Wishart 行列のゆらぎモーメントによる回帰モデルの残差分析

大谷 有美　(指導教員：吉田 裕亮)

1　はじめに

カーネル法はカーネル関数を利用し, 観測データを高次元のベクトル空間に写像し, 変換後のデータに線形の手法を用いることで, 非線形の統計解析を実現するものである.

本研究では, カーネル回帰の正則化パラメーター λ に関して, 残差からのリサンプリングにより大量の Wishart 行列を構成し, そのゆらぎモーメントと比較することにより最適な λ の値を選択することを試みる.

2　カーネル回帰とリッジ正則化

2.1　リッジ正則化

線形回帰において残差平方和の最小化のみを評価関数とした場合, 過学習が発生する. そこで回帰係数 $\{\beta_j\}$ の L^2 ノルムを罰則項として加えた

$$\sum_{i=1}^{N}\Big(y_i-\sum_{j=1}^{p}\beta_j x_{ij}\Big)^2+\lambda\sum_{j=1}^{p}\beta_j^2 \quad (\lambda\geq 0)$$

を最小化することにより, 過学習を抑える手法がリッジ (Ridge) 正則化である. ここで $\lambda>0$ は罰則項を制御するパラメータである.

2.2　カーネル関数

$\phi_1,...,\phi_l$ という非線形関数で特徴抽出されたベクトルを特徴ベクトルといい,

$$\Phi(x)=(\phi_1(x),..,\phi_l(x))^T$$

と書くとする. X の 2 つの要素 $x,\acute{x}$ に対して, カーネル関数 $k(x,\acute{x})$ は $x,\acute{x}$ それぞれの特徴ベクトル同士の内積

$$k(x,\acute{x})=\Phi(x)^T\Phi(\acute{x})=\sum_{m=1}^{l}\phi_m(x)\phi_m(\acute{x})$$

と定義される.
本研究では, カーネル関数としてよく使用されるガウスカーネル

$$k(x,\acute{x})=\exp\left(-\beta\|x-\acute{x}\|^2\right)$$

を用いる. ここで, $\|\cdot\|^2$ はユークリッド距離 2 乗, $\beta>0$ はあらかじめ決めておく非線形性を調整するパラメータと考えられる.

線形回帰の内積をカーネルにおきかえ, 非線形化することにより, カーネル回帰は与えられる.

2.3　カーネル回帰における L^2-正則化

一般にパラメータの次元が高くなると, 関数の表現能力が指数関数的に増大するため汎化能力が落ちる. すなわち回帰での過学習が発生する. カーネル回帰でも, 高次元に保ったまま関数の表現を抑える手法である正則化を用いる.

正則化は, K を $x_1,\ldots,x_n$ のカーネル行列, $\lambda>0$ を正則化項を制御するパラメータとし, サンプルに対する誤差の他に罰則項を付け加えた

$$R_{k,\lambda}(\alpha)=(y-K\alpha)^T(y-K\alpha)+\lambda\alpha^T K\alpha$$

を最小化することによって, 過学習を抑えるという方法である. この最適解を与える α は

$$\alpha=(K+\lambda I_N)^{-1}y$$

で表せる.

3　Wishart 行列

確率変数を要素とする行列は, 一般にランダム行列と呼ばれる. 特に, $G_{N,M}$ が独立に平均 0 分散 1 の標準 Gauss 分布に従う確率変数を要素に持つ $N\times M$ ランダム行列のとき Ginibre 行列と呼ばれ, それらの $N\times N$ の共分散行列

$$W_N=\frac{1}{N}G_{N,M}{}^tG_{N,M}$$

を Wishart 行列と呼ぶ.

W_N は N 個の固有値を持つ. この $N\to\infty$ における固有値漸近分布を考える. ここで極限 $N\to\infty$ の際 $N\times M$ Ginibre 行列の漸近縦横比は $M/N\to\gamma$ とする. この極限における固有値漸近分布は, パラメータ γ の Marchenko-Pastur 分布と呼ばれ

$$\begin{aligned} d\pi_\gamma(x)=&\frac{\sqrt{-(x-\gamma_-)(x-\gamma_+)}}{2\pi\gamma x}1_{[\gamma_-,\gamma_+]}(x)dx\\ &+\max\{0,1-\gamma\}\delta_0(x)\end{aligned}$$

で与えられる. ただし,$\gamma_\pm(\gamma)=(1\pm\sqrt{\gamma})^2$.

この極限分布は Wishart 行列のモーメントの極限から定まるものである. すなわち, Wishart 行列の k 次モーメント m_k

$$m_k=\lim_{N\to\infty}\frac{1}{N}\mathrm{Tr}\big(W_N^k\big)=\lim_{N\to\infty}\mathrm{tr}\big(W_N^k\big)$$

が分布 π_γ の k 次モーメントで与えられる.

3.1　Catalan 数

Marchenko-Pastur 分布において, $\gamma=1$ の時, つまり Wishart 行列が正方の Ginibre 行列のとき, そのモーメント列は Catalan 数 C_k で与えられる. ここで C_k は,

$$\begin{aligned} C_n&=\frac{1}{n+1}\binom{2n}{n}=\frac{(2n)!}{(n+1)!n!}\quad(n\geq 0)\\ &=1,1,2,5,14,42,132,429,\cdots\end{aligned}$$

となる自然数列である.

