

線形回帰における馬蹄型事前分布を用いた ベイズ予測のリスクの不偏推定量

数理情報学専攻 48216229 福田 健一郎
指導教員 駒木 文保 教授

1 はじめに

線形回帰モデルは、目的変数 $y \in \mathbb{R}^n$, 説明変数 $x_i \in \mathbb{R}^n$ ($i = 1, \dots, p$) を並べた行列である計画行列 $X \in \mathbb{R}^{n \times p}$, 未知パラメータ $\beta \in \mathbb{R}^p$, および誤差項 $\varepsilon \in \mathbb{R}^n$ を用いて次式で表される:

$$y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n).$$

パラメータ β をスパースな量と仮定し推定する手法は多く考えられており、ベイズ統計では縮小事前分布を用いる方法がある。その中でも、馬蹄型事前分布 [2] は優れた推定が可能な事前分布として知られている [2, 4]。また、予測のリスクの推定量として、[5] は未知である真のパラメータを含まない計算可能な量を導き、これはリスクの不偏推定量 (Stein's Unbiased Risk Estimate, SURE) として知られている。

本論文では、SURE を一般のノルムにおけるリスクの不偏推定量へと拡張するとともに、線形回帰において馬蹄型事前分布を用いたベイズ予測について、計画行列が別の計画行列に変わるという問題設定における SURE を導出し、それを用いたリスクの評価について提案する。

2 既存研究

[5] は、未知パラメータを含まない、観測値 y に基づいて計算可能なリスクの不偏推定量を導いた。

定理 1 (Stein's unbiased risk estimate (SURE)). $y \sim N(\mu, \sigma^2 I_n)$ を n 次元の観測値、弱微分可能な関数 $\hat{\mu} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ を未知パラメータ μ の推定量とする。このとき、リスク $E[\|\mu - \hat{\mu}\|^2]$ の不偏推定量 (SURE) は、

$$\text{SURE} = -n\sigma^2 + \|y - \hat{\mu}\|^2 + 2\sigma^2 \sum_{i=1}^n \frac{\partial \hat{\mu}_i}{\partial y_i}.$$

馬蹄型事前分布を用いたときの SURE についての研究 [1] について概説する。計画行列 X について $X = UDW^\top$ と特異値分解し、 $Z = UD, \alpha = W^\top \beta$ とおくと、もとの線形回帰モデルは次式で表される:

$$y = Z\alpha + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n).$$

α の最小二乗推定量を $\hat{\alpha}$ とすると、リッジ回帰や主成分回帰といった大域的な縮小推定手法の SURE は、 $\text{SURE} = \sum_{i=1}^n \text{SURE}_i$ として次式で与えられる:

$$\text{SURE}_i = \frac{\hat{\alpha}_i^2 d_i^2}{(1 + \tau^2 \lambda_i^2 d_i^2)^2} + 2\sigma^2 \frac{\tau^2 \lambda_i^2 d_i^2}{(1 + \tau^2 \lambda_i^2 d_i^2)}.$$

大域的な縮小推定手法において λ_i は定数であり、単一のパラメータ τ によって各 SURE_i の値を適切に小さくすることは難しい。そこで、 τ だけでなく局所的な縮小パラメータ λ_i を持つ馬蹄型事前分布の適用を考える。

定理 2 (馬蹄型事前分布を用いたときの SURE [1]). $\text{SURE} = \sum_{i=1}^n \text{SURE}_i$ として、

$$\text{SURE}_i = 2\sigma^2 - \sigma^4 d_i^{-2} \left\{ \frac{m'(\hat{\alpha}_i)}{m(\hat{\alpha}_i)} \right\}^2 + 2\sigma^4 d_i^{-2} \frac{m''(\hat{\alpha}_i)}{m(\hat{\alpha}_i)}.$$

