数理解析研究所講究録
第2078巻 2018年 32-38

レプリカ解析を用いた線形回帰誤差関数の理論評価

新里　隆 *
Takashi Shinzato
玉川大学工学部マネジメントサイエンス学科
Tamagawa University

1　はじめに

研究背景

- 回帰分析は主成分分析や因子分析などと並んで最も基本的な多変量解析手法の 1 つであり，広く実用されている統計手法である．
- 回帰分析には 1 つの被説明変数 y を説明する変数が 1 つである単回帰分析と複数ある重回帰分析に分けることができる．

$$\text{単回帰分析} \qquad y = ax + b \tag{1}$$

$$\text{重回帰分析} \qquad y = \sum_{i=1}^{N} a_i x_i + b \tag{2}$$

- 重回帰分析では多重共線性や過剰適合などの問題を内包していることが知られている [1-3].

先行研究と問題点

- Cun らは偏回帰係数を求める最急降下法の効率 (スピード) を評価するために，計画行列で定義される分散共分散行列 (Wishart 行列) の固有値分布を解析した [4].
- Hoyle らは主成分分析に対してレプリカ解析を用いて上記と同様に固有値分布を解析し，Marčenko-Pastur 分布や Tracy-Widom 分布と比較した [5].

しかしながら重回帰分析の線形近似の精度について十分な議論がなされているとは言い難い．

研究目的

- Hoyle らが用いたレプリカ解析を用いて，重回帰分析の線形近似の精度について，統一的な解析手法を提案する．
  1. はじめに線形回帰モデルを定式化する．
  2. Lagrange 未定乗数法を用いて最適な偏回帰係数や残差平方和を求める．

* shinzato@eng.tamagawa.ac.jp

3. 上記の手順の困難を解決するために，事後確率を用いた MAP 推定問題に再定式化する．
4. レプリカ解析を用いてキュムラント母関数を評価し，残差平方和を解析的に導出する．

- 提案手法を用いて残差平方和とデータ数の関係について議論する．

2 重回帰分析

2.1 モデル設定

被説明変数を $y \in \mathbf{R}$，説明変数を N 次元ベクトル $\vec{x} = (x_1, x_2, \cdots, x_N)^{\mathrm{T}} \in \mathbf{R}^N$ で表す．さらに説明変数 $\vec{x}$ と被説明変数 y の組が p 組あるとする．

$$\mathcal{D}^p = \{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \cdots, (\vec{x}_p, y_p)\} \tag{3}$$

このとき被説明変数 y_μ と説明変数 $\vec{x}_\mu = (x_{1\mu}, x_{2\mu}, \cdots, x_{N\mu})^{\mathrm{T}} \in \mathbf{R}^N$ の間に次の関係が成り立つと仮定する．

$$y_\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_i x_{i\mu} \tag{4}$$

式 (4) は多変数に対する線形回帰 (重回帰分析) を表しているが，一般的に p 組の入出力関係 $\mathcal{D}^p$ を満たす偏回帰係数ベクトル $\vec{w} = (w_1, w_2, \cdots, w_N)^{\mathrm{T}} \in \mathbf{R}^N$ を見つけることは容易ではなく，

$$y_\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_i x_{i\mu} + noise_\mu \tag{5}$$

のようになる．さらに $p > N$ とする．線形回帰式 $y_\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_i x_{i\mu} + noise_\mu$ から，最小 2 乗法より，

$$\mathcal{H}(\vec{w}|\mathcal{D}^p) = \frac{1}{2} \sum_{\mu=1}^{p} \left(y_\mu - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_i x_{i\mu} \right)^2 \tag{6}$$

の残差平方和 $\mathcal{H}(\vec{w}|\mathcal{D}^p)$ を最小にする偏回帰係数 $\vec{w}$ を求める必要がある．ここで中心極限定理に倣い，N 個の確率変数 $x_{i\mu}$(に偏回帰係数 w_i をかけた) の和 $\sum_{i=1}^{N} w_i x_{i\mu}$ に対して，スケーリングの $\frac{1}{\sqrt{N}}$ を掛けた．(注) 理論的な性質を調べる第一歩として，極端に簡単な設定からスタートさせる[*1]．線形関係を満たす偏回帰係数 $\vec{w}$ を求めるために，次の残差平方和の最小化問題を解く必要

[*1] 十分大きな N のとき，平均 0，分散 1 の i.i.d. の N 個の確率変数 $X_1, X_2, \cdots, X_N$ に対して，$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i$ は平均 0，分散 1 の正規分布に従う．

