DP-means 法の非線形化とノイズ推定

理学専攻　情報科学コース　北口景子　(指導教員：吉田裕亮)

1　はじめに

データをまとまりごとに集めてグループ分けをするクラスタリングという手法がある．非線形なデータをグループ分けするには，カーネル k-means 法やスペクトラルクラスタリングが一般的であるが，前者は初期値依存が強く適切な解を得るのに何度も初期値を変えて繰り返さなければならない場合がある．また，どちらも事前にユーザがクラスタ数を設定する必要があり，設定されたクラスタ数により結果が大きく変化する．最近，ディリクレ過程混合モデルと k-means 法を組み合わせた DP-means 法という手法に対し，カーネル法を用いて非線形化することで，初期値に依存せず，クラスタ数を推定する非線形クラスタリング手法が考えられた．しかし，この場合実験データによっては適切なクラスタリング結果を得るのが困難な場合がある．そこで本研究では，データから得られるカーネル行列の固有値分布を，ランダム行列理論で知られている固有値分布と，モーメント法を用いて照らし合わせてデータの構造部を推定する．手法を援用し，構造部の固有値のみを用いて，カーネル DP-means 法によるクラスタリングを行うことにより，その精度を向上させることを試みる．

2　ディリクレ過程

ベイズ統計モデルの一つに，ノンパラメトリックベイズ法がある．ノンパラメトリックベイズ法は，データに応じてモデルの複雑さを無限に伸縮し自動的に学習する統計モデルであり，入力データに応じて適切な混合数を決定する．ノンパラメトリックベイズ法の最も基本となるモデルとしてディリクレ過程が挙げられる．ディリクレ過程は，基底測度と呼ばれるある確率分布 G_0 から，それに似た無限次元の離散分布 G を生成する確率過程であり，次のように表記される．

$$G \sim DP(\gamma, G_0)$$
$$G = (\theta_1, \theta_2, \theta_3, \ldots)$$

ここで，γ>0 は G が平均的にどれくらい G_0 に似ているかを制御する，学習可能なパラメータである．

2.1　ディリクレ過程混合モデル

ディリクレ過程では，それぞれのデータにパラメータ θ が対応付けられているため，データ数が n であれば，最大 n 混合モデルまで実現することができる．しかし，ディリクレ過程は，データが観測される毎に要素分布数が必要に応じて増える柔軟なデータ生成過程となっているので，現実には n 混合モデルになることはほとんどない．換言すると，$\theta_1 \sim \theta_n$ は全て異なるわけでなく，K 個の適応的に値の異なるパラメータ $\theta_{(1)} \sim \theta_{(K)}$ から成り立っている $(1 \leq K \leq n)$．このとき，データは K 個のクラスタに分割される．これらの性質を用いて構成するモデリングが，ディリクレ混合過程モデルである．ディリクレ過程混合モデルのデータ生成手順は，以下で表記される．

(1) $G \sim DP(\gamma, G_0)$,

(2) $\theta_i \mid G \sim G, for\ i = 1, \ldots, n$,

(3) $x_i \sim p(x \mid \theta_i), for\ i = 1, \ldots, n$.

3　DP(Dirichlet Process)-means 法

DP-means 法 [1] は以下のアルゴリズムでクラスタリングを行う手法である．
初期値設定：$k = 1, l_1 = \boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$ とし，$\boldsymbol{\mu_1}$は全データの平均とする．λ の値は適当に設定しておく．
$\boldsymbol{x_i}(i = 1, \ldots, n)$ について各パラメータの値が収束するまで以下の 1.〜3. を繰り返す．

1. $\boldsymbol{x_i}(i = 1, \ldots, n)$ について，
   - $d_{ic} = \| \boldsymbol{x_i} - \boldsymbol{\mu_i} \|^2\ (c = 1, \ldots, k)$,
   - $\min_c d_{ic} > \lambda$ の場合，$k = k + 1, z_i = k$, $\boldsymbol{\mu_k} = \boldsymbol{x_i}$,
   - それ以外の場合 $z_i = \mathrm{argmin}_c d_{ic}$.
2. $z_1, \ldots, z_k$ に基づき，クラスタ $l_1, \ldots, l_k$ に各データを割り当てる．
3. $l_j(j = 1, \ldots, k)$ について $\boldsymbol{\mu_j} = \sum_{x \in l_j} \boldsymbol{x}$ を求める．

このとき k はクラスタ数，l_i は i 番目のクラスタ，$\boldsymbol{\mu_i}$ はクラスタ l_i の平均，z_i は $\boldsymbol{x_i}$ が所属するクラスタ番号を表す．また，λ はクラスタ数 k を制御するパラメータであり，一般に λ の値が大きければ k の値は小さく，λ の値が小さければ k の値は大きくなる傾向にある．適切なクラスタリング結果を得るには，適切な λ を選択する必要がある．

4　カーネル DP-means 法

カーネル法 [3] では，非線形なデータを一度入力空間から高次元の特徴空間上に写像することで解析しやすいデータに変換し，その特徴空間上で線形なモデルを組み立てて問題を解く．カーネル法を用いて DP-means 法を非線形化したものをカーネル DP-means 法という．

一般に k-means 法をカーネル法により非線形化したカーネル k-means 法は次のように固有値問題として解くことができる．このとき K はカーネル行列を表し，Z は各データが所属するクラスタを表す行列である．

$$Z = \underset{\{Y \mid Y^T Y = I\}}{\arg\max}\ \mathrm{tr}(Y^T K Y), Y = Z(Z^T Z)^{-\frac{1}{2}}$$

