Vol.2010-MPS-81 No.40
Vol.2010-BIO-23 No.40
2010/12/17

カーネル非線形回帰へのランダム行列の応用

小　林　由　香[†1]

１次元データのカーネル非線形回帰を行う場合, ユーザーが最適な正則化パラメータ λ を与えなければならない. 本研究では, 非線形回帰においてランダム行列理論を用いて, λ を推定する手法の一つを提示する.

An application of random matrices to kernel non-linear regression

Yuka Kobayashi[†1]

When we use the kernel non-linear regression, we must give a suitable regularization parameter λ. In this study, we show the giving the parameter λ by applying the theory of random matrices.

1. はじめに

カーネル法における非線形回帰では, 最適なガウスカーネルのパラメータ β と, 正則化パラメータ λ をユーザーが与える必要があるが, この最適な β, λ を決めることは難しい. そこで本研究では, 通常ノイズが含まれていないとされる構造にノイズを入れることにより, 正則化パラメータ λ をランダム行列の理論を用いて推定する方法を提示する.

2. カーネル非線形回帰

カーネル関数とは, ２つの入力 $x = (x_1, \ldots, x_d)^T, x' = (x'_1, \ldots, x'_d)$ から計算される関数 $k(x, x')$ である. ここではガウスカーネル,

†1 お茶の水女子大学大学院 人間文化創成科学科
Graduate school of Humanities and Sciences, Ochanomizu University

$$k(x, x') = \exp(-\beta \parallel x - x' \parallel^2) \tag{1}$$

を用いることにする. ただし, $\parallel \quad \parallel^2$ は $z = (z_1, \ldots, z_d)$ に対して $\parallel z \parallel = \sum_{m=1}^d z_m^2$ とする. 関数 k は, $x = x'$ のとき最大値 1 を取り, 直観的には x と x' の近さを表す量になる. また, パラメータ β はあらかじめ適当に決めておく.

カーネル回帰では, x に対して

$$y = \sum_{j=1}^n \alpha_j k(x^{(j)}, x) \tag{2}$$

という関数をあてはめる. 与えられた x に対して, 各サンプル $x^{(j)}$ との近さを測った $k(x^{(j)}, x)$ を一つの成分と見て, それらを α_j という重みで足し合わせたモデルである. 線形の場合と同様に二乗誤差

$$r_k(y, x; \alpha) = (y - \sum_{j=1}^n \alpha_j k(x^{(j)}, x))^2 \tag{3}$$

を最小にするように $\alpha = (\alpha_1, ..., \alpha_n)^T$ を決める. ここで, (i, j) 成分が $K_{ij} = k(x^{(j)}, x^{(i)})$ である行列を

$$K = \begin{pmatrix} k(x^{(1)}, x^{(1)}) & k(x^{(2)}, x^{(1)}) & \ldots & k(x^{(n)}, x^{(1)}) \\ k(x^{(1)}, x^{(2)}) & k(x^{(2)}, x^{(2)}) & \ldots & k(x^{(n)}, x^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ k(x^{(1)}, x^{(n)}) & k(x^{(2)}, x^{(n)}) & \ldots & k(x^{(n)}, x^{(n)}) \end{pmatrix} \tag{4}$$

とおくと, 二乗誤差の総和は

$$R_k(\alpha) = \sum_{i=1}^n r_k(y_{(i)}, x_{(i)}; \alpha) = (y - K\alpha)^T (y - K\alpha) \tag{5}$$

と得られる. 解は（K が正則行列なら）

$$\alpha = (K^T K)^{-1} K^T y \tag{6}$$

である. さらに, 任意の x, x' に対して $k(x, x') = k(x', x)$ が成り立ち, K は対称行列となり, $K^T = K$ より $\alpha = K^{-1} y$ と書ける.

カーネル関数を使って拡張した式 (2) へのモデルのあてはめは「カーネル回帰」と呼ばれ

Vol.2010-MPS-81 No.40
Vol.2010-BIO-23 No.40
2010/12/17

る．パラメーター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．そこで, カーネル法では表現能力を抑える正則化を用いる．正則化は, サンプルに対する誤差のほかに余分な項を付け加えたものを最小化することによって, カーネル関数の表現能力を落とすといういう方法である．ここでは, $\alpha^T K\alpha$ という α の 2 次形式を λ という正数で重みづけて加え,

$$R_{k,\lambda}(\alpha) = (y - K\alpha)^T (y - K\alpha) + \lambda\alpha^T K\alpha, \qquad \lambda > 0 \tag{7}$$

