}$$

$$\therefore\quad K^2E[\beta'(\mathbf{S}+K\mathbf{I})^{-2}\beta]=K^2\beta'(\mathbf{S}+K\mathbf{I})^{-2}\beta=K^2\sum_{i=1}^{m}\left(\frac{r_i}{\lambda_i+K}\right)^2$$

对（4）式第二项

$$
\begin{aligned}
&E[\mathbf{e}'X(\mathbf{S}+K\mathbf{I})^{-2}X'\mathbf{e}]\\
&=t_r[X(\mathbf{S}+K\mathbf{I})^{-2}X']\sigma^2\\
&=t_r[(\mathbf{S}+K\mathbf{I})^{-2}X'X]\sigma^2\\
&=t_r[(\mathbf{S}+K\mathbf{I})^{-2}(\mathbf{S}+K\mathbf{I}-K\mathbf{I})]\sigma^2\\
&=[t_r(\mathbf{S}+K\mathbf{I})^{-1}-Kt_r(\mathbf{S}+K\mathbf{I})^{-2}]\sigma^2\\
&=\left[\sum_{i=1}^{m}\frac{1}{\lambda_i+K}-K\sum_{i=1}^{m}\left(\frac{1}{\lambda_i+K}\right)^2\right]\sigma^2\\
&=\sigma^2\sum_{i=1}^{m}\frac{\lambda_i}{(\lambda_i+K)^2}
\end{aligned}
$$

$$\therefore\quad E[\mathbf{e}'X(\mathbf{S}+K\mathbf{I})^{-2}X'\mathbf{e}]=\sum_{i=1}^{m}\frac{\lambda_i\sigma^2}{(\lambda_i+K)^2}$$

其中$t_r[X(\mathbf{S}+K\mathbf{I})^{-2}X']$表示矩阵$X(\mathbf{S}+K\mathbf{I})^{-2}X'$的迹，即$X(\mathbf{S}+K\mathbf{I})^{-2}X'$全部特征根之和。

又当$0<K\leqslant\sigma^2/\max r_i^2$，有$\max r_i^2\leqslant\frac{\sigma^2}{K}$

$$\therefore\quad r_i^2\leqslant\frac{\sigma^2}{K},\ i=1,\ 2,\ \cdots,\ m$$

$$\text{故}\quad L(\hat{\beta}_{(K)})=K^2\sum_{i=1}^{m}\left(\frac{r_i}{\lambda_i+K}\right)^2+\sum_{i=1}^{m}\frac{\lambda_i\sigma^2}{(\lambda_i+K)^2}\tag{5}$$

$$
\begin{aligned}
&=\sum_{i=1}^{m}\frac{K^2r_i^2+\lambda_i\sigma^2}{(\lambda_i+K)^2}\leqslant\sum_{i=1}^{m}\frac{K^2\sigma^2/K+\lambda_i\sigma^2}{(\lambda_i+K)^2}\\
&=\sum_{i=1}^{m}\frac{K+\lambda_i}{(\lambda_i+K)^2}\sigma^2=\sum_{i=1}^{m}\frac{1}{\lambda_i+K}\sigma^2\\
&<\sigma^2\sum_{i=1}^{m}\frac{1}{\lambda_i}=L(\hat{\beta})
\end{aligned}
$$

注：由（5）式，当$K=0$时，$L(\hat{\beta}_{(0)})=\sum_{i=1}^{m}\frac{\lambda_i\sigma^2}{\lambda_i^2}=\sigma^2\sum_{i=1}^{m}\frac{1}{\lambda_i}$

即是$L(\hat{\beta})=\sigma^2\sum_{i=1}^{m}\frac{1}{\lambda_i}$

即得到$L(\hat{\beta}_{(K)})<L(\hat{\beta})$

〔证毕〕

三、实　例

生产某种产品，从原料到成品需要 6 道工序，各道工序的精度均会不同程度地影响成品的精度。现有原始数据见表 1 。分析各工序精度（质量）对成品的影响，怎样才能有效地提高成品精度。

