パラメータの変動に伴うガウス型カーネル行列の固有値の挙動

理学専攻　情報科学コース　1840667 根本優花　(指導教員：吉田裕亮)

1　はじめに

カーネル法は非線形データ解析の手法の一つである．近年はサポートベクトルマシンの発明によって一躍注目されるようになった．カーネル法を扱う際には，カーネル関数と呼ばれる非線形の関数を用いる．カーネル法の最も重要な特徴は，カーネル関数に線形手法を用いて非線形問題を解くという点である．この方法は複雑な最適化問題を避け，大規模なデータ解析をすることを可能にしている一方で，従来のパターン認識とは逆転の発想を用いてデータを一旦高い次元の空間に移してから処理を行うというカーネルトリックを用いる．

本研究では，カーネルアルゴリズムの多くが固有値問題に帰着していることを踏まえ，カーネル関数によって生成されるカーネル行列とその固有値に関する実験をもとに，クラスタリングにおけるクラスタ数の推定ならびにカーネルパラメータの推定，データ次元の削減に関する提案を行う．

2　カーネル法とカーネル行列

2.1　カーネル関数

一般に，$\phi_1, \cdots, \phi_d$ 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，$\Phi(x) = (\phi_1(x), \cdots, \phi_d(x))^T$ と書くことにする．X の二つの要素 x, x' に対し，カーネル関数 $k(x, x')$ は x, x' それぞれの特徴ベクトルどうしの内積

$$k(x, x') = \Phi(x)^T \Phi(x') = \sum_{m=1}^{d} \phi_m(x)\phi_m(x')$$

として定義される．

2.2　ガウスカーネル

以下の式によって表される関数をガウスカーネルという．ただし，$||\cdot||$ はベクトルノルムを表し，$\beta > 0$ はあらかじめ適当に定められるパラメータである．

$$k(x, x') = \exp(-\beta||x - x'||^2)$$

この式は正規分布の確率密度関数と同じ形をしている．このカーネルをガウスカーネルという．一般にカーネル行列は正定値性をもつ．

2.3　カーネル行列

カーネル関数を要素にもつカーネル行列

$$K = \begin{pmatrix} k(x_1, x_1) & k(x_2, x_1) & \cdots & k(x_n, x_1) \\ k(x_1, x_2) & k(x_2, x_2) & \cdots & k(x_n, x_2) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_1, x_n) & k(x_2, x_n) & \cdots & k(x_n, x_n) \end{pmatrix}$$

を考えたとき，その二次形式が常に非負，すなわち

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j K_{ij} \geq 0$$

が任意の n 次元ベクトル $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)^T$ について成り立つとき，カーネル関数 $k(x, x')$ は正定値性をもつ．

3　Wishart 行列と Marchenko-Pastur 則

3.1　Wishart 行列

G を $N \times M$ の独立に平均 0 分散 1 の標準ガウス分布に従う変数を要素に持つランダム行列，その共分散行列を，

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

とする．この $N \times N$ のランダム行列 W_N を Wishart 行列という．

3.2　Marchenko-Pastur 則

Wishart 行列の固有値漸近分布を与える，Marchenko-Pastur 分布がある．$N \times M$ の Ginibre 行列の漸近的縦横比を λ とおくとき，その固有値漸近分布は

$$d\pi_\lambda(x) = \frac{\sqrt{-(x-\lambda_-)(x-\lambda_+)}}{2\pi\lambda x} 1_{[\lambda_-, \lambda_+]}(x)dx + \max\{0, 1-\lambda\}\delta_0(x)$$

で与えられる．ただし，$\lambda_\pm(\lambda) = (1 \pm \sqrt{\lambda})^2$ である．

3.3　Catalan 数

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

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

となる自然数列である．

4　クラスタ数と優固有値の関係性について

4.1　実験概要

正規乱数を使ってクラスタを生成し，これに対しパラメータ β を動かしながらガウス型カーネル行列の固有値を求める．このとき，クラスタ数や配置を様々に用意し，パラメータ β の値に伴う優固有値の挙動を考察する．