という関数の最小化を行う．この式も, α の 2 次関数なので,

$$-K(y - K\alpha) + \lambda K\alpha = 0 \tag{8}$$

となる．ここで K が正則とすれば,

$$\alpha = (K + \lambda I_n)^{-1} y \tag{9}$$

という解が得られる．(I_n は n 次の単位行列) このような方法は「正則化」と呼ばれる．正則化を行なう際に加えた $\lambda\alpha^T K\alpha$ を正則化項と呼び, その強さを調節している λ を正則化パラメータと呼ぶ．汎化能力を最適化するように λ を決めるのは難しい．そこで, 本研究ではランダム行列の理論を用いて, この λ を求める手法を提示する．

3. 提案手法

3.1 実験概要

非線形構造データの, カーネル回帰を行うことにおいて, ノイズを含まない説明変数 x に異なる分散で正規乱数を加えたものを N 個 $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N)$ 用意する．ただし, 各回分散は適当に取る．各 $\tilde{x}_i$ を用いてカーネルを計算し, スペクトルを一つにまとめ, ヒストグラムを描き Marcenko-Pastur 分布により雑音部と構造部の推定しカーネル回帰を行う．

3.2 数値実験例

モデル $y = \sin x + \sin 7x + \epsilon$, $\epsilon \sim N(0, 0.3)$ において, π を 256 の等分したものを x とする．x の各成分に N(0,0.12) で発生させた正規乱数を加える．分散を 0.01 刻みで変化させ N(0, 0.21) まで行い, $N = 10$ で各カーネル行列を求める．求めたカーネル行列それぞれの固有値を調べ, 一つにまとめヒストグラムを描く (図 1)．ここで, カーネル関数の β は 10.0 と与える．図 1 に Marcenkok-Pastur 分布を重ね合わせ, 構造部として取り出す部分を決定する．重ね合わせたものが図 2 である．本実験では,Marcenkok-Pastur 分布の重ね合わせから固有値の上位から 30 個抽出することになった．ノイズを含まない x において, 上位 30 個の固有値と固有ベクトルで, 正則化パラメータ λ が定まる．図 3 は今回近似するデータであり, 図 4 は提案手法で正則化行った場合の結果である．

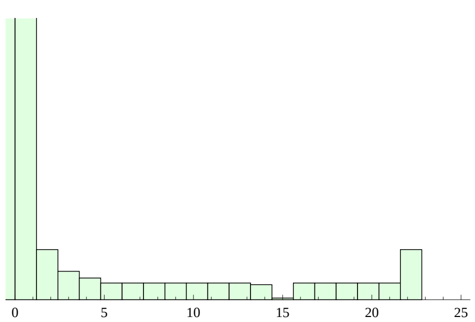

図 1 x_i のスペクトル
Fig. 1 nonliner structure

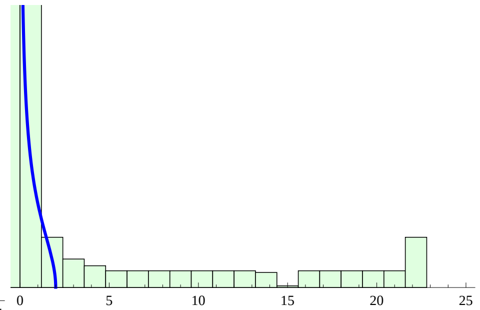

図 2 x のスペクトルと Marcenko-Pastur 分布
Fig. 2 nonliner structure and Marcenko-Pastur distribution

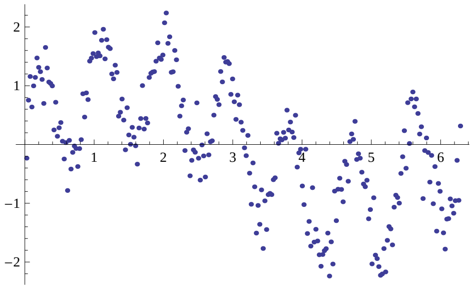

図 3 非線形構造
Fig. 3 nonliner structure

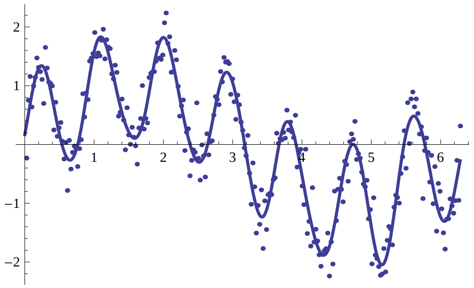

図 4 平滑化
Fig. 4 smoothing

4. まとめ

ランダム行列理論を援用し, 正則化パラメータ λ を推定し平滑化を行った．λ の推定における累積寄与率との関係を調べることを今後の課題としたい．

参考文献

1) 赤穂昭太郎:カーネル多変量解析 非線形データ解析の新しい展開 (2008)