これと同様に，カーネル DP-means 法は，以下の式のように Y を求めることでクラスタリング結果を得ることができる．

$$Z = \underset{\{Y \mid Y^T Y = I\}}{\arg\max}\ \mathrm{tr}(Y^T (K - \lambda I) Y)$$

5　ランダム行列理論

一般に，ランダム行列とは確率変数を要素に持つ行列であり，代表例に Wishart 行列が挙げられる．

5.1　Wishart 行列

各成分が独立に $N(0,1)$ の標準正規分布に従う変数をもつ $n \times p$ の行列を C とする．このランダム行列 C から

$$S = \frac{1}{n} CC^T$$

で求められる $n \times n$ 対称ランダム行列 S を Wishart 行列という．$p/n = \lambda$ を保ちながら，$n \to \infty, p \to \infty$ の極限をとると，Wishart 行列 S の固有値経験分布は，$\lambda_{min} \leq t \leq \lambda_{max}$ のときに以下の確率密度関数に収束することが知られている．

$$p(t) = \frac{1}{2\pi} \frac{\sqrt{-(t-\lambda_{max})(t-\lambda_{min})}}{\lambda t},$$
$$\lambda_{min}^{max} = (1 \pm \sqrt{\lambda})^2$$

また，このような確率密度関数を持つ分布は Marcenko-Pastur 分布と呼ばれている．

5.2　ガウスカーネル

様々なカーネルの中でもガウスカーネル

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

を用いて特徴空間に写像した行列は，相関行列と同じような振る舞いをすることが知られている．また，Wishart 行列の固有値分布と，ガウスカーネルで写像した特徴空間における内積行列の固有値分布は等価であると知られており，これによりガウスカーネル行列におけるノイズ部に相当する固有値分布も Marchenko-Pastur 分布と同様の性質を持つことがわかる．

6　モーメント法

$f(x)$ を連続確率変数 X の密度関数とするとき，原点まわりの k 次モーメント m_k は，

$$m_k = \int_{-\infty}^{\infty} x^k f(x) dx$$

と表せ，これらの値は確率分布の特徴を与える．また，先にランダム行列理論で述べた Marcenko-Pastur 分布のモーメントは，

$$m_k = \frac{2k!}{k!(k+1)!} m_1^k$$

で与えられる．

本研究では，観測データからの標本モーメント列を理論値と比較することにより，最適なノイズ部の推定を行う．

7　提案手法

以下の手順でカーネル DP-means 法によるクラスタリングを行うことを提案する．また，本研究ではより精度の高いクラスタリング結果を得るために，カーネル行列 K を直接用いるのでなく，K から以下の類似度行列 W を算出し，K の代わりに W を用いてクラスタリングを行うこととする．このとき，$k \in \mathbb{N}$.

$$W_{i,j} = \begin{cases} 1 \text{ if } \boldsymbol{x_i} \in kNN(\boldsymbol{x_j}) \text{ or } \boldsymbol{x_j} \in kNN(\boldsymbol{x_i}) \\ 0 \text{ otherwise} \end{cases}$$

1. サンプルデータのガウスカーネル行列 K を構成し，固有値分布を求める．
2. Marcenko-Pastur 分布のモーメント列に最も適合するように，1 で求めた固有値からモーメント法を用いてノイズ部と構造部を推定する．
3. 構造部の固有値のみを用いてカーネル行列 K' を再構成する．
4. 新たな K' を用いて類似度行列 W を構成し，カーネル DP-means 法によるクラスタリングを行う．

8　実験例

このデータに対し提案手法によってクラスタリングを行った結果は以下の通りである．線形で分けることの出来ないデータを用意する．サンプルデータとして，1 群が 300 点，2 群が 100 点，3 群が 100 点の合計 500 点で構成されている．λ の値とガウスカーネルにおける β の値はあらかじめ適当な値に設定しておく．

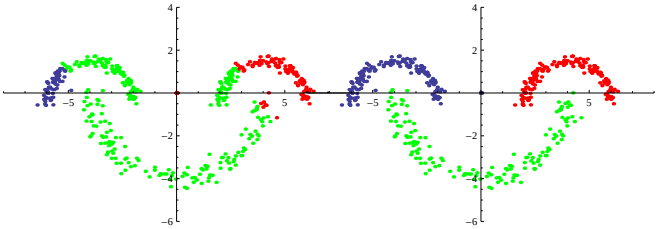

図 1: ノイズ推定を行わない場合　　図 2: ノイズ推定を行った場合

9　結果とまとめ

モーメント法によるノイズ推定を行った場合のクラスタリング結果は図 2 のようになった．このとき，λ の値は 12.3 であり，クラスタ数 k は 3 となった．

サンプルデータについて，期待するクラスタリング結果，クラスタ数を得ることができたが，λ の多少の変化によって結果が異なってしまうため，λ の値は細かい調整が必要で，適切な λ を選択することは難しい．今後は，適切な λ の定め方を検討する課題が残されていると考える．

参考文献

[1] Jordan, M.I. and EDU, B.: *Revisiting k-means: New Algorithms via Bayesian Nonparametrics*, arXiv preprint arXiv:1111.0352(2011).

[2] 伊藤里江:ランダム行列理論を用いた Gaussian カーネルにおける雑音の推定，お茶の水女子大学大学院理学専攻情報科学コース修士論文,2009 年.

[3] 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，岩波書店.