また、 $m'(\hat{\alpha}_i)/m(\hat{\alpha}_i) = -\hat{\alpha}_i d_i^2 E[Z_i]/\sigma^2$, および $m''(\hat{\alpha}_i)/m(\hat{\alpha}_i) = -d_i^2 E[Z_i^2]/\sigma^2 + \hat{\alpha}_i^2 d_i^4 E[Z_i^2]/\sigma^4$ が成り立つ。ただし、 Z_i は compound confluent hypergeometric 分布 [3] $\text{CCH}(p = 1, q = 1/2, r = 1, s = \hat{\alpha}_i^2 d_i^2/2\sigma^2, \nu = 1, \theta = 1/\tau^2 d_i^2)$ に従う確率変数である。

3 提案手法

計画行列が変化する線形回帰における SURE を導く。

定理 3 (SURE の一般のノルムへの拡張). 正定値対称行列 B に対して、ノルム $\|\hat{\mu} - \mu\|_B := (\hat{\mu} - \mu)^\top B(\hat{\mu} - \mu)$ とする。 $y \sim N(\mu, \sigma^2 I_n)$ を n 次元の観測値、弱微分可能な関数 $\hat{\mu} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ を未知パラメータ μ の推定量とする。また、正定値対称行列 B は直交行列 P を用いて、 $B = P^\top D P$ と対角化されるとする。このとき、リスク $E[\|\hat{\mu} - \mu\|_B]$ の不偏推定量は、

$$\begin{aligned} \text{SURE} = & - \sum_{i=1}^n \sigma^2 d_i + \|D^{\frac{1}{2}} P y - D^{\frac{1}{2}} P \hat{\mu}\|^2 \\ & + 2\sigma^2 \sum_{i=1}^n d_i \frac{\partial (D^{\frac{1}{2}} P \hat{\mu})_i}{\partial (D^{\frac{1}{2}} P y)_i}. \end{aligned}$$

定理 4 (計画行列が変化する線形回帰モデルへの適用). $y = Z\hat{\alpha} \sim N(Z\alpha, \sigma^2 I_n)$ を n 次元の観測値、未知パラメータ α の推定量を $\hat{\alpha}$ として $\tilde{y} = Z\hat{\alpha}$ と表記する。

また、行列 $(Z^+)^T \tilde{Z}^T \tilde{Z} Z^+$ は直交行列 P を用いて、 $(Z^+)^T \tilde{Z}^T \tilde{Z} Z^+ = P^T D' P$ と対角化されるとする。このとき、リスク $E[\|\tilde{Z}\tilde{\alpha} - \tilde{Z}\alpha\|^2]$ の不偏推定量は、

$$\begin{aligned} \text{SURE} = & - \sum_{i=1}^n \sigma^2 d'_i + \|D'^{\frac{1}{2}} P y - D'^{\frac{1}{2}} P \tilde{y}\|^2 \\ & + 2\sigma^2 \sum_{i=1}^n d'_i \frac{\partial(D'^{\frac{1}{2}} P \tilde{y})_i}{\partial(D'^{\frac{1}{2}} P y)_i}. \end{aligned}$$

定理 5 (馬蹄型事前分布の適用). 馬蹄型事前分布を用いたとき、リスク $E[\|\tilde{Z}\tilde{\alpha} - \tilde{Z}\alpha\|^2]$ の不偏推定量は、

$$\begin{aligned} \text{SURE} = & \sum_{i=1}^n \sigma^2 d'_i \\ & + \sigma^4 (\nabla_{\hat{\alpha}} \log m(\hat{\alpha}))^T D^{-2} \tilde{Z}^T \tilde{Z} D^{-2} \nabla_{\hat{\alpha}} \log m(\hat{\alpha}) \\ & + 2\sigma^4 \sum_{j=1}^n \frac{\partial^2}{\partial \hat{\alpha}_j^2} \log m(\hat{\alpha}_j) \sum_{i=1}^n (D'^{\frac{1}{2}} P U D^{-1})_{i,j}^2. \end{aligned}$$