4.2　実験結果

図１－２，図２－２，図３－２の固有値分布の全てのグラフに関して、横軸はパラメータ β の値，縦軸は固有値の値を示す．

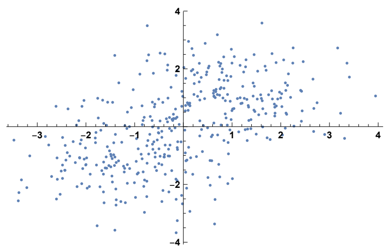

図１－１：クラスタ数２の正規分布

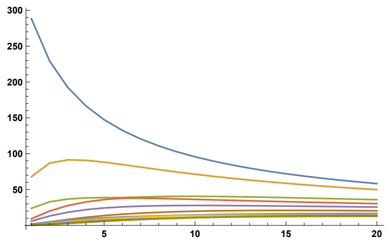

図１－２：図１－１の固有値分布

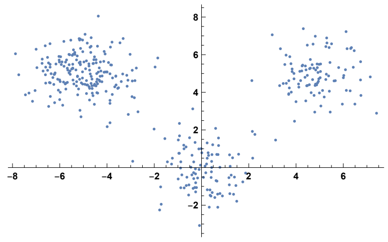

図２－１：クラスタ数３の正規分布

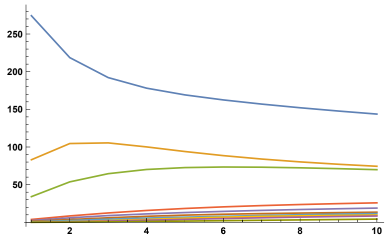

図２－２：図２－１の固有値分布

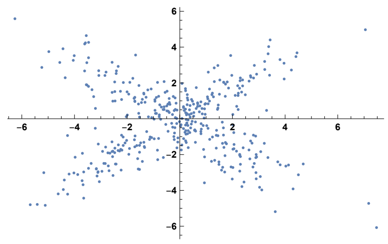

図３－１：２つの正規分布がクロスした１つのクラスタ

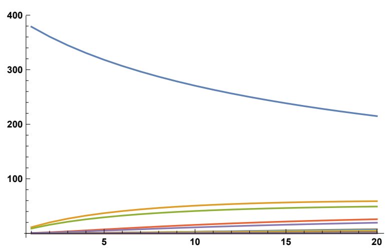

図３－２：図３－１の固有値分布

図３－１に関しては，２つの正規分布が交差したクラスタである．実験前は分散が２方向に伸びているため，優固有値の個数は２つになるのではないかとの予測を立てていたが，図３－２の結果を見ると実際には優固有値の個数は明らかに１つになった．それぞれの固有値分布のグラフを見ると，上から数本のグラフが明確に分かれていることがわかる．そしてその本数は，クラスタ数に対応していることがわかる．これらの結果により，クラスタ数が複数個ある場合について、優固有値とクラスタ数は一致するということが充分にいえる．

5　ガウスカーネルの最適なパラメータ β

先の実験結果で得られたガウス型カーネル行列の固有値とパラメータ β のグラフにより，カーネル関数を扱う際に重要となるパラメータ β の決定方法の提案を行う．

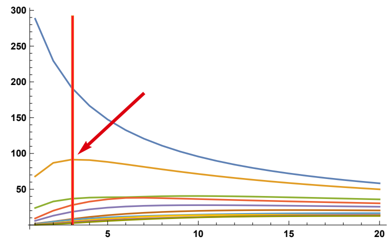

図４：クラスタ数２の正規分布の固有値分布

図４は先述したクラスタ数が２つの正規分布のパラメータ β の変動に伴う固有値のグラフである．図４中の矢線で示された第２固有値を観察する．すなわち第２固有値が極大点を与える β を最適な β として与える．

6　データの次元の縮約の提案

先に紹介した方法で最適な β を決定したのち，ガウス型カーネル行列の固有値を求め，その優固有値を上からひとつずつ取り除き、そのモーメント列を求める．モーメント列が Catalan 数にできるだけ近くなるまで優固有値を取り除く．取り除かれた優固有値の数が有効なデータの次元になる．それ以降の固有値はランダムな正方行列の積の Wishart 行列の固有値，すなわち $\lambda = 1$ の Marchenko-Pastur 則に対応すると考えられるので，取り除いた部分が本質的な主要部になる．この方法を用いれば，データの次元を大幅に縮約し，計算量を大幅に削減することも可能となる．

7　考察と今後の展望

本研究では、カーネル行列とその固有値の挙動を観察することにより得られた３つのトピックスについて述べた．これらの実験結果や提案をカーネル法を扱う際に適用することで，より分析の高精度化と効率化が期待できる．それぞれの提案について補足を述べると，クラスタ数と優固有値の個数の関係性については，クラスタ数が目視で明らかに決定できない正規分布の場合，優固有値が明らかに同じ個数になったとは断定できなかった．また，ガウスカーネルの最適なパラメータ β の決定方法に関しては，クラスタ数が１つの場合には適用できないため，別の方法でパラメータを推定する必要があるといえる．またこれらの実験については，自ら推定しやすい正規分布を作成したので実データに適用するまでに至っていない．しかし，いずれも理論的に推測したものを実験で示すことに成功しているので，今後実験を続けることで更なる発展も期待できるといえる．

参考文献

[1] 赤穂昭太郎「カーネル多変量解析」岩波書店,2008

[2] A.M. Tulino and S. Verdu「Random Matrix Theory and Wireless Communications」Now Pub,2004