がある.

$$\mathcal{H}(\vec{w}|\mathcal{D}^p) = \frac{1}{2}\sum_{\mu=1}^{p}\left(y_\mu - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} w_i x_{i\mu}\right)^2 \tag{7}$$

ここでは簡単のために次の制約条件を課す.

$$\sum_{i=1}^{N} w_i = N \times \tau \tag{8}$$

ちなみに式 (8) は τ で指定される超平面を表す. また観測値ベクトル $\vec{y} = (y_1, \cdots, y_p)^{\mathrm{T}} \in \mathbf{R}^p$ と計画行列 $X = \left\{\frac{x_{i\mu}}{\sqrt{N}}\right\} \in \mathbf{R}^{N\times p}$, Wishart 行列 $J = XX^{\mathrm{T}} \in \mathbf{R}^{N\times N}$ を用いて,

$$\mathcal{H}(\vec{w}|\mathcal{D}^p) = \frac{1}{2}\left(\vec{y} - X^{\mathrm{T}}\vec{w}\right)^2 = \frac{1}{2}\vec{y}^{\mathrm{T}}\vec{y} - \vec{w}^{\mathrm{T}}X\vec{y} + \frac{1}{2}\vec{w}^{\mathrm{T}}J\vec{w} \tag{9}$$

2.2 Lagrange 未定乗数法

Lagrange 関数 $L(\vec{w}, k)$ は

$$\begin{aligned} L(\vec{w}, k) &= \mathcal{H}(\vec{w}|\mathcal{D}^p) + k(N\tau - \vec{w}^{\mathrm{T}}\vec{e}) \\ &= \frac{1}{2}\vec{y}^{\mathrm{T}}\vec{y} - \vec{w}^{\mathrm{T}}X\vec{y} + \frac{1}{2}\vec{w}^{\mathrm{T}}J\vec{w} + k(N\tau - \vec{w}^{\mathrm{T}}\vec{e}) \end{aligned} \tag{10}$$

と定義できる. ただしここでは定数ベクトル $\vec{e} = (1, 1, \cdots, 1)^{\mathrm{T}} \in \mathbf{R}^N$ と Lagrange 未定乗数 $k \in \mathbf{R}$ を用いた. この最適解は,

$$k = \frac{\tau - \frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}X\vec{y}}{\frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}\vec{e}} \tag{11}$$

$$\vec{w}^* = J^{-1}X\vec{y} + kJ^{-1}\vec{e} \tag{12}$$

となり, 平方残差和 $\min_{\vec{w}\in\mathcal{W}}\mathcal{H}(\vec{w}|\mathcal{D}^p) = \mathcal{H}(\vec{w}^*|\mathcal{D}^p)$ は,

$$\mathcal{H}(\vec{w}^*|\mathcal{D}^p) = \frac{1}{2}\vec{y}^{\mathrm{T}}\vec{y} - \frac{1}{2}\vec{y}^{\mathrm{T}}X^{\mathrm{T}}J^{-1}X\vec{y} + \frac{Nk^2}{2}\frac{\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}{N} \tag{13}$$

となる. ただし $\mathcal{W} \subseteq \mathbf{R}^N$ は偏回帰係数 $\vec{w}$ の実行可能部分空間とする. 得られた結果 $\mathcal{H}(\vec{w}^*|\mathcal{D}^p) = \frac{1}{2}\vec{y}^{\mathrm{T}}\vec{y} - \frac{1}{2}\vec{y}^{\mathrm{T}}X^{\mathrm{T}}J^{-1}X\vec{y} + \frac{Nk^2}{2}\frac{\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}{N}$ と $k = \frac{\tau - \frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}X\vec{y}}{\frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}$ から, 1 自由度当たりの残差平方和 $\varepsilon = \frac{1}{N}\mathcal{H}(\vec{w}^*|\mathcal{D}^p)$ は,

$$\begin{aligned} \varepsilon = \frac{1}{N}\mathcal{H}(\vec{w}^*|\mathcal{D}^p) &= \frac{1}{2}\frac{\vec{y}^{\mathrm{T}}\vec{y}}{N} - \frac{1}{2}\frac{\vec{y}^{\mathrm{T}}X^{\mathrm{T}}J^{-1}X\vec{y}}{N} + \frac{k^2}{2}\frac{\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}{N} \\ &= \frac{1}{2}\frac{\vec{y}^{\mathrm{T}}\vec{y}}{N} - \frac{1}{2}\frac{\vec{y}^{\mathrm{T}}X^{\mathrm{T}}J^{-1}X\vec{y}}{N} + \frac{1}{2}\frac{\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}{N}\left(\frac{\tau - \frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}X\vec{y}}{\frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}\right)^2 \end{aligned} \tag{14}$$

となる.

