

距離の分布に基づくカーネルk-平均法のパラメータ探索法

柳 信 一

北海道情報大学

A Search Algorithm for Gaussian Parameter of kernel
k-means method based on distribution of distance

Shinichi YANAGI

Hokkaido Information University

平成26年11月

北海道情報大学紀要 第26巻 第1号別刷

〈論文〉

距離の分布に基づくカーネル k -平均法のパラメータ探索法

柳 信一*

A Search Algorithm for Gaussian Parameter of kernel k -means method based on distribution of distance

YANAGI SHINICHI*

要旨

本論文では、データ間の距離分布に基づきカーネル k -平均法のガウス関数のパラメータを逐次的に計算する手法を提案する。カーネル k -平均法は、代表的なクラスタリング手法である k -平均法に基づいており、クラスタ同士の境界が非線形であるデータの分類が可能である。しかし、計算結果は距離計算に用いるガウス関数のパラメータに強く依存し、パラメータ調整が難しいという問題点がある。提案手法は、パラメータの更新と従来のカーネル k -平均法の計算を交互に繰返し行う。いくつかの人工データと実データを用いて計算機実験を行い、クラスタ間が十分な距離で分離しているデータに対する提案手法の有効性を示す。

Abstract

This paper propose a kernel k -means algorithm to compute a Gaussian parameter based on distribution of distance in the original kernel k -means algorithm repeatedly. The kernel k -means algorithm is based on the k -means algorithm which is a well known clustering algorithm, and make possible to identify clusters that are non-linearly separable in input space. However, a conclusion identifying clusters with kernel k -means depends heavily on a chosen Gaussian parameter which used for computing distance between data and determining an optimal Gaussian parameter is difficult. The proposed algorithm identifies clusters with the original kernel k -means algorithm and search a Gaussian parameter repeatedly. An experimental comparison in some artificial and real data set shows that a performance of the proposed algorithm is efficient for a data that each cluster is sepelated with adequate gap.

キーワード

クラスタリング, k -平均法, カーネル法, ガウス関数, パラメータ

1. まえがき

クラスタリングとは、データ集合をある基準によって分類することである。 k -平均法 (MacQueen, 1967) は代表的なクラスタリング手法であり、アルゴリズムが単純で、事前に必要とするパラメータがクラスタ数のみであるという利点のため、クラスタリングを必要とする計算過程において利用されることが多い。しかし、距離計算にユークリッド距離を使用するため、各クラスタの分離境界が非線形となるデータの分類に適していない。

カーネル k -平均法 (Girolami, 2002) は、 k -平均法に対してカーネル法 (Muller, Mika, Ratsch, Tsuda and Scholkopf, 2001) を適用した手法であり、非線形に分離されているクラスタの分類が可能である。しかし、計算結果が距離計算に使用するカーネル関数のパラメータに敏感であり、カーネル関数のパラメータの調整が困難であるという問題点がある。

他にカーネル k -平均法と同じ問題点がある手法としては、重み付き距離を適用する重み付きカーネル k -平均法 (Dhillon, Guan and Kulis, 2004), 重み付きカーネル k -平均法と等価であるスペクトラルクラスタリング (Ng, Jordan and Weiss, 2002), 機械学習手法の一つであるサポートベクターマシン (Vapnik Vladimir,

* 北海道情報大学経営情報学部システム情報学科講師,
Lecturer, Department of Business and Information Systems, HIU

1998) を利用するサポートベクタークラスタリング (Ben-Hur, Horn, Siegelmann and Vapnik, 2002) などがある。上記の手法で使用するカーネル関数は、ほとんどの場合がガウス関数であり、パラメータはデータ間の距離に関するスケール調整の意味を持つ。適切なパラメータを選択する手法としては、Yu ら (Yu, Tranchevent, De Moor and Moreau, 2011) が、スペクトラルクラスタリングとパラメータの最適化を交互に行う手法を提案しているが、事前に解となるパラメータの有限列を用意する必要があり、依然としてパラメータ選択の困難さが残る。また、文献 (柳, 2012) では、Yu らの手法と同様、カーネル k -平均法とパラメータの更新を交互に行う手法を提案し、計算機実験によりいくつかのデータに対して有効性を示しているものの、入力であるガウス関数の初期パラメータへの依存が強いという問題点がある。

本論文では、文献 (柳, 2012) の手法において、データ間の距離の分布に基づきカーネル関数のパラメータを計算するように改良したアルゴリズムを提案する。いくつかの人工データと実データを用いて文献 (柳, 2012) の手法との比較実験を行い、提案手法はクラスタ間が十分な距離で分離しているデータに対して有効であることを示す。

2. カーネル k -平均法

クラスタリングとは、対象とする n 個のデータからなるデータ集合 $X = \{x_i | i = 1, 2, \dots, n, \forall x_i \in R^d\}$ を何らかの基準により k 個の部分集合 $C_j = \{x | \forall x \in X\}$ に分類することであり、この部分集合 $C_j (j = 1, 2, \dots, k)$ をクラスタと呼ぶ。一般的にクラスタリングを必要とする場面では、データ集合のみが与えられ、クラスタ数 k は未知であることが多いが、本論文で k は既知のものとして取り扱う。

k -平均法は各データ $x_i \in X (i = 1, 2, \dots, n)$ と、クラスタの中心 $m_j \in R^d (j = 1, 2, \dots, k)$ との距離の総和を最小とする最適化問題として式 (1) で定式化される。

$$\begin{aligned} \min_{C_j} & \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - m_j\|^2, \\ \text{s.t.} & \bigcup_{j=1}^k C_j = X, \\ & C_j \cap C_l = \emptyset, 1 \leq j, l \leq k, j \neq l. \end{aligned} \quad (1)$$

式 (1) の $\|\cdot\|$ は L_2 ノルムである。式 (1) の解はクラスタ $C_j (j = 1, 2, \dots, k)$ であり、クラスタ中心 m_j は式 (1) より、 C_j に属するデータ数 $|C_j|$ を用いて式 (2) のように求まる。

$$m_j = \frac{1}{|C_j|} \sum_{x \in C_j} x, (j = 1, 2, \dots, k). \quad (2)$$

式 (1) の定式化は暗に各クラスタが凸形状となっていることを仮定しているため、線形分離不可能なクラスタからなるデータの分類に適していない。そこで、 k -平均法にカーネル法を適用するカーネル k -平均法 (Girolami, 2002) が提案されている。カーネル法は、無限次元まで含めた、現在の空間の次元数 d よりはるかに大きい $D (\gg d)$ 次元空間上に、写像 ϕ により $x \in R^d$ を非線形写像し、空間 R^D 上でデータの線形分離を試みる方法である。最大の特徴は、計算量削減のため実際の写像 $\phi(x)$ 自体の計算は行わず、式 (3) のカーネル関数を利用して $\phi(x)$ と $\phi(x')$ の内積のみを使用することで問題を解くことである。

$$K(x, x') = \phi(x) \cdot \phi(x') \quad (3)$$

具体的なカーネル関数としては、カーネル関数が高次元空間における内積となっていることを保証するマーセルカーネル (Máté, 1989) が利用されることが多く、特に多くの研究では、代表的なマーセルカーネルである式 (4) のガウスカーネルが使用されている。

$$K(x, x') = \exp\left(\frac{-\|x - x'\|^2}{\delta^2}\right). \quad (4)$$

式 (4) において、 δ はスケールパラメータを表す。カーネル k -平均法では、式 (1) の最適化問題は式 (5) のように表現される。

$$\begin{aligned} \min_{C_j} & \sum_{j=1}^k \sum_{x_i \in C_j} \|\phi(x_i) - M_j\|^2, \\ \text{s.t.} & \bigcup_{j=1}^k C_j = X, \\ & C_j \cap C_l = \emptyset, 1 \leq j, l \leq k, j \neq l. \end{aligned} \quad (5)$$

また、式 (2) のクラスタ中心は式 (6) のように表現される。

$$M_j = \frac{1}{|C_j|} \sum_{x \in C_j} \phi(x), M_j \in R^D (j = 1, 2, \dots, k). \quad (6)$$

写像 ϕ が未定義の場合、式 (6) のクラスタ中心を計算することはできないが、クラスタ中心 $M_j \in R^D$ とデータ $\phi(x) \in R^D$ の距離は式 (7) で計算することができる。

$$\begin{aligned} & \| M_j - \phi(x) \|^2 \\ &= \| M_j \|^2 - 2M_j \cdot \phi(x) + \phi(x) \cdot \phi(x) \\ &= \frac{1}{|C_j|^2} \sum_{x, x' \in C_j} K(x, x') - \frac{2}{|C_j|} \sum_{x' \in C_j} K(x, x') \\ &\quad + K(x, x), (j = 1, 2, \dots, k). \quad (7) \end{aligned}$$

式 (5) の問題は、可能な C_j の選択に関して組合せ的な探索を必要とする離散最適化であるため NP 困難となる。したがって、局所解の出力を認める以下の発見的アルゴリズムを用いて計算する。

Algorithm 1 入力: $X, k, \delta, C_1^{(1)}, C_2^{(1)}, \dots, C_k^{(1)}$

- 1 繰返し番号 t を $t = 1$ で初期化する。
 - 2 各データに対して各クラスタ $C_j^{(t)}$ の中心との距離を式 (7) により計算し、最も近いクラスタ中心のクラスタに割り当てる。
 - 3 2の結果から式 (5) の目的関数を再計算し、変化が無ければ $C_j^{(t)}$ ($j = 1, 2, \dots, k$) を出力して計算を終了する。そうでなければ $t = t + 1$ として2へ戻る。
-

Algorithm 1 の入力 $C_j^{(1)}$ ($j = 1, 2, \dots, k$) は初期クラスタである。通常は、対象データ $x \in X$ をランダムに割り当てることが多い。