また、 $m'(\hat{\alpha}_i)/m(\hat{\alpha}_i) = -\hat{\alpha}_i d_i^2 E[Z_i]/\sigma^2$ 、および $m''(\hat{\alpha}_i)/m(\hat{\alpha}_i) = -d_i^2 E[Z_i^2]/\sigma^2 + \hat{\alpha}_i^2 d_i^4 E[Z_i^2]/\sigma^4$ が成り立つ。ただし、 Z_i は $\text{CCH}(p=1, q=1/2, r=1, s=\hat{\alpha}_i^2 d_i^2/2\sigma^2, \nu=1, \theta=1/\tau^2 d_i^2)$ に従う確率変数である。

以上より、将来の計画行列を含んだリスクの評価が可能となり、またこの SURE を小さくするパラメータを用いることで優れた予測が可能になると考えられる。

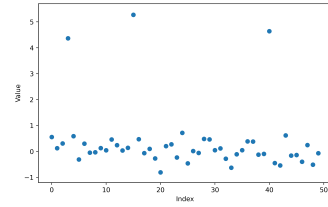
4 数値実験

計画行列が変化するような人工データについて、既存手法および提案手法による SURE をそれぞれ最小化するようなパラメータ τ を用いてパラメータ α の推定、および予測の二乗誤差を計算した。表 1 の結果から、提案手法による予測の性能が優れていることが確認できる。また、図 1 から、提案手法による推定値が真のシグナルをうまく捉えられている様子を読み取れる。

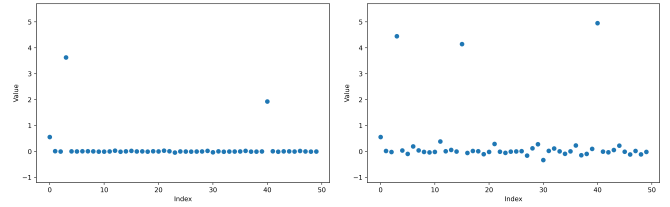
表 1. 既存手法、提案手法それぞれの SURE および予測の二乗誤差 (二乗誤差の標準偏差) の結果の比較。

既存手法		提案手法	
SURE	二乗誤差	SURE	二乗誤差
163.35	11203.03	1311.28	1551.05
	(214.25)		(72.35)

また、実データとして、癌細胞株についての研究データを用いた。癌細胞株に使用した化合物および薬剤の濃度レベルを説明変数とし、癌細胞株におけるヒト ABC



(a) 真の値



(b) 既存手法

(c) 提案手法

図 1. (a) は真のパラメータ α の値、(b), (c) は各手法による α の推定値 (横軸: 要素番号, 縦軸: 値).

トランスポーター遺伝子の発現量を目的変数として予測を行った。表 2 の結果から、いずれの遺伝子についても提案手法による予測が優れていることが確認された。

表 2. 既存手法、提案手法による予測の平均二乗誤差 (平均二乗誤差の標準偏差) の結果の比較の一部。

目的変数	MSE (既存手法)	MSE (提案手法)
A1	3.28	2.16
	(0.72)	(0.30)
A2	1.13	0.74
	(0.18)	(0.17)
A3	25.88	23.64
	(4.32)	(3.16)

参考文献

- [1] A. Bhadra, J. Datta, Y. Li, N. G. Polson, and B. Willard. Prediction risk for the horseshoe regression. *The Journal of Machine Learning Research*, Vol. 20, No. 1, pp. 2882–2920, 2019.
- [2] C. M. Carvalho, N. G. Polson, and J. G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, Vol. 97, No. 2, pp. 465–480, 2010.
- [3] M. B. Gordy. A generalization of generalized beta distributions. In *Finance and Economics Discussion Series*. Division of Research and Statistics, Division of Monetary Affairs, Federal Reserve Board, 1998.
- [4] N. G. Polson and J. G. Scott. Shrink globally, act locally: sparse Bayesian regularization and prediction. *Bayesian Statistics*, Vol. 9, pp. 501–538, 2010.
- [5] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, Vol. 9, No. 6, pp. 1135–1151, 1981.