3.2 Wishart 行列のモーメントのゆらぎ

Wishart 行列のゆらぎモーメントとは, m_p と m_q の共分散の情報で

$$\alpha_{p,q} = \lim_{N\to\infty} \mathrm{Cov}\big[\mathrm{Tr}\big(W_N^p\big), \mathrm{Tr}\big(W_N^q\big)\big]$$

と定義される．この $\alpha_{p,q}$ は 2 次の漸近挙動で，モーメントより詳細な漸近情報を含む．例えば $\alpha_{k,k}$ は 極限 $N\to\infty$ における $\mathrm{tr}\big(W_N^k - m_k \cdot I\big)$ の収束のゆらぎを与えることになる．Wishart 行列の漸近的縦横比 $\gamma = 1$ の, ゆらぎモーメントは, 以下の式で与えられることが知られている．

$$\alpha_{p,q} = 2\frac{2pq}{p+q}\binom{2p-1}{p}\binom{2q-1}{q}$$

6 次までを, 以下に行列として書いておく．

$$\begin{pmatrix} 2 & 8 & 30 & 112 & 420 & 1584 \\ 8 & 36 & 144 & 560 & 2160 & 8316 \\ 30 & 144 & 600 & 2400 & 9450 & 36960 \\ 112 & 560 & 2400 & 9800 & 39200 & 155232 \\ 420 & 2160 & 9450 & 39200 & 158760 & 635040 \\ 1584 & 8316 & 36960 & 155232 & 635040 & 2561328 \end{pmatrix}$$

4 本研究での手法

実験の目標はある関数にノイズを加えたシュミレーションデータに対して, 最適なカーネル回帰モデルを見つけること, すなわち最適なカーネルパラメータ β と正則化パラメータ λ の値を推定することである．

1. まず β の選択を行う

   β の値を変えながらそのカーネル行列の優固有値の数を順次調べ, その優固有値を取り除きノイズ部分とみなせるようにカーネル行列のモーメント列が Catalan 数列に出来る限り近くなるような β を選択する．

2. 選択されたそれぞれの β について, 最適な正則化パラメータ λ を以下のように求める．

   (a) 設定された各 λ に対して, カーネル回帰モデルにおける残差の集合を求める．

   (b) 残差の集合からランダムリサンプリングにより 50×50 の Ginibre 行列の標本を作り, $\gamma = 1$ の Wishart 行列の標本を得る．

   (c) この Wishart 行列の 6 次までのモーメント列を求める．

   (d) 手順 (b),(c) を多数 ($N = 10,000$) 回繰り返す．

   (e) N サンプルの平均によりモーメントを求め, Catalan 数と比較する．

   (f) N サンプルの共分散行列を求め, 6 次までのゆらぎモーメントの行列と比較する．

以上の手順で最適な β と λ の値を選択することが可能かを実験する．また同データに対して 5-fold 交差検査を行い, 本手法と比較をする．

5 実験例

関数 $\sin 4\pi x(-0.5 \le x \le 0.5)$ の 50 標本点にノイズを加えたシミュレーションデータを用いる．まず β の選択を行う．

構造部を 1 次元として, β が 0.32 と選択された場合, $\lambda = 0.1$ では, 罰則項の影響が強すぎるため図 1 のように過度に滑らかな曲線を描いてしまう．

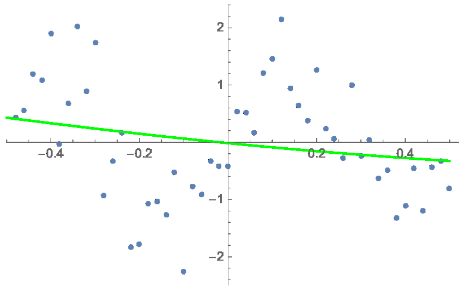

図 1: $\lambda = 0.1$ の場合

この時 6 次までのモーメント列は,
$(0.980, 1.934, 4.783, 13.286, 39.653, 124.322)$ になる．

様々な λ で比較し最適と推定されたのは, $\lambda = 10^{-8}$ であり, 6 次までのモーメント列は,
$(0.980, 1.943, 4.835, 13.528, 40.700, 128.715)$ であり
ゆらぎモーメントは, 以下のようになった．

$$\begin{pmatrix} 2.0026 & 7.9332 & 29.615 & 110.60 & 416.56 & 1583.4 \\ 7.9332 & 35.375 & 141.05 & 549.55 & 2132.3 & 8286.5 \\ 29.615 & 141.05 & 586.73 & 2355.1 & 9343.2 & 36943. \\ 110.60 & 549.55 & 2355.1 & 9665.9 & 39018. & 156456 \\ 416.56 & 2132.3 & 9343.2 & 39018. & 159734 & 648023 \\ 1583.4 & 8286.5 & 36943. & 156456 & 648023 & 2655260 \end{pmatrix}$$

またこの λ でのカーネル回帰曲線は図 2 のようになる．

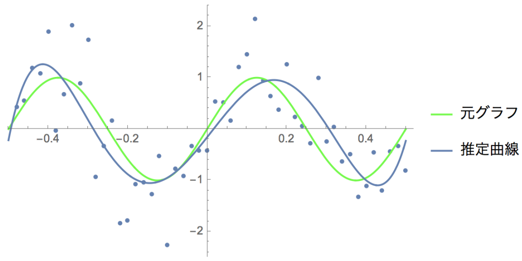

図 2: $\lambda = 10^{-8}$ の場合

また 5-fold 交差検査でも $\lambda = 10^{-8}$ が最適な λ 値として選択された．

6 まとめ

本研究で行った数値実験から, カーネル行列のノイズ部の推定に Wishart 行列を用いるとで最適な β の候補が見つけられること, 選択された β に対して最適な正則化パラメータ λ が, 残差集合からのリサンプリングによる大量の Wishart 行列の標本のゆらぎモーメントで選択可能であることが分かった．またこれらの選択は交差検証と両立することもわかった．

大谷 有美, Residual analysis of regression models by fluctuation moments of Wishart Matrices, *Non-Commutative Probability and Related Topics 2022*, Nov. 8, 2022, Hokkaido Univ., (Sapporo, Japan).