3. カーネル k -平均法の問題点と従来法

カーネル k -平均法は δ に対して最適化していないため、 δ の変化に関して式 (5) の目的関数を最小とする C_j ($j = 1, 2, \dots, k$) が最適とは限らない。そのため、異なる δ に対するクラスタリング結果の評価が難しい。また、最適化のパラメータとして δ を加えて C_j ($1 \leq j \leq k$) と δ に関して式 (5) を解くことも困難である。まず、カーネル k -平均法による分類可能性の基準を考える。与えられたデータ集合 X において、分類結果として期待するクラスタ集合が存在すると仮定し、 $C^t = \{C_j^t \mid C_j^t \subseteq X, j = 1, 2, \dots, k\}$ とする。次に、 $0 < \delta < \infty$ に対して Algorithm 1 が出力する任意のクラスタ集合を $C_{All} \subset 2^X$ とし、全 C_{All} からなる集合を $C_{All}^* \subset 2^{2^X}$ とする。したがって、Algorithm 1 がある計算で C^t を出力する必要条件は、 $C^t \in C_{All}^*$ となる。

次に δ の条件を考える。まず、簡単のため二つのクラスタに限定し、 d 次元空間上のデータ集合 $X = C_a^t \cup C_b^t$ ($C_a^t \cap C_b^t = \emptyset$) を考える。このとき、 D 次元空間上においてカーネル k -平均法で二つのクラスタを分離できる条件は、 D 次元空間上での C_a^t と C_b^t のクラスタ中心点 $M_a \in R^D$ と $M_b \in R^D$ を垂直二等分する超平面が各クラスタと交差しないようなパラメータ δ が存在することである。すなわち、任意の $x \in C_a^t, x' \in C_b^t$ に対して式 (8) を満たす δ が存在する必要がある。

$$\begin{aligned} & \left((M_a - M_b) \cdot \phi(x) - \frac{\|M_a\|^2 + \|M_b\|^2}{2} \right) \times \\ & \left((M_a - M_b) \cdot \phi(x') - \frac{\|M_a\|^2 + \|M_b\|^2}{2} \right) < 0. \quad (8) \end{aligned}$$

式 (8) の左辺を $g_{ab}(x, x', \delta)$ という関数に置き換え、式 (8) を満たす δ の集合 Δ_{ab} を式 (9) で定義する。

$$\Delta_{ab} = \{ \delta \mid g_{ab}(x, x', \delta) < 0, \forall x \in C_a^t, \forall x' \in C_b^t \}. \quad (9)$$

したがって、 $\{C_a^t, C_b^t\} \in C_{All}^*$ となる必要十分条件は $\Delta_{ab} \neq \emptyset$ となる。

次に、 k 個のクラスタ C_i^t ($i = 1, 2, \dots, k$) で構成されるデータ集合 $X = C_1^t \cup C_2^t \cup \dots \cup C_k^t$, $C_i^t \cap C_j^t = \emptyset$ ($i \neq j, 1 \leq i, j \leq k$) を考える。式 (9) より、任意の二つのクラスタ C_i^t と C_j^t の Δ_{ij} は、残りのクラスタに所属するデータ $x \in X - C_i^t - C_j^t$ には依存しない。そのため、単純に二つのクラスタ C_i^t と C_j^t に式 (9) を適用でき、各クラスタの対に関して独立に Δ_{ij} ($1 \leq i < j \leq k$) を定義できる。したがって、 $C^t = \{C_j^t \mid j = 1, 2, \dots, k\} \in C_{All}^*$ となる必要十分条件は式 (10) となる。

$$\Delta = \bigcap_{1 \leq i < j \leq k} \Delta_{ij} \neq \emptyset. \quad (10)$$

与えられた $X = C_1^t \cup C_2^t \cup \dots \cup C_k^t$ に対して、式 (10) を満たす δ が存在しない場合は、Algorithm 1 は、どのような初期クラスタを与えても、 C^t を出力することはできない。

文献(柳, 2012)の手法は、与えられたデータ集合に対して $\Delta \neq \emptyset$ を仮定し、Algorithm 1 を再帰的に利用しながら δ の更新を行う。十分小さな δ_1 をガウス関数のパラメータの初期値として、Algorithm 1 の実行とパラメータの更

新を、後述の停止条件を満たすまで繰り返す。 $q+1$ 番目のパラメータである δ_{q+1} の更新は、式 (11) のように、単純に増加させていくという戦略を用いる。

$$\delta_{q+1} = \delta_q + \epsilon_q, (\epsilon_q \geq 0). \quad (11)$$

停止条件に関しては、まず、クラスタ C_a とクラスタ C_b の最小距離 $d_{\min}(C_a, C_b)$ を式 (12) で定義する。

$$d_{\min}(C_a, C_b) = \min_{x \in C_a, x' \in C_b} \|x - x'\|^2. \quad (12)$$

式 (12) を満たす $x \in C_a$ を x_a , $x' \in C_b$ を x_b とし、 δ の更新により、 $C_a - \{x_a\}$ と $C_b \cup \{x_a\}$ が発生すると仮定する。 x_a の移動が発生する前のクラスタ間の中心点の距離 $\|M_a - M_b\|^2$ と移動が発生した後の距離 $\|M_a^{\text{new}} - M_b^{\text{new}}\|^2$ の比率 $d_{\text{rate}} = \|M_a^{\text{new}} - M_b^{\text{new}}\|^2 / \|M_a - M_b\|^2$ が、事前に決めた下限 r_{low} と上限 r_{up} による区間 $[r_{\text{low}}, r_{\text{up}}]$ に対して、 $d_{\text{rate}} \notin [r_{\text{low}}, r_{\text{up}}]$ となったとき δ_q の更新を終了し、その時点でのクラスタリングの結果を最終的な計算結果とする。 $\|M_a - M_b\|^2$ と $\|M_a^{\text{new}} - M_b^{\text{new}}\|^2$ は式 (13) と式 (14) により計算する。

$$\begin{aligned} \|M_a - M_b\|^2 &= \|M_a\|^2 - 2M_a \cdot M_b + \|M_b\|^2 \\ &= \frac{1}{|C_a|^2} \sum_{x, x' \in C_a} K(x, x') \\ &\quad - \frac{2}{|C_a||C_b|} \sum_{x \in C_a, x' \in C_b} K(x, x') \\ &\quad + \frac{1}{|C_b|^2} \sum_{x, x' \in C_b} K(x, x'). \end{aligned} \quad (13)$$

$$\begin{aligned} \|M_a^{\text{new}} - M_b^{\text{new}}\|^2 &= \\ &\frac{|C_a|^2 \|M_a\|^2}{(|C_a| - 1)^2} + \frac{|C_b|^2 \|M_b\|^2}{(|C_b| + 1)^2} \\ &\quad - 2 \frac{|C_a||C_b|}{(|C_a| - 1)(|C_b| + 1)} (M_a \cdot M_b) \\ &\quad - 2 \frac{(|C_a| + |C_b|)}{(|C_a| - 1)^2 (|C_b| + 1)} \sum_{x \in C_a} K(x, x_a) \\ &\quad + 2 \frac{(|C_a| + |C_b|)}{(|C_a| - 1)(|C_b| + 1)^2} \sum_{x \in C_b} K(x, x_a) \\ &\quad + \left(\frac{1}{(|C_a| - 1)} + \frac{1}{(|C_b| + 1)} \right)^2 K(x_a, x_a). \end{aligned} \quad (14)$$

ただし、式 (13) と式 (14) はともに δ を用いて

計算する。

次に、更新幅に関して、式 (12) を満たす $x_a \in C_a$ と $x_b \in C_b$ に関して、 $\delta_{q+1} = \delta_q + \epsilon_q$ のパラメータ更新により、 $C_a^{(q+1)} = C_a^{(q)} - \{x_a\}$, $C_b^{(q+1)} = C_b^{(q)} \cup \{x_a\}$ の更新が発生すると仮定し、 ϵ_q を推定する。これは、以下の式 (15), (16) を満たす必要がある。

$$\|M_a^{(q)} - \phi_q(x_a)\|^2 \leq \|M_b^{(q)} - \phi_q(x_a)\|^2, \quad (15)$$

$$\|M_a^{(q+1)} - \phi_{q+1}(x_a)\|^2 \geq \|M_b^{(q+1)} - \phi_{q+1}(x_a)\|^2. \quad (16)$$

$M_a^{(q)}$ と $M_a^{(q+1)}$ の違いは、あくまで内積計算時のガウスカーネルのパラメータが、それぞれ δ_q , δ_{q+1} となることだけであり、クラスタに所属するデータは変化していない。また、 $\phi_q(x_a)$ はデータ x_a の D 次元空間上への写像であり、ガウスカーネルによる内積計算時はパラメータとして δ_q を使用することを意味する。 $A_q = \|M_a^{(q)}\|^2 + 2M_b^{(q)} \cdot \phi_q(x_a)$, $B_q = \|M_b^{(q)}\|^2 + 2M_a^{(q)} \cdot \phi_q(x_a)$ として、式 (15) と式 (16) より、 ϵ_q は式 (17) のように下から抑えることができる。

$$\epsilon_q \geq \frac{\delta_q^3}{2 \max\{d_1, d_2\}} \log \frac{B_q}{A_q}. \quad (17)$$