表 1　原始数据表

绝对偏差 工序 / 序号	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ	成品
1	−0.6414	−0.2070	0.0895	−0.0429	−0.0241	−0.0140	−0.6480
2	0.6238	−0.0587	−0.0955	0.0202	−0.0121	−0.0197	0.6731
3	0.5465	0.3316	0.2899	0.0293	−0.0082	−0.0039	0.6075
4	−0.5523	0.3966	−0.1210	0.0466	−0.0447	0.0024	−0.4759
5	−0.6117	−0.2779	0.0466	0.2156	0.0162	−0.0022	−0.6216
6	−0.5742	0.3839	−0.0347	−0.0107	0.1371	−0.0538	−0 5070
7	0.4050	−0.3111	−0.2196	0.0777	0 2640	0.1112	0.3585
8	−0.3105	0.3727	0.1084	−0.2166	−0.2354	−0.2702	−0.2704
9	−0.2741	0.1632	0.5751	−0.0948	−0.2574	−0.0770	0.5380
10	0.1079	−0.2454	−0.1033	0.4967	0.0911	0.3373	0.5279
11	0.3198	−0.2517	−0.2853	0.0984	0.5408	0.0937	0 5325
12	0 1250	−0.2862	−0.0691	0.4031	0.0907	0.5342	0.5819

解：

1．用Jacobi法求相关矩阵$X'X$的特征值及相应的标准正交特征向量，见表 2

表 2　特征值及特征向量表

序　号	1	2	3	4	5	6
特征值	4.345	1.070	0.403	0.142	0.096	0.043
特	−0.435	−0.274	0.403	−0.427	−0.355	−0.485
征	0.457	−0.080	−0.430	0.202	−0.178	−0.646
向	0.371	0.453	0.732	0.292	−0.119	−0.125
量	−0.375	0.542	−0.305	0.157	−0.658	0.082
P_i	−0.415	−0.380	0.118	0.809	0.074	−0.072
	−0.390	0.525	−0.087	−0.040	0.626	−0.376

由此可把实对称阵$X'X$对角化：

$$P(X'X)P'=\Lambda \quad \text{即} X'X=P'\Lambda P$$

其中$P=(P_1P_2\cdots P_6)'$为正交矩阵，

$$\Lambda=diag(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_6)=\begin{pmatrix}\lambda_1 & & & 0\\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_6\end{pmatrix}$$

$\mathbf{P}_i$为属于特征值λ_i的规范特征向量

2. 求LS估计

LS估计式$\hat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}=(\mathbf{P}'\Lambda\mathbf{P})^{-1}\mathbf{X}'\mathbf{X}$

$\qquad\qquad=\mathbf{P}'\Lambda^{-1}\mathbf{P}\mathbf{X}'\mathbf{Y}$

利用已知数据求得LS估计值为

$$\hat{\beta}_1=1.5545\qquad \hat{\beta}_2=0.5705\qquad \hat{\beta}_3=1.2024$$

$$\hat{\beta}_4=-0.8865\qquad \hat{\beta}_5=0.1417\qquad \hat{\beta}_6=3.8480$$

由此　$\hat{\sigma}_2=\mathbf{e}'\mathbf{e}/(n-m)=(\mathbf{Y}-\mathbf{X}\hat{\beta})'\ (\mathbf{Y}-\mathbf{X}\hat{\beta})/(n-m)$

$\qquad=3.4182/(12-6)=0.5697$

3. 计算$\sigma^2/\max\limits_{1<i<m}r_i^2$

由于σ^2和β在实际中不可测，故用$\hat{\sigma}^2$代σ^2，$\hat{\beta}$代β，则有

$\mathbf{r}=\mathbf{P}\hat{\beta}=(-1.1954\quad 1.5594\quad 1.2121\quad -0.3919\quad 2.2068\quad -2.7248)$

令$K_C=\hat{\sigma}^2/\max\limits_{1<i<6}\gamma_i^2=0.5697/(-2.7248)^2=0.0767$

4. 计算岭回归估计

岭估计为$\hat{\beta}_{(K)}=(\mathbf{X}'\mathbf{X}+K\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$，对于$K$在0到$K_C$内取值不同算得相应岭估计值及估计的均方误差和回归离差　见表3