問題点

- Wishart 行列 $J = XX^{\mathrm{T}} \in \mathbf{R}^{N\times N}$ の逆行列を求めるには，$O(N^3)$ の計算量が必要であり，十分大きな N(理論的な性質を議論しやすい) において，実行することは容易ではない．
- 1 自由度当たりの残差平方和 $\varepsilon = \frac{1}{N}\mathcal{H}(\vec{w}^*|\mathcal{D}^p)$ は $\mathcal{D}^p = (X, \vec{y})$ に依存しており，ε の典型値を求めるためには，ε を計画行列 $X \in \mathbf{R}^{N\times p}$ や観測値ベクトル $\vec{y} \in \mathbf{R}^p$ で平均化する必要がある．

1 自由度当たりの平方残差和の典型値 $\varepsilon = E_{X,\vec{y}}\left[\frac{1}{N}\mathcal{H}(\vec{w}^*|\mathcal{D}^p)\right]$ を求めるには，$\frac{\vec{y}^{\mathrm{T}}\vec{y}}{N}$, $\frac{\vec{y}^{\mathrm{T}}X^{\mathrm{T}}J^{-1}X\vec{y}}{N}$, $\frac{\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}{N}$, $\frac{\vec{e}^{\mathrm{T}}J^{-1}X\vec{y}}{N}$，$k = \frac{\tau - \frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}X\vec{y}}{\frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}$ などの典型値を求める必要がある．

$$\frac{1}{N}\mathcal{H}(\vec{w}^*|\mathcal{D}^p) = \frac{1}{2}\frac{\vec{y}^{\mathrm{T}}\vec{y}}{N} - \frac{1}{2}\frac{\vec{y}^{\mathrm{T}}X^{\mathrm{T}}J^{-1}X\vec{y}}{N} + \frac{1}{2}\frac{\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}{N}\left(\frac{\tau - \frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}X\vec{y}}{\frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}\right)^2 \tag{15}$$

問題点

- Wishart 行列 $J = XX^{\mathrm{T}}$ の平均値 $E_X[XX^{\mathrm{T}}]$ の逆行列 $(E_X[XX^{\mathrm{T}}])^{-1}$ と Wishart 行列の逆行列 $(XX^{\mathrm{T}})^{-1}$ の平均値 $E_X[(XX^{\mathrm{T}})^{-1}]$ は一致するか？
- $k = \frac{\tau - \frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}X\vec{y}}{\frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}$ や $\frac{\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}{N}\left(\frac{\tau - \frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}X\vec{y}}{\frac{1}{N}\vec{e}^{\mathrm{T}}J^{-1}\vec{e}}\right)^2$ の典型値を評価することは容易か？

3 Boltzmann 分布

3.1 MAP 推定

MAP 推定は事後確率 $P(\vec{w}|\mathcal{D}^p)$ を最大にする $\vec{w}$ を求めることであり，機械学習や情報理論，最適化問題などに共通する数理構造を用いて解析する手法である [6-8].

MAP 推定

MAP 推定=事後確率 $P(\vec{w}|\mathcal{D}^p)$ を最大にする $\vec{w}$ を求める．

$$\Longrightarrow \qquad \vec{w}^* = \arg\max_{\vec{w}\in\mathcal{W}} P(\vec{w}|\mathcal{D}^p) \tag{16}$$

3.2 事後確率

事後確率 $P(\vec{w}|\mathcal{D}^p)$ は事前確率 $P_0(\vec{w})$ と条件付き確率 (もしくは尤度関数) $e^{-\beta\mathcal{H}(\vec{w}|\mathcal{D}^p)}$ の積で以下のように定義できる．

$$P(\vec{w}|\mathcal{D}^p) = \frac{P_0(\vec{w})e^{-\beta\mathcal{H}(\vec{w}|\mathcal{D}^p)}}{Z(\mathcal{D}^p)} \tag{17}$$

ただし $Z(\mathcal{D}^p)=\int_{-\infty}^{\infty}P_0(\vec{w})e^{-\beta\mathcal{H}(\vec{w}|\mathcal{D}^p)}d\vec{w}$ は分配関数と呼ばれる規格化定数であり，$\beta>0$ は逆温度と呼ぶ．