ここで、 d_1 はクラスタ C_a に所属するデータ間の最大距離、 d_2 はデータ x_a とクラスタ C_b に所属するデータ間の最大距離とする。

式 (17) の不等式は、かなり緩い評価であり、初期値 δ_1 をかなり小さな値に設定する場合は、 ϵ_1 も相当小さくなるため収束速度がかなり遅くなり、実用的に好ましくない。そこで、クラスタ C_a に所属するデータ間の距離の平均を d_3 、データ x_a とクラスタ C_b に所属するデータとの距離の平均を d_4 、クラスタ C_a に所属するデータ間の最短距離を d_5 、そして、データ x_a とクラスタ C_b に所属するデータとの最短距離を d_6 として d_{opt} を式 (18) のように定義する。

$$d_{\text{opt}} = \arg \max_{d_i} \{d_i \mid \epsilon_q(d_i) \leq p\delta_q, 1 \leq i \leq 6\} \quad (18)$$

式 (18) において $\epsilon_q(d_i)$ は式 (17) の右辺の $\max\{d_1, d_2\}$ を d_i で置き換えた式であり、 $p(0 < p \leq 0.5)$ は事前に設定する安全係数である。式 (17) の右辺の $\max\{d_1, d_2\}$ をパラ

メータ d_{opt} で置き換えた式 (19) により更新幅 ϵ_q を $\epsilon_q = \epsilon_q(d_i)$ として決定する。

$$\epsilon_q(d_{opt}) = \frac{\delta_q^3}{2d_{opt}} \log \frac{B_q}{A_q}. \quad (19)$$

したがって、 $\delta_{q+1} = \delta_q + \epsilon_q(d_{opt})$ により δ_{q+1} の値を更新する。

二つ以上の複数クラスタに対しては、特定の二つのクラスタに着目し δ_q の値を更新していく。二つのクラスタの選択方法に関しては、現在得られている k 個の各クラスタ C_i に関して、他のクラスタ C_j とのクラスタ中心の距離を計算し、その最小値となるクラスタ番号 $\min(i) = \arg \min_{j \neq i} \|M_i - M_j\|^2$ を求める。求めた k 個のクラスタ番号の対 i と $\min(i)$ に関して、 R^D 上でのクラスタ中心と所属するデータとの距離の平均 $d_m(C_i)$ を求め、 $d_m(C_i)$ と $d_m(C_{\min(i)})$ の比率が 1 以上となるように分母と分子を定め、比率の大きな二つのクラスタを δ_q 更新のためのクラスタとして採用する。

$$u = \arg \max_i \left\{ \frac{\max\{d_m(C_i), d_m(C_{\min(i)})\}}{\min\{d_m(C_i), d_m(C_{\min(i)})\}} \right\},$$

$$v = \min(u). \quad (20)$$

最終的に $a = \arg \max_{u,v} \{d_m(C_u), d_m(C_v)\}$ を対象クラスタの番号とする。従来法のアルゴリズムを以下に示す。

Algorithm 2 入力: X, k, δ

- 1 繰返し番号 q を 1 とし、 $\delta_q = \delta$ とする。
 - 2 入力データ $x_i \in X (i = 1, 2, \dots, n)$ をランダムに k 個のクラスタ $C_j^{(q)} \subset X (j = 1, 2, \dots, k)$ に分類する。
 - 3 $X, C_1^{(q)}, C_2^{(q)}, \dots, C_k^{(q)}, \delta_q$ を入力として **Algorithm 1** を実行し、クラスタ $C_1^{(q)}, C_2^{(q)}, \dots, C_k^{(q)}$ を再計算する。
 - 4 終了フラグが立っていれば 9 へ進み、それ以外の場合は 5 へ進む。
 - 5 式 (20) より C_a と C_b を選択する。
 - 6 $d_{rate} = \|m_a^{new} - m_b^{new}\|^2 / \|m_a - m_b\|^2$ を計算し、 $d_{rate} < r_{low}$ なら 9 へ、 $d_{rate} > r_{up}$ なら終了フラグを立てる。
 - 7 式 (19) より、 ϵ_q を計算する。
 - 8 $\delta_q = \delta_q + \epsilon_q$, $q = q + 1$ とし 3 へ戻る。
 - 9 $C_1^{(q)}, C_2^{(q)}, \dots, C_k^{(q)}$ を出力し終了する。
-

4. 提案法

従来法である文献(柳, 2012)の手法はいくつかの人工データと実データに対して、ある程度有効性を示しているものの、以下のような問題点がある。

- (1) 計算の終了が、ガウス関数のパラメータ、下限 r_{low} , 上限 r_{up} に依存しており、データによっては終了がやや不安定となる。
- (2) 反復法の更新幅の評価が緩く、安全係数 p で制御する必要がある。
- (3) 探索が正方向のみであるため、初期クラスタ依存の高いデータに対しては有効性が低い。
- (4) 探索が正方向のみであるため、適切な初期 δ を選択する必要がある。

本論文では、上記の問題点の改良を検討する。基本的には、従来法と同様の考え方により、Algorithm 1 を再帰的に利用しながら δ の更新を反復して行う手法である。従来法との大きな相違点は二点ある。一点目は、従来法が (1) パラメータの更新, (2) Algorithm 1 の実行という 2 つの計算過程であったのに対して、提案法は (1) クラスタの更新, (2) パラメータの更新, (3) Algorithm 1 の実行という 3 つの計算過程となることである。二点目は、パラメータの更新に関して、従来法が正方向のみの探索であったのに対して、提案法は正負両方向の探索を行うことである。まず、探索の終了条件として、カーネル関数の写像元の空間の距離を利用し、それに適した更新幅を検討する。次に、データの距離の分布情報を利用して探索方向を決める。二つのクラスタ C_a と C_b に対して、 $d_{\min}(C_a, C_b)$ を満たす $x \in C_a$ を $x = \text{idx}_a(C_a, C_b)$, 同様に、 $d_{\min}(C_a, C_b)$ を満たす $x \in C_b$ を $x = \text{idx}_b(C_a, C_b)$ として、 C_a の部分集合となる系列 $X_a(i), (i \leq |C_a|)$ を以下のように定義する。

$$X_a(0) = \emptyset,$$

$$X_a(1) = \{x_1 | x_1 = \text{idx}_a(C_a, C_b)\},$$

$$X_a(i) = X_a(i-1) \cup \{x_i\},$$

$$x_i = \text{idx}_a(C_a - X_a(i-1), C_b \cup X_a(i-1)). \quad (21)$$

式 (21) より $|X_a(i)| = i$ となり、 $d_{\min}(C_a, C_b)$ を満たす $x \in C_a$ を、クラスタ C_a からクラスタ C_b へと順番に変更していくことを意味する。

また、定義より、 $X_a(|C_a|) = C_a$ となる。次に、 $C_a^{\text{new}} = C_a - X_a(i), C_b^{\text{new}} = C_b \cup X_a(i), (1 \leq i \leq |C_a| - 1)$ としたときのデータの移動数 i の変化に対する最大クラスタ間距離 d_a と対応する $X_a = X_a(i)$ を以下のように定義する。

$$\begin{aligned} d_a &= \max_{0 \leq i \leq |C_a| - 1} d_{\min}(C_a - X_a(i), C_b \cup X_a(i)), \\ X_a &= X_a(t), \\ t &= \arg \max_{0 \leq i \leq |C_a| - 1} d_{\min}(C_a - X_a(i), C_b \cup X_a(i)). \end{aligned} \quad (22)$$

式(21)と同様に、 $d_{\min}(C_a, C_b)$ を満たす $x \in C_b$ を、クラスタ C_b からクラスタ C_a へと順番に変更していくときの i 回目の部分集合を $X_b(i)$ とするとき、 $C_b^{\text{new}} = C_a \cup X_b(i), C_b^{\text{new}} = C_b - X_b(i), (1 \leq i \leq |C_b| - 1)$ におけるデータの移動数 i の変化に対する最大クラスタ間距離 d_b と対応する $X_b = X_b(i)$ は式(23)となる。

$$\begin{aligned} d_b &= \max_{0 \leq i \leq |C_b| - 1} d(C_a \cup X_b(i), C_b - X_b(i)), \\ X_b &= X_b(t), \\ t &= \arg \max_{0 \leq i \leq |C_b| - 1} d(C_a \cup X_b(i), C_b - X_b(i)). \end{aligned} \quad (23)$$

式(22)と式(23)より、 d_a と d_b を比較し、大きい方に対応する C_a^{new} と C_b^{new} を新しいクラスタとする。式(22)と式(23)において、 $t=0$ のときは $C_a^{\text{new}} = C_a, C_b^{\text{new}} = C_b$ となる。これ以降、煩雑さを避けるため、一般性を失わず $d_a > d_b$ と仮定して説明を行う。したがって、 $C_a^{\text{new}} = C_a - X_a, C_b^{\text{new}} = C_b \cup X_a$ とし、 C_a^{new} と C_b^{new} に分類可能である δ を求める。

4-1 パラメータの更新比率

前節より、 $x_a = \text{id}x_a(C_a^{\text{new}}, C_b^{\text{new}}), x_b = \text{id}x_b(C_a^{\text{new}}, C_b^{\text{new}})$ とし、後の計算の簡易さを考慮して δ^{new} を以下の式(24)で定義する。

$$\delta^{\text{new}} = \rho \delta. \quad (24)$$

式(24)は、 $\delta^{\text{new}} = \rho \delta = \delta + \Delta \delta$ より、 $(\rho - 1)\delta = \Delta \delta$ とすることで、式(11)と等価となる。 δ を増加させていく場合は $\rho \geq 1$ となり、減少させていく場合は $\rho < 1$ となる。また、逐次的な計算により δ を求めていくため、以下のように q 回目の計算結果となる δ_q と ρ_q を用いて $q+1$ 回目の δ_{q+1} を求める。

$$\delta_{q+1} = \rho_q \delta_q. \quad (25)$$

以降は、簡単さを考慮して式(24)を用いて $\rho \geq 1$ について説明を行う。式(22)と式(23)より、以下の条件が成立している。

$$\begin{aligned} C_a \cup C_b &= X, C_a \cap C_b = \emptyset, \\ C_a^{\text{new}} \cup C_b^{\text{new}} &= X, C_a^{\text{new}} \cap C_b^{\text{new}} = \emptyset. \end{aligned} \quad (26)$$