表3　岭估计值及相应参数值

		0	$\frac{1}{8}k_c$	$\frac{2}{8}k_c$	$\frac{3}{8}k_c$	$\frac{4}{8}k_c$	$\frac{5}{8}k_c$	$\frac{6}{8}k_c$	$\frac{7}{8}k_c$	$\frac{8}{8}k_c$
$\hat{\beta}_{(k)}=$	$\hat{\beta}_{(k)1}$	1.5545	1.3791	1.2543	1.1601	1.0859	1.0254	0.9749	0.9318	0.8944
	$\hat{\beta}_{(k)2}$	0.5705	0.3595	0.2177	0.1172	0.0430	-0.0134	-0.0572	-0.0920	-0.1200
	$\hat{\beta}_{(k)3}$	1.2023	1.1272	1.0670	1.0164	0.9724	0.9332	0.8978	0.8654	0.8355
	$\hat{\beta}_{(k)4}$	-0.8865	-0.8208	-0.7680	-0.7238	-0.6859	-0.6525	-0.6229	-0.5962	-0.5719
	$\hat{\beta}_{(k)5}$	0.1417	0.1363	0.1358	0.1379	0.1412	0.1450	0.1490	0.1529	0.1567
	$\hat{\beta}_{(k)6}$	3.8480	3.1813	2.7148	2.3699	2.1045	1.8936	1.7221	1.5798	1.4599
$L(\hat{\beta}_{(k)})$		25.6732	19.5718	15.6277	12.8768	10.8546	9.3099	8.0951	7.1175	6.3162
e_k^2		3.4182	2.0316	1.3320	0.9564	0.7482	0.6332	0.5730	0.5454	0.5388

其中$L(\beta_{(K)})=\sum_{i=1}^{m}\frac{K^2\gamma_i^2+\lambda_i\sigma^2}{(\lambda_i+K)^2}$，$e_K^2=(Y-X\hat{\beta}_{(K)})'(Y-X\hat{\beta}_{(K)})$由计算结果可知：对于$0<K\leqslant K_C=\sigma^2/\max\gamma_i^2$，相应于$K$的岭回归估计对于估计均方误差$L(\hat{\beta})$而言优于LS估计。

设　$LE_{(K)}=\alpha\cdot e_K^2/e_0^2+(1-\alpha)L(\hat{\beta}_{(K)})/L(\hat{\beta}_{(0)})$

权数α满足$0\leqslant\alpha\leqslant1$，根据实际确定$\alpha$，若侧重于回归离差，则$\alpha$取大于0.5；反之，$\alpha$取小于0.5。当然$\alpha=0$或1作为极端情形也是允许的。

再根据$\min_K LE(K)$来确定应选取哪个值。

对此例，由于要求刻画各道工序精度对成品精度的影响。所以β的估计$\hat{\beta}$的均方误差特别重要，因此取$\alpha=0$的极端情形。

即　$LE(K)=L(\hat{\beta}_{(K)})/L(\hat{\beta}_{(0)})$。

则　$\min_{K=\frac{i}{8}K_C,i=0,1,2\cdots,8} LE(K)=LE\left(\frac{8}{8}K_C\right)=6.3162/25.6732=0.2460$

所以，最后接受$K=\frac{8}{8}K_C=K_C$相应的岭估计值$\hat{\beta}(K_C)$为$\hat{\beta}$即：

$$\hat{\beta}=(0.8944\quad -0.1200\quad 0.8355\quad -0.5719\quad 0.1587\quad 1.4599)'$$
$$Y=0.8944x_1-0.1200x_2+0.8355x_3-0.5719x_4+0.1587x_5+1.4599x_6$$

当然也可不取$\alpha=0$而取α为0.8，0.9…等大于0.5的数，不过对此例特别，由于e_K^2，$L(\beta_{(K)})$均为单调减少，因此取为任何0，1间的数，都不会影响为最优选择这一结果。

鉴此，对成品影响最大者是第6道工序，所以我们应着重提高第6道工序的精度，即把好最后一道工序的质量关，其次是第1，第3道工序，这样便能有效地提高成品的质量。

作者感谢重庆大学应用数学系段虞荣教授与何中市同志的帮助。

参 考 文 献

〔1〕 罗积玉，邢瑛，经济统计分析方法及预测，北京，清华大学出版社，1987

〔2〕 陈希孺，王松桂，近代回归分析——原理方法及应用，北京，科学出版社，1974

〔3〕 陈希孺，陈桂景，线性模型参数的估计理论，北京，科学出版社，1983

〔4〕 王启应，回归系数岭估计的相合性，数理统计与应用概率，1988，(1)，

〔5〕 A.E.Horel, R.W.Kennard, *Ridge regression biased estimation for non-orthogonal problems*, Technometrics, 1970, (12), 55—68