最適化

$$\vec{w}^*=\arg\max_{\vec{w}\in\mathcal{W}}P(\vec{w}|\mathcal{D}^p)=\arg\min_{\vec{w}\in\mathcal{W}}\mathcal{H}(\vec{w}|\mathcal{D}^p) \tag{18}$$

逆温度極限として，

$$\lim_{\beta\to\infty}P(\vec{w}|\mathcal{D}^p)=\begin{cases}1 & \vec{w}=\vec{w}^*\\ 0 & \vec{w}\neq\vec{w}^*\end{cases} \tag{19}$$

が知られている．

さらに $Z(\mathcal{D}^p)=\int_{-\infty}^{\infty}P_0(\vec{w})e^{-\beta\mathcal{H}(\vec{w}|\mathcal{D}^p)}d\vec{w}$ を用いて示される，

$$E_{\vec{w}}[\mathcal{H}(\vec{w}|\mathcal{D}^p)]=\int_{-\infty}^{\infty}\mathcal{H}(\vec{w}|\mathcal{D}^p)P(\vec{w}|\mathcal{D}^p)d\vec{w}=-\frac{\partial}{\partial\beta}\log Z(\mathcal{D}^p) \tag{20}$$

の恒等式を用いると，

$$\mathcal{H}(\vec{w}^*|\mathcal{D}^p)=\lim_{\beta\to\infty}E_{\vec{w}}[\mathcal{H}(\vec{w}|\mathcal{D}^p)]=\lim_{\beta\to\infty}\left\{-\frac{\partial}{\partial\beta}\log Z(\mathcal{D}^p)\right\} \tag{21}$$

となる．さらに成分数 N の大きい極限で $\log Z(\mathcal{D}^p)$ が自己平均性 $\log Z(\mathcal{D}^p)\simeq E_{X,\vec{y}}[\log Z(\mathcal{D}^p)]$ を満たすため，1 自由度当たりの残差平方和 ε は，

$$\begin{aligned}\varepsilon&=\lim_{N\to\infty}\frac{1}{N}\mathcal{H}(\vec{w}^*|\mathcal{D}^p)=\lim_{N\to\infty}\frac{1}{N}E_{X,\vec{y}}[\mathcal{H}(\vec{w}^*|\mathcal{D}^p)]\\&=\lim_{\beta\to\infty}-\frac{\partial}{\partial\beta}\left\{\lim_{N\to\infty}\frac{1}{N}E_{X,\vec{y}}[\log Z(\mathcal{D}^p)]\right\}\end{aligned} \tag{22}$$

3.3 問題整理

- 1 自由度当たりの残差平方和 ε は，

$$\varepsilon=\lim_{N\to\infty}\frac{1}{N}\mathcal{H}(\vec{w}^*|\mathcal{D}^p)=\lim_{\beta\to\infty}-\frac{\partial}{\partial\beta}\left\{\lim_{N\to\infty}\frac{1}{N}E_{X,\vec{y}}[\log Z(\mathcal{D}^p)]\right\} \tag{23}$$

  で求まる．
- キュムラント母関数

$$\phi=\lim_{N\to\infty}\frac{1}{N}E_{X,\vec{y}}[\log Z(\mathcal{D}^p)] \tag{24}$$

  を用いて，$\varepsilon=-\lim_{\beta\to\infty}\frac{\partial\phi}{\partial\beta}$ で求めることができる．

本発表では被説明変数 y_μ は各々独立に平均 $E[y_\mu]=0$，分散 $V[y_\mu]=t$ の分布に従うとし，説明変数 $x_{i\mu}$ は各々独立に平均 $E[x_{i\mu}]=0$，分散 $V[x_{i\mu}]=v_i$ の分布に従うとする．

4 レプリカ解析

4.1 レプリカ対称解

レプリカ解析を用いて，キュムラント母関数 ϕ は以下のようになる．

$$
\begin{aligned}
\phi &= \lim_{N\to\infty} \frac{1}{N} E_{X,\vec{y}}[\log Z(\mathcal{D}^p)] \\
&= \underset{k,\chi_s,q_s,\tilde{\chi}_s,\tilde{q}_s}{\mathrm{Extr}} \left\{ -\frac{\alpha}{2}\log(1+\beta\chi_s) - \frac{\alpha\beta(q_s+t)}{2(1+\beta\chi_s)} + \frac{1}{2}(\chi_s+q_s)(\tilde{\chi}_s-\tilde{q}_s) \right. \\
&\quad \left. +\frac{1}{2}q_s\tilde{q}_s - k\tau - \frac{1}{2}\langle \log v\rangle - \frac{1}{2}\log\tilde{\chi}_s + \frac{\tilde{q}_s + k^2\langle v^{-1}\rangle}{2\tilde{\chi}_s} \right\}
\end{aligned} \tag{25}
$$