このとき、クラスタ C_a^{new} と C_b^{new} の各中心、および、データ x_a, x_b との距離に関して以下の式が成立するように δ^{new} を決定する。

$$\begin{aligned} \|M_a^{\text{new}} - \phi(x_a)\|^2 &\leq \|M_b^{\text{new}} - \phi(x_a)\|^2, \\ \|M_a^{\text{new}} - \phi(x_b)\|^2 &\geq \|M_b^{\text{new}} - \phi(x_b)\|^2. \end{aligned} \quad (28)$$

ここで、式(28)と式(29)における M_a^{new} と M_b^{new} は、それぞれ、 C_a^{new} と C_b^{new} のクラスタ中心である。式(28)と式(29)は、 a と b を置き換えると同じ形になっているため、一般性を失わず、式(28)について解析する。まず、式(28)を式(30)のように変形する。

$$\begin{aligned} \|M_a^{\text{new}}\|^2 + 2M_b^{\text{new}}\phi(x_a) &\leq \\ \|M_b^{\text{new}}\|^2 + 2M_a^{\text{new}}\phi(x_a). \end{aligned} \quad (30)$$

次に、式(30)を満たす ρ の条件を考える。

任意の x, x' に関してパラメータを δ とするカーネル関数を $K(x, x')$ とし、パラメータを δ^{new} とするカーネル関数を $K^{\text{new}}(x, x')$ とする。このとき、 $K^{\text{new}}(x, x')$ と $K(x, x')$ の差分 $\Delta K(x, x')$ は式(31)となる。

$$\begin{aligned} \Delta K(x, x') &= K^{\text{new}}(x, x') - K(x, x') \\ &= \exp\left(-\frac{\|x - x'\|^2}{\delta^{\text{new}2}}\right) - \exp\left(-\frac{\|x - x'\|^2}{\delta^2}\right) \\ &= \exp\left(-\frac{\|x - x'\|^2}{(\rho\delta)^2}\right) - \exp\left(-\frac{\|x - x'\|^2}{\delta^2}\right). \end{aligned} \quad (31)$$

式(31)の最大値を考えるために、 $s > t > 0$ となる定数 s と t を用いて式(32)の関数 $f(x)$ を定義する。

$$f(x) = e^{-\frac{x}{s}} - e^{-\frac{x}{t}} \quad (32)$$

式(32)の導関数 $f'(x)$ は式(33)となる。

$$f'(x) = \frac{df(x)}{dx} = -\frac{1}{s}e^{-\frac{x}{s}} + \frac{1}{t}e^{-\frac{x}{t}} \quad (33)$$

式 (33) より, $f'(x_0) = 0$ を満たす x_0 は式 (34) となる。

$$x_0 = \frac{st}{s-t} \log \frac{s}{t} \quad (34)$$

$s > t > 0$ の仮定より, $x_0 > 0, f'(0) = \frac{s-t}{st} > 0$ となるため, 関数 $f(x)$ は $x = x_0$ で式 (35) の最大値を持つ。

$$\max_x f(x) = f(x_0) = \left(\frac{s}{t}\right)^{-\frac{t}{s-t}} - \left(\frac{s}{t}\right)^{-\frac{s}{s-t}} \quad (35)$$

例として, $s = 0.5, t = 0.2$ の関数 $f(x)$ の $0 < x \leq 2$ のグラフの概形を図 1 に示す。

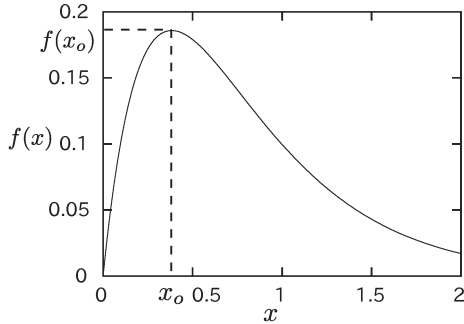


図 1 $s = 0.5, t = 0.2$ の関数 $f(x)$ のグラフの概形

以上より, まず, 式 (30) の右辺を上から抑えることを考える。 $K(x, x) = K^{\text{new}}(x, x) = 1$ より, $\|m_b^{\text{new}}\|^2$ は式 (36) を満たす。

$$\begin{aligned} \|M_b^{\text{new}}\|^2 &= \frac{1}{|C_b^{\text{new}}|^2} \sum_{x, x' \in C_b^{\text{new}}} K^{\text{new}}(x, x') \\ &= \frac{1}{|C_b^{\text{new}}|^2} \sum_{x, x' \in C_b^{\text{new}}} K(x, x') \\ &\quad + \frac{1}{|C_b^{\text{new}}|^2} \sum_{x, x' \in C_b^{\text{new}}, x \neq x'} \Delta K(x, x') \\ &\leq \frac{1}{|C_b^{\text{new}}|^2} \sum_{x, x' \in C_b^{\text{new}}} K(x, x') \\ &\quad + \frac{(|C_b^{\text{new}}|^2 - |C_b^{\text{new}}|)}{|C_b^{\text{new}}|^2} \Delta K_{\max} \\ &= \|M_b^{(\text{new})}\|^2 + \frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} \Delta K_{\max}. \end{aligned} \quad (36)$$

ここで, 式 (36) における $\|M_b^{(\text{new})}\|^2$ は, カーネル関数の計算時にパラメータとして δ^{new} ではなく δ を用いることを意味する。また, ΔK_{\max} は, 任意の x, x' における $\Delta K(x, x')$ の最大値とする。 δ の増加比率 ρ に関しては, $\rho \geq 1$ を仮定しているため, $s = \delta^{\text{new}} = (\rho\delta)^2, t = \delta^2$ とすると, $s/t = \rho^2 > 1, s/(s-t) = \rho^2/(\rho^2-1) >$

$1, t/(s-t) = 1/(\rho^2-1)$ となる。従って, 式 (32), 式 (35) より, L_2 ノルム $\|x - x'\|^2$ を変数とする $f(\|x - x'\|^2) = \Delta K(x, x')$ の最大値 ΔK_{\max} は式 (37) を満たす。

$$\begin{aligned} \Delta K_{\max} &= (\rho^2)^{-\frac{1}{\rho^2-1}} - (\rho^2)^{-\frac{\rho^2}{\rho^2-1}} \\ &= (\rho^2)^{-\frac{\rho^2}{\rho^2-1}} (\rho^2 - 1) \\ &\leq (\rho^2)^{-1} (\rho^2 - 1) = \left(1 - \frac{1}{\rho^2}\right). \end{aligned} \quad (37)$$

式 (36) と同様の手順により, 式 (30) の右辺の $2M_a^{\text{new}}\phi(x_a)$ は式 (38) の不等式を満たす。

$$\begin{aligned} 2M_a^{\text{new}}\phi(x_a) &\leq 2M_a^{(\text{new})}\phi(x_a) \\ &\quad + 2\frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|} \Delta K_{\max}. \end{aligned} \quad (38)$$

式 (36) と式 (38) より, 式 (30) の右辺は式 (39) の不等式で上から抑えることができる。

$$\begin{aligned} \|M_b^{\text{new}}\|^2 + 2M_a^{\text{new}}\phi(x_a) &\leq \|M_b^{(\text{new})}\|^2 + 2M_a^{(\text{new})}\phi(x_a) \\ &\quad + \left(\frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} + 2\frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|}\right) \Delta K_{\max}. \end{aligned} \quad (39)$$

次に式 (30) の左辺を下から抑えることを考える。式 (36), 式 (38), 式 (39) と同様の手順により, 式 (30) の左辺は式 (40) となる。

$$\begin{aligned} \|M_a^{(\text{new})}\|^2 + 2M_b^{(\text{new})}\phi(x_a) &+ \frac{1}{|C_a^{\text{new}}|^2} \sum_{x, x' \in C_a^{\text{new}}, x \neq x'} \Delta K(x, x') \\ &+ 2\frac{1}{|C_b^{\text{new}}|} \sum_{x \in C_b^{\text{new}}} \Delta K(x, x_a) \\ &= \|M_a^{\text{new}}\|^2 + 2M_b^{\text{new}}\phi(x_a). \end{aligned} \quad (40)$$

ここで, ϵ を式 (41) のように定義する。

$$\begin{aligned} \epsilon &= \frac{1}{|C_a^{\text{new}}|^2} \sum_{x, x' \in C_b^{\text{new}}, x \neq x'} \Delta K(x, x') \\ &\quad + 2\frac{1}{|C_b^{\text{new}}|} \sum_{x \in C_b^{\text{new}}} \Delta K(x, x_a). \end{aligned} \quad (41)$$

$\epsilon \ll \left(\frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} + 2\frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|}\right) \Delta K_{\max}$ と仮定し, 式 (40) を式 (42) のように下から抑える。

$$\begin{aligned} \|M_a^{(\text{new})}\|^2 + 2M_b^{(\text{new})}\phi(x_a) &\leq \|M_a^{\text{new}}\|^2 + 2M_b^{\text{new}}\phi(x_a). \end{aligned} \quad (42)$$

式(39), 式(42)より, 式(30)の不等式を上下から抑え, 式(43)の不等式を得る。

$$\begin{aligned}
& \|M_a^{(\text{new})}\|^2 + 2M_b^{(\text{new})}\phi(x_a) \\
& \leq \|M_a^{\text{new}}\|^2 + 2M_b^{\text{new}}\phi(x_a) \\
& \leq \|M_b^{\text{new}}\|^2 + 2M_a^{\text{new}}\phi(x_a) \\
& \leq \|M_b^{(\text{new})}\|^2 + 2M_a^{(\text{new})}\phi(x_a) + \\
& \left(\frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} + 2 \frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|} \right) \Delta K_{\max}.
\end{aligned} \tag{43}$$

式(43)のカーネル関数の計算の際に用いるパラメータは, 全て $\delta^{\text{new}} = \rho\delta$ ではなく δ とできる。 δ を用いた D 次元空間上でのクラスタ $C \subseteq X$ の中心 M と $x \in X$ の L_2 ノルムを $d(M, x) = \|M - \phi(x)\|^2$ とおくと, 式(37), 式(43)より, 式(44)のように ρ の下限を得る。

$$\begin{aligned}
\Delta K_{\max} & \geq \frac{d(M_a^{(\text{new})}, x_a) - d(M_b^{(\text{new})}, x_a)}{\left(\frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} + 2 \frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|} \right)} \\
1 - \frac{1}{\rho^2} & \geq \frac{d(M_a^{(\text{new})}, x_a) - d(M_b^{(\text{new})}, x_a)}{\left(\frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} + 2 \frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|} \right)}, \\
\therefore \rho & \geq \sqrt{\frac{1}{\left(1 - \frac{d(M_a^{(\text{new})}, x_a) - d(M_b^{(\text{new})}, x_a)}{\left(\frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} + 2 \frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|} \right)} \right)}}.
\end{aligned} \tag{44}$$

式(44)の右辺を用いて $\delta^{\text{new}} = \rho\delta$ と更新していくが, ρ の下限を用いて更新するため, 実際には, 一度で適切な δ^{new} が求まるとは限らず, 式(28)と式(29)を満たす δ_q が求まるまで, 逐次的に $\delta_q = \rho_q \rho_{q-1} \dots \rho_1 \delta$ と計算していくことになる。 q 回目の計算過程において, 式(44)の $x_a \in C_a^{\text{new}}$ は, d 次元空間上で $d_{\min}(C_a^{\text{new}}, C_b^{\text{new}})$ を満たすデータではあるものの, D 次元空間上では, $\|M_a^{(\text{new})} - \phi(x)\|^2 \geq \|M_b^{(\text{new})} - \phi(x)\|^2$ を満たす $\exists x \in X_a$ が存在する可能性がある。その場合 δ_q をパラメータとするカーネル関数を用いた計算では, $x \notin C_a^{\text{new}}$ となる。そこで, q 回目の各計算段階において, 式(45)のように x_a と x_b を求める。

$$\begin{aligned}
x_a & = \min_{x \in C_a^{\text{new}}} \|M_b^{(\text{new})} - \phi(x)\|, \\
x_b & = \min_{x \in C_b^{\text{new}}} \|M_a^{(\text{new})} - \phi(x)\|.
\end{aligned} \tag{45}$$

4-2 探索方向

$\delta_{q+1} < \delta_q$ となる負方向への探索は $\rho < 1$ となるようにする。これは, $\rho \geq 1$ のときは式(30)の右辺の増加が左辺の増加を上回る仮定で定式化を行ったのに対して, 式(30)の左辺の減少が右辺の減少を下回る仮定で定式化を行う。すなわち, 式(30)の右辺の減少分である $\Delta K(x, x')$ の総和の絶対値が式(30)の左辺の $\Delta K(x, x')$ の総和の絶対値に比して無視できるくらい小さいと仮定する。式(39), 式(40), 式(41)と同様の考え方により, 式(30)の左辺の減少分を, 任意の x, x' における $\Delta K(x, x')$ の最小値 ΔK_{\min} を用いて表し, 式(46)を得る。

$$\begin{aligned}
& \|M_a^{(\text{new})}\|^2 + 2M_b^{(\text{new})}\phi(x_a) + \\
& \left(\frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|} + 2 \frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} \right) \Delta K_{\min} \\
& \leq \|M_b^{(\text{new})}\|^2 + 2M_a^{(\text{new})}\phi(x_a).
\end{aligned} \tag{46}$$

ΔK_{\min} の増減の変化に関しては, 式(32)の関数 $f(x)$ において, $t > s > 0$ とした場合が対応しており, そのグラフは図1の横軸に対して反転した概形となる。したがって, $\rho < 1$ より, 式(35)の $f(x)$ の極大値を, そのまま, 極小値として利用でき, $\Delta K_{\min} \geq (1 - 1/\rho^2)$ より, ρ の下限は式(47)となる。

$$\rho \geq \sqrt{\frac{1}{\left(1 + \frac{d(M_a^{(\text{new})}, x_a) - d(M_b^{(\text{new})}, x_a)}{\left(\frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|} + 2 \frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} \right)} \right)}}. \tag{47}$$

正方向と負方向の切り替えには, データ X の二点間の距離の分布情報を利用する。以下, 簡単のため, $\rho \geq 1$ の場合で説明するが, $\rho < 1$ の場合も同様の考え方である。まず, 式(44)より, ρ の探索区間を検討する。 ρ の下限に関しては, 式(44)において, 仮に $d(M_a^{(\text{new})}, x_a) - d(M_b^{(\text{new})}, x_a) < 0$ を満たした場合, すでに現在の δ が式(30)を満たしているため, これ以上 δ を更新する必要はない。したがって, 式(30)を δ 更新の終了条件の一つとすることで, $\rho \geq 1$ を強制できる。次に, 任意の C_a^{new} と C_b^{new} に対して, 必ずしも式(30)を満たす δ が存在するとは限らず, その場合 $\lim_{\delta \rightarrow \infty} \rho = 1 + \epsilon (0 < \epsilon \ll 1)$ となり, 式(30)の条件では計算が終了しないことになる。そこで, $\rho = 1 + \epsilon (0 < \epsilon \ll 1)$ を二つ目の終了条件とする。最後に, ρ の上限に関して,

$0 < \|M\|^2 < 1, 0 < M \cdot \phi(x) < 1$ より, 式 (44) において, $d(M_a^{(\text{new})}, x_a) - d(M_b^{(\text{new})}, x_a) < 3$ となり, さらに, 十分大きな C_a^{new} と C_b^{new} に対して, $\left(\frac{(|C_b^{\text{new}}|-1)}{|C_b^{\text{new}}|} + 2\frac{(|C_a^{\text{new}}|-1)}{|C_a^{\text{new}}|}\right) \approx 3$ となるため, 上記二つの終了条件を満たさなかった場合, ρ がかなり大きくなる可能性がある。そこで, $K(x, x')$ の特性に着目し, $\|x - x'\|^2, x, x' \in X$ の分布を考慮した制約を設ける。図 1 より, $f(x)$ のグラフを $\Delta K(x, x')$ に置き換えて考えると, 横軸が L_2 ノルム $\|x - x'\|^2, x, x' \in X$ を表す変数となるため, 現在の δ_q に関して, 極大となる点 x_0 の近辺に多くの $\|x - x'\|^2, x, x' \in X$ が分布しているとき, ΔK_{\max} と近い値となる $\Delta K(x, x')$ の数が増加すると考えられる。また, $\|x - x'\|^2, x, x' \in X$ の分布が x_0 から正の方向に離れている場合は, δ の増加に対して ΔK_{\max} と近い値となる $\Delta K(x, x')$ の数の増加は見込めないと考えられる。上記より, まず, クラスタ C の任意の二点間の距離の集合を D_C , データ x とクラスタ C の任意の点との距離の集合を $D_{C,x}$, 集合 A の最小値を $\min(A)$, 中央値を $\text{med}(A)$, 最大値を $\max(A)$ とし, 式 (30) の右辺に関わる距離の分布における最大値, 最小値, および, 中央値を式 (48) で定義する。

$$\begin{aligned} d_{\max} &= \min\{\max(D_{C_b}), \max(D_{C_a, x_a})\}, \\ d_{\min} &= \begin{cases} \min(D_{C_b}) & (d_{\max} = \max(D_{C_b})) \\ \min(D_{C_a, x_a}) & (\text{otherwise}) \end{cases}, \\ d_{\text{med}} &= \begin{cases} \text{med}(D_{C_b}) & (d_{\max} = \max(D_{C_b})) \\ \text{med}(D_{C_a, x_a}) & (\text{otherwise}) \end{cases}. \end{aligned} \quad (48)$$

三つ目の終了条件として, δ の探索区間を式 (49) とする。

$$d_{\min} \leq \delta \leq d_{\max}. \quad (49)$$

次に方向の切り替えは, $\Delta K(x, x')$ の極大点, つまり, 式 (34) を用いて, $x_0 \leq d_{\text{med}}$ なら正方向, $x_0 > d_{\text{med}}$ なら負方向の探索を行う。式 (34) において $s = \rho\delta, t = \delta$ とすると, 極大点は $x_0 = \frac{\rho \log \rho}{\rho - 1} \delta$ となり, ρ が定まらなければ基準 x_0 を決定できないため, ρ が +1 から僅かに正方向, あるいは, 負方向に変化した場合で考える。この場合, $\lim_{n \rightarrow \pm\infty} (1 + 1/n)^n = e$ より, $\lim_{\rho \rightarrow 1+0} x_0 = \lim_{\rho \rightarrow 1-0} x_0 = \delta$ となる。ただし, q 回目の計算過程毎に方向を変更すると, 収束しない可能性があるため, 1 回のクラ