ただし $\alpha = p/N \sim O(1)$ と

$$
\langle f(v)\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f(v_i) \tag{26}
$$

とし，$\underset{m}{\mathrm{Extr}}\, g(m)$ は関数 $g(m)$ の変数 m に対する極値を表す．$\frac{\partial\phi}{\partial k} = \frac{\partial\phi}{\partial\chi_s} = \frac{\partial\phi}{\partial q_s} = \frac{\partial\phi}{\partial\tilde{\chi}_s} = \frac{\partial\phi}{\partial\tilde{q}_s} = 0$ より，

$$
k = \beta(\alpha-1)\frac{\tau}{\langle v^{-1}\rangle} \tag{27}
$$

$$
\chi_s = \frac{1}{\beta(\alpha-1)} \tag{28}
$$

$$
q_s = \frac{t}{\alpha-1} + \frac{\alpha}{\alpha-1}\frac{\tau^2}{\langle v^{-1}\rangle} \tag{29}
$$

$$
\tilde{\chi}_s = \beta(\alpha-1) \tag{30}
$$

$$
\tilde{q}_s = \beta^2(\alpha-1)\left(t + \frac{\tau^2}{\langle v^{-1}\rangle}\right) \tag{31}
$$

となる．これより $\varepsilon = -\frac{\partial\phi}{\partial\beta} = \frac{\alpha\chi_s}{2(1+\beta\chi_s)} + \frac{\alpha q_s}{2(1+\beta\chi_s)^2}$ から，

$$
\varepsilon = \lim_{\beta\to\infty}\left(\frac{\alpha\chi_s}{2(1+\beta\chi_s)} + \frac{\alpha(q_s+t)}{2(1+\beta\chi_s)^2}\right) = \frac{\alpha-1}{2}\left(t + \frac{\tau^2}{\langle v^{-1}\rangle}\right) \tag{32}
$$

となる．

- $E[y_\mu] = 0, V[y_\mu] = t, E[x_{i\mu}] = 0, V[x_{i\mu}] = v_i$ のとき，1 自由度当たりの残差平方和 $\varepsilon = \lim_{N\to\infty}\frac{1}{N}\mathcal{H}(\vec{w}^*|\mathcal{D}^p)$ は，

$$
\varepsilon = \frac{\alpha-1}{2}\left(t + \frac{\tau^2}{\langle v^{-1}\rangle}\right) \tag{33}
$$

となる．

5 まとめと今後の課題

まとめとして，

- レプリカ解析を用いて 1 自由度当たりの残差平方和を推定する方法を議論した．
- Lagrange 未定乗数法を用いて求めた 1 自由度当たりの残差平方和には Wishart 行列 $J = XX^{\mathrm{T}} \in \mathbf{R}^{N\times N}$ の逆行列 J^{-1} を求める必要があるが，逆行列の計算量は $O(N^3)$ であるため，別の解析手法を提案した．
- Boltzmann 分布を用いた最適化問題の解析手法をまとめ，1 自由度当たりの残差平方和はキュムラント母関数 ϕ の β 微分で求められることを説明した．
- レプリカ解析を用いてキュムラント母関数を評価し，1 自由度当たりの残差平方和を解析的に求めることができた．

今後の課題として，

- 決定係数も議論する必要があるだろう．
- マルチファクターモデルにも拡張する必要があるだろう．

Reference

[1] 竹村彰通，現代数理統計学，創文社，(1991).

[2] 鈴木武，山田作太郎，数理統計学，内田老鶴圃，(1996).

[3] 稲垣宣生，数理統計学，裳華房，(1990).

[4] Y. L. Cun, I. Kanter and S. A. Solla, Phys. Rev. Let. Vol. **66**, No. 18, pp. 2396-2399, (1991).

[5] D. C. Hoyle and M. Rattray, Phys. Rev. E, Vol. **69**, No. 2, 026124, (2004).

[6] A. Engel and C. V. Broeck, Statistical Mechanics of Learning, Cambridge Univ. Press, (2001).

[7] H. Nishimori, Statistical Physics of Spin Glasses and Information Processing, Oxford Univ. Press, (2001).

[8] A. K. Hartmann and M. Weigt, Phase Transitions in Combinatorial Optimization Problems, Wiley, (2005).