スタの変更では探索方向の変更は行わないこととする。以上より, δ を更新するためのアルゴリズムを Algorithm 3 に示す。

Algorithm 3 入力: C_a, C_b, δ

- 1 式 (48) より, $d_{\max}, d_{\min}, d_{\text{med}}$ を求める。
 - 2 $\delta \leq d_{\text{med}}$ なら, 式 (44), そうでなければ, 式 (47) を ρ の計算方法とする。
 - 3 繰返し番号を $q = 1$ とし, $\delta_q = \delta$ とする。
 - 4 式 (45) より x_a と x_b を求める。
 - 5 C_a, C_b, x_a, δ_q より 2 で ρ_q を計算する。
 - 6 以下のいずれかであれば, δ_q を出力し計算を終了する。
 - 式 (28) と式 (29) を同時に満たす
 - $\rho_q = 1$
 - 式 (49) を満たしていない
 - 7 $\delta_q = \rho_q \delta_q, q = q + 1$ とし, 4 へ戻る。
-

Algorithm 3 の 6 行目の最後の停止条件において, 実際には, 初期 δ が $\delta < d_{\min}$ となる場合もあるため, δ の正方向への探索の終了条件は $\delta \leq d_{\max}$ を満たさないことのみとする。同様に, 初期 δ が $d_{\max} < \delta$ となる場合もあるため, δ の負方向への探索の終了条件は $d_{\min} \leq \delta$ を満たさないことのみとする。初期 δ が $\delta < d_{\min}$ となる場合は強制的に $\delta = d_{\min}$ とし, 初期 δ が $\delta > d_{\max}$ となる場合は強制的に $\delta = d_{\max}$ とする方法も有力であるが, $\exists \delta \in \Delta, \delta < d_{\min}, \delta > d_{\max}$ となるデータも存在する可能性を考慮し, 採用しなかった。

4-3 最近傍点情報の利用

本論文では, 分類を期待するクラスタ同士が十分離れていると仮定しているが, 実際には, クラスタ同士が分離していないデータに関しても, クラスタリングを必要とする場合がある。また, 各クラスタにおけるクラスタ内のデータの分布が, 全てのクラスタでほぼ同じになるとも限らない。上記のような場合, d 次元空間上の L_2 ノルムのみをクラスタ分離の基準に使用すると, うまく分類できない可能性がある。そこで, データ x との距離が近い上位 K 個のデータからなる K -近傍点を $\text{nn}(K, x)$ と定義し, 式 (22) と式 (23) におけるクラスタ間の移動データ X_a と X_b の決定条件を, K -近傍点の条件を付加した式 (50) のように変更する。

$$\begin{aligned}
d_a &= \max_{0 \leq i \leq |C_a| - 1} d_{\min}(C_a - X_a(i), C_b \cup X_a(i)), \\
X_a &= X_a(t), \\
t &= \arg \max_{0 \leq i \leq |C_a| - 1} d_{\min}(C_a - X_a(i), C_b \cup X_a(i)), \\
\forall x \in \text{nn}(K, \text{idx}_b(C_a - X_a(t), C_b \cup X_a(t))), \\
x &\notin C_a - X_a(t), \\
d_b &= \max_{0 \leq i \leq |C_b| - 1} d_{\min}(C_a \cup X_b(i), C_b - X_b(i)), \\
X_b &= X_b(t), \\
t &= \arg \max_{0 \leq i \leq |C_b| - 1} d_{\min}(C_a \cup X_b(i), C_b - X_b(i)), \\
\forall x \in \text{nn}(K, \text{idx}_b(C_a \cup X_b(t), C_b - X_b(t))), \\
x &\notin C_b - X_b(t). \tag{50}
\end{aligned}$$

式 (50) は、式 (22) と式 (23) の条件に加えて、 $\text{idx}_a(C_a^{\text{new}}, C_b^{\text{new}})$ と $\text{idx}_b(C_a^{\text{new}}, C_b^{\text{new}})$ の K -近傍点がすべて他方のクラスタには含まれないという条件を付加することを意味する。パラメータを $K = 0$ としたときは、式 (50) ではなく、式 (22) と式 (23) を用いて C_a^{new} と C_b^{new} を決定することと等価となる。

4-4 複数クラスタへの拡張とアルゴリズム

二つ以上の複数クラスタに対しては、全ての2対のクラスタに関して、クラスタ間距離が最小となる順に、式 (50) を評価して $C_a^{\text{new}}, C_b^{\text{new}}$ への更新を行っていく。1対の更新後は、影響のあるクラスタを含む対のみ、クラスタ間距離の更新を行い、式 (50) を評価する順序を並べ換える。全ての評価を終えたあと、最終的なクラスタ間距離に対して降順に $k(k-1)/2$ 回 Algorithm 3 を実行し δ の更新を行う。上記のクラスタ更新と δ の更新を、クラスタ更新が発生しなくなるまで繰り返す。クラスタ更新アルゴリズムを Algorithm 4 に示す。

Algorithm 4 入力: C_1, C_2, \dots, C_k, K

- 1 $L = \emptyset$ とする。
 - 2 $\min_{(i,j) \notin L} \{d_{\min}(C_i, C_j)\}$ を満たす i と j を求める。
 - 3 式 (50) より、 $d_i \geq d_j$ なら $a = i, b = j$ とし、そうでなければ $a = j, b = i$ とする。また、 X_a も求める。
 - 4 $C_a = C_a - X_a, C_b = C_b \cup X_a$ とする。
 - 5 $L = L \cup \{(a, b)\}$ とする。
 - 6 $|L| = k(k-1)/2$ なら L を出力して終了、そうでなければ、2 へ戻る。
-

Algorithm 4 では、クラスタの更新毎に δ の更新を行う方法も有力と考えられるが、計算時間の増加と δ の収束への悪影響の懸念から採用しなかった。最終的な提案法のアルゴリズムを Algorithm 5 に示す。

Algorithm 5 入力: X, k, δ, K

- 1 初期目的関数値を $f_0 = 0$ とする。
 - 2 繰返し番号を $l = 1$ とし、 $\delta_l = \delta$ とする。
 - 3 対象データ $x_i \in X (i = 1, 2, \dots, n)$ をランダムにクラスタ $C_j^{(l)} \subset X (j = 1, 2, \dots, k)$ に分類し初期クラスタとする。
 - 4 $X, k, C_1^{(l)}, C_2^{(l)}, \dots, C_k^{(l)}, \delta_l$ を入力として Algorithm 1 を実行し、クラスタ $C_1^{(l)}, C_2^{(l)}, \dots, C_k^{(l)}$ を再計算する。
 - 5 式 (5) より、目的関数値 f_l を計算する。
 - 6 $f_l = f_{l-1}$ なら、 $C_1^{(l)}, C_2^{(l)}, \dots, C_k^{(l)}$ を出力して計算を終了する。
 - 7 $C_1^{(l)}, C_2^{(l)}, \dots, C_k^{(l)}, K$ を入力として、Algorithm 4 を実行し、二対のクラスタの番号のリスト L を得る。
 - 8 $\max_{(u,v) \in L} \{d_{\min}(C_u^{(l)}, C_v^{(l)})\}$ を満たす u と v を求める。
 - 9 $C_u^{(l)}, C_v^{(l)}, \delta_l$ を入力として Algorithm 3 を実行し、 δ_l を再計算する。
 - 10 $L = L - \{(u, v)\}$ とし、 $L \neq \emptyset$ なら、8 へ戻る。
 - 11 $l = l + 1$ とし、4 へ戻る。
-

4-5 計算量

Algorithm 1 のカーネル k -平均法の計算量のオーダーは Algorithm 1 の繰返し数 t 、クラスタ数 k 、データ数 n 、データの次元数 d を用いて $\mathcal{O}(tkn^3d)$ と表すことができる。従来法の計算量のオーダーは Algorithm 2 の繰返し数 q を q_0 とすると $\mathcal{O}(q_0tkn^3d)$ となる (柳, 2012)。一方、提案法の計算量のオーダーに関しては、Algorithm 3 の繰返し数 q を q_p とすると、そのオーダーは $\mathcal{O}(q_p n^2 d)$ となり、Algorithm 5 の繰返し数 l に対して $lk(k-1)/2$ 回の呼び出しが発生する。Algorithm 4 のオーダーは $\mathcal{O}(k^2 n^2 d(1 + ndK \log K))$ で、Algorithm 5 の中で l 回の呼び出しが発生する。したがって、Algorithm 5 のオーダーは $\mathcal{O}(lk^2 q_p n^2 d + lk^2 n^2 d(1 + ndK \log K) + ltkn^3 d)$ となり、 $q_0 \gg l$ でない限り、クラスタ数 k の増加に伴い従来

法に比して提案法の計算量が増加していくと考えられる。

5. 実験

実験は、教師ラベル付きの人工データと実データに対して行う。実験環境は、Intel Xeon 2.00 GHz 4 コア × 2 のデュアルプロセッサ、および、24GB RAM のワークステーションで、OS は 64 ビット版 Linux を使用した。また、コンパイラとして gcc 4.6.3 を使用して C++ により Algorithm 1 ~ 5 を実装した。

5-1 クラスタリング結果の評価方法

教師ラベル付きのデータのクラスタリング結果に対して評価を行う代表的な指標として、Adjusted Rand Index(ARI)(Hubert and Arabie, 1985) や Normalized Mutual Information(NMI)(Strehl and Ghosh, 2003) がある。本論文では NMI を評価指標として用いる。あらかじめ教師ラベルにより決定されている k 個のクラスタ集合を $C^t = \{C_1^t, C_2^t, \dots, C_k^t\}$ 、実際にクラスタリングすることにより得られる k 個のクラスタ集合を $C = \{C_1, C_2, \dots, C_k\}$ 、そして、 $n = |C_1^t \cup C_2^t \cup \dots \cup C_k^t| = |C_1 \cup C_2 \cup \dots \cup C_k|$ をデータ数とするとき、 C^t と C のエントロピーは式 (51) で定義される。

$$H(C^t) = - \sum_{j=1}^k \frac{|C_j^t|}{n} \log \frac{|C_j^t|}{n},$$

$$H(C) = - \sum_{i=1}^k \frac{|C_i|}{n} \log \frac{|C_i|}{n}. \quad (51)$$

以降、本論文では C^t を正解クラスタ集合と呼ぶことにする。また、 C^t と C の結合エントロピーは式 (52) で定義される。

$$H(C^t, C) = - \sum_{i=1}^k \sum_{j=1}^k \frac{|C_i \cap C_j^t|}{n} \log \frac{|C_i \cap C_j^t|}{n}. \quad (52)$$

NMI の計算式は式 (51) と式 (52) を用いて、式 (53) により定義される。

$$\text{NMI} = \frac{H(C^t) + H(C) - H(C^t, C)}{\sqrt{H(C^t)H(C)}}. \quad (53)$$

NMI は 0 と 1 の間の値をとり、正解クラスタと実際に得られたクラスタが完全に一致するとき、 $H(C^t) = H(C) = H(C^t, C)$ より、最大値の 1 をとる。

5-2 実験条件

まず、Algorithm 2 の停止基準のパラメータ r_{low} と r_{up} に関しては、文献(柳, 2012) 同様、それぞれ $r_{low} = 0.8$, $r_{up} = 1.2$ とした。また、式 (18) において、 $d_1 \sim d_6$ のいずれかを選択するための安全係数 p に関しては、文献(柳, 2012) の結果より、0.5 とした。Algorithm 2 の最大繰返し回数を 100 回とし、100 回の繰返しを越えても計算を終了しない場合は強制的に計算を終了するようにした。さらに、Algorithm 2 は探索が正方向のみであるため、適切な初期 δ を選択する必要がある。しかし、計算環境には限界があるため、 δ の値が小さすぎると、ガウス関数の値が計算可能な有効桁以下となり距離計算が不能となる(柳, 2012)。そこで、文献(柳, 2012) のような統計的な情報からではなく、計算環境に依存した適切な初期値 δ_{min} を検討する。使用コンパイラにおける倍精度浮動小数型変数の正の最小値は 2.25074×10^{-308} であるため、計算対象データのデータ間の最大距離を d_{max} とし、 δ_{min} を式 (54) で定める。

$$2.25074 \times 10^{-308} < e \times 10^{-308} = e^{-\frac{d_{max}^2}{\delta_{min}^2}},$$

$$\therefore \delta_{min} = \frac{d_{max}}{\sqrt{308 \log_e 10 - 1}}. \quad (54)$$

次に、Algorithm 5 の入力 K に関しては $K = 0 \sim 4$ と変化させた。Algorithm 5 の最大繰返し回数を 20 回とし、20 回の繰返しを越えても計算を終了しない場合は強制的に計算を終了するようにした。

5-3 使用データの詳細

人工データは全て 2 次元とし、適切な δ を与えたとき、カーネル k -平均法により、正確に分類できるデータを用意した。 $\Delta \neq \emptyset$ の判定は、文献(柳, 2012) と同様、Algorithm 1 の入力を $(X, k, \delta_q, C_j^t (j = 1, 2, \dots, k))$ として Q 回 ($1 \leq q \leq Q$) の試行を行う。 δ_Q をデータ間の最大距離の 2 倍、 $\delta_1 = 0.01$, $Q = 200$ とし、各データの試行を行った結果、 $\delta_Q = 2$ までの試行で $\Delta \neq \emptyset$ の推定が可能であると判断できたため、 $\delta_1 = 0.01, \delta_Q = 2.001, Q = 200$ とした。表 1 に各データの詳細を示す。表 1 の最終列は、前節で述べた実験で使用する計算不能とならない初期 δ の値であり、その前列が式 (54) で初期 δ の計算に必要な最大距離である。図 2 ~ 15 に各データの分布状態と対

表1 人工データの詳細

データ名	図番号	サンプル数	クラス数	NMIが最大となる δ の区間の推定	距離の最大値	初期 δ
データ1	図2	400	2	図9より0.47 ~ 1.26	1.984	0.075
データ2	図3	400	2	図10より0.16 ~ 0.54	4.396	0.165
データ3	図4	601	2	図11より0.09 ~ 0.15	2.445	0.092
データ4	図5	600	3	図12より0.16 ~ 0.54	4.396	0.165
データ5	図6	600	3	図13より0.47 ~ 0.78	3.583	0.135
データ6	図7	300	3	図14より0.20 ~ 0.40	2.034	0.076
データ7	図8	800	4	図15より0.47 ~ 0.54	4.396	0.165

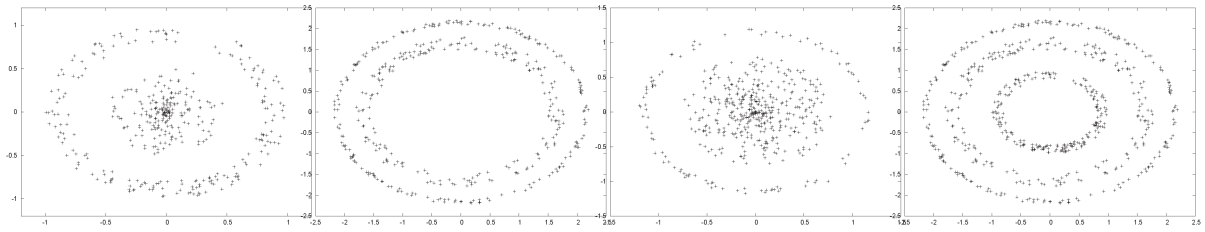


図2 データ1

図3 データ2

図4 データ3

図5 データ4

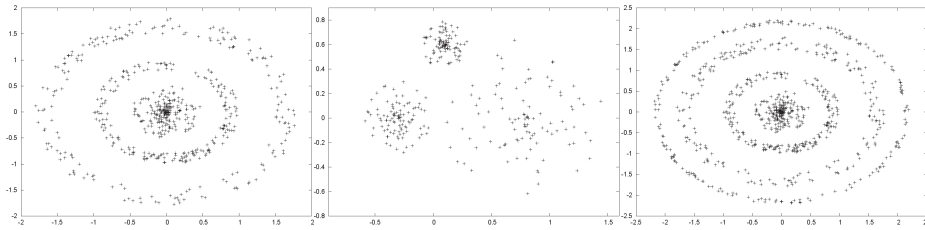
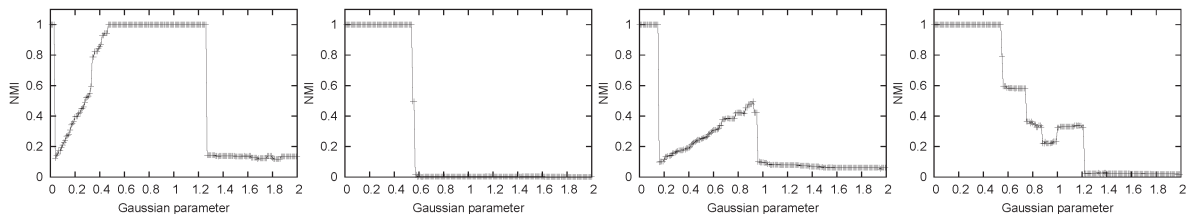
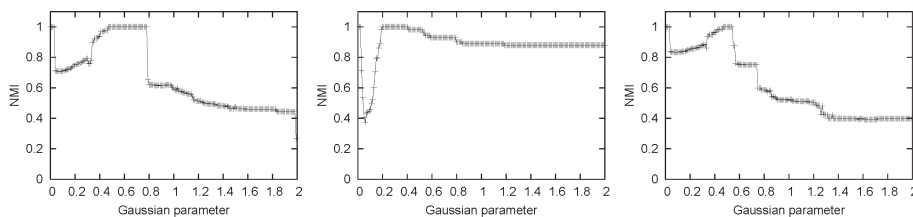


図6 データ5

図7 データ6

図8 データ7

図9 1の δ 対NMI図10 2の δ 対NMI図11 3の δ 対NMI図12 4の δ 対NMI図13 5の δ 対NMI図14 6の δ 対NMI図15 7の δ 対NMI

応する δ 対NMI値のグラフを示す。

実データに関してはUCI machine learning repository(Frank and Asuncion, 2010)より, iris, wine, satimage, pen digitの4つのデータを使用した。satimageとpen digitに関しては, 元のデータのデータ数が非常に多いため, 各クラスからランダムに80個を選択した。各データの詳細を表2に示す。表2の各データにおけるNMIが最大となる δ の区間は,

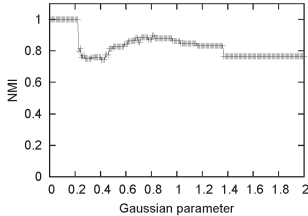
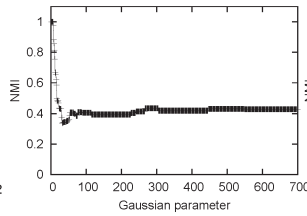
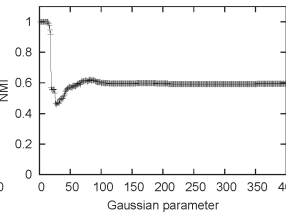
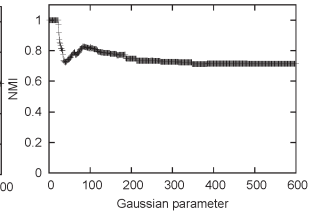
前述の人工データと同じ方法で推定した。結果を図16~19に示す。図16~19より, 全ての実データは $\Delta = \emptyset$ と推定できる。

5-4 従来法と提案法の比較

従来法であるAlgorithm 2と提案法であるAlgorithm 5を用いて, 各データに対して50回の試行を行い, δ の更新回数に関する平均と最小と最大, δ の更新毎のAlgorithm 2にお

表 2 実データの詳細

データ名	次元数	サンプル数	クラス数	NMIが最大となる δ の区間の推定	距離の最大値	初期 δ
iris	4	150	3	図16より0.8 ~ 0.82	7.09	0.27
wine	13	178	3	図17より265 ~ 285	1402.1	52.7
satimage	36	480	6	図18より81 ~ 83	384.2	14.4
pen digit	16	800	10	図19より84 ~ 86	286.3	10.8

図 16 iris の δ 対 NMI図 17 wine の δ 対 NMI図 18 sat. の δ 対 NMI図 19 pen digit の δ 対 NMI

ける繰返し回数の平均と最小と最大，提案法である Algorithm 5 における繰返し回数の平均と最小と最大，計算時間の平均と最小と最大，および，NMI 値の平均と最小と最大を調べた。また，人工データに関しては NMI 値が 1 となった回数も調べた。表 3 に人工データの実験結果を示す。表 3 の最初の列項目の番号はデータ名である。2 番目の列項目の K は提案法における K -近傍点数であり，各データにおける最初の行の「-」は従来法を意味する。そして，最後の列項目の「= 1」は NMI 値が 1 となった回数である。表 3 より，全体的に提案法の NMI の最大値は， K の値のいずれかで 1 となっており，その割合は異なるものの，カーネル k -平均法で分類可能なデータは，提案法で分類可能であると考えられる。また，計算時間に関しては，全体的にクラス数数の増加に比して提案法の計算時間の平均も増加していくのが見て取れる。各データで NMI 値が 1 となる回数が多い K に関して， δ の平均が表 1 の δ の推定区間の下限に近い値となっており，下から効率的に推定できていると考えられる。データ 1 では，提案法のすべての K の値に対して NMI 値は 1 となっている。4-5 節より，クラス数数の増加に伴い提案法の計算時間が大きくなると考えられるが，データ 1 のクラス数数は 2 であり，Al. 5 の繰返し数が 2 回であることから，アルゴリズムの特性上，1 回で $\delta \in \Delta$ となる δ の探索に成功していると考えられる。そのため， δ の更新回数は多いものの，Al. 2 の繰返し数が少なくなり，結果として計算時間が少なくなったと考えられる。データ 2 とデータ 4

はクラス数数の違いはあるものの，各クラス数が細長い環状に分布しており，初期クラス数保存度が高いと考えられる。従来法は 100 回の上限で収束できず停止しており，NMI 値の平均もかなり低い。特に，データ 2 に関しては δ の平均が $\delta \in \Delta$ となっていることから， δ の更新毎に Al. 2 が行うクラスタリングにおいて，結合したデータ集合を再分割していると考えられる。一方で提案法では， δ の探索を正負両方向に対して行えるため，探索回数に制限を設けていない。そのため， δ の更新回数の最大数はかなり高くなっているが， $0 \leq K \leq 2$ で平均 100 回程度であり，探索回数に制限を設けていないことが優位性につながったとは考えにくい。Al. 4 はクラス数間が十分な距離で分離している場合は，かなり高い確率で，正解クラス数に分離することができると推測できる。また，そのときの δ が $\delta \in \Delta$ であれば，初期値として $\delta \in \Delta$ と正解クラス数を与えて Al. 2 を実行した場合と同じになる。このことが，精度よく分類できた理由と考えられる。実際に，データ 2 では，提案法の δ は，ほぼ初期値の状態であり， $K = 0$ の場合は一度も δ の更新をしていないことに顕著に表れている。ただし， K の値が大きくなると，NMI が 1 となる回数が減少している。図 2，図 3 より，データ 1 に比してクラス数同士の分離がやや不十分であるためと考えられる。データ 3 は，データ数が異なる二つのクラス数からなっている。従来法の δ の平均は 0.49 で $\Delta = [0.09, 0.15]$ を大きく上回っており，うまく推定できなかったのに対して，提案法は全ての K において δ の平均が初期 δ

表3 人工データ(1~7)の実験結果

K	δ の更新回数			A1.5の繰返し数			A1.2の繰返し数			計算時間			δ			NMI			= 1	
	最小	平均	最大	最小	平均	最大	最小	平均	最大	最小	平均	最大	最小	平均	最大	最小	平均	最大		
1	-	1	50.5	56	-	-	-	2	2.8	11	0.13	1.78	2.02	0.07	0.87	0.95	0.00	0.90	1.00	45
	0	77	77.5	78	2	2.0	2	2	2.0	2	1.16	1.19	1.25	0.47	0.48	0.49	1.00	1.00	1.00	50
	1	77	77.5	78	2	2.0	3	2	2.0	2	1.17	1.20	1.27	0.47	0.48	0.49	1.00	1.00	1.00	50
	2	77	77.5	78	2	2.0	2	2	2.0	2	1.16	1.20	1.27	0.47	0.48	0.49	1.00	1.00	1.00	50
	3	77	77.5	78	2	2.0	2	2	2.0	2	1.16	1.20	1.32	0.47	0.48	0.49	1.00	1.00	1.00	50
4	77	77.5	78	2	2.0	2	2	2.0	2	1.17	1.20	1.35	0.47	0.48	0.49	1.00	1.00	1.00	50	
2	-	100	100.0	100	-	-	-	2	2.1	13	2.42	2.51	2.87	0.17	0.23	0.63	0.00	0.02	0.09	0
	0	0	0.0	0	3	3.0	3	2	2.0	2	0.32	0.34	0.35	0.17	0.17	0.17	1.00	1.00	1.00	50
	1	27	84.0	141	3	3.0	4	2	2.1	12	0.33	0.40	3.16	0.17	0.17	0.48	1.00	1.00	1.00	50
	2	27	84.0	141	3	3.0	4	2	2.1	12	0.33	0.40	3.16	0.17	0.17	0.48	1.00	1.00	1.00	50
	3	22	236.3	1945	2	3.0	4	2	2.9	25	0.34	2.26	29.29	0.10	0.19	0.50	0.02	0.41	1.00	8
4	54	363.0	1763	1	1.5	4	2	3.2	18	0.15	2.31	26.14	0.17	0.21	0.42	0.00	0.07	1.00	3	
3	-	5	47.2	49	-	-	-	2	2.8	9	0.34	2.73	2.99	0.09	0.49	0.50	0.00	0.23	0.24	0
	0	0	0.0	0	2	2.0	2	2	2.0	2	0.51	0.56	0.64	0.09	0.09	0.09	1.00	1.00	1.00	50
	1	0	0.0	0	2	2.0	3	2	2.0	2	0.50	0.56	0.72	0.09	0.09	0.09	1.00	1.00	1.00	50
	2	0	0.0	0	2	2.0	2	2	2.0	2	0.51	0.56	0.62	0.09	0.09	0.09	1.00	1.00	1.00	50
	3	47	52.0	110	2	2.4	20	2	3.0	22	0.52	1.12	27.76	0.09	0.10	0.27	0.54	0.99	1.00	49
4	0	0.0	0	2	2.0	3	2	2.0	2	0.51	0.56	0.69	0.09	0.09	0.09	1.00	1.00	1.00	50	
4	-	17	88.9	100	-	-	-	2	2.6	34	3.09	5.93	8.41	0.17	0.58	1.54	0.01	0.08	0.35	0
	0	2	64.0	1869	2	5.5	20	2	6.4	32	1.06	12.65	60.66	0.05	0.33	0.86	0.59	0.97	1.00	46
	1	1	79.4	1335	3	4.7	12	2	6.2	42	1.25	10.92	89.10	0.05	0.31	0.48	0.68	0.99	1.00	48
	2	2	119.3	1656	2	4.9	9	2	6.7	43	1.00	12.54	45.88	0.05	0.27	0.54	0.68	0.99	1.00	48
	3	3	158.1	3107	2	5.8	14	2	7.6	42	0.89	26.15	89.54	0.04	0.21	0.49	0.60	0.94	1.00	40
4	1	188.6	1974	2	6.0	12	2	8.7	42	0.91	31.22	93.00	0.05	0.21	0.54	0.58	0.95	1.00	43	
5	-	16	31.3	95	-	-	-	2	4.1	32	2.01	3.32	7.75	0.97	1.15	3.22	0.21	0.58	0.61	0
	0	1	101.6	1632	2	4.4	20	2	7.2	30	1.06	12.89	50.76	0.02	0.64	1.27	0.50	0.96	1.00	46
	1	1	99.5	1619	2	4.8	20	2	7.1	25	1.06	12.81	39.50	0.01	0.65	1.27	0.50	0.94	1.00	44
	2	1	114.5	2143	2	4.7	20	2	7.8	27	1.16	14.37	70.25	0.01	0.70	1.27	0.22	0.89	1.00	39
	3	1	102.5	1231	2	5.6	17	2	7.7	30	0.87	16.62	75.92	0.01	0.54	1.27	0.50	0.83	1.00	30
4	1	103.2	990	2	4.2	20	2	7.3	33	0.98	11.10	32.80	0.01	0.57	1.27	0.50	0.82	1.00	29	
6	-	8	14.3	19	-	-	-	2	4.0	18	0.28	0.37	0.51	0.22	0.25	0.26	0.94	1.00	1.00	46
	0	4	12.6	243	2	5.1	20	2	5.1	9	0.24	0.93	4.50	0.06	0.13	0.16	0.28	0.68	0.75	0
	1	1	14.6	405	2	13.2	20	2	6.7	17	0.18	2.84	4.57	0.06	0.13	0.37	0.28	0.61	1.00	23
	2	1	11.4	320	20	20.0	20	2	7.0	12	3.46	4.02	5.52	0.07	0.35	1.12	0.28	0.45	0.88	0
	3	1	11.5	214	20	20.0	20	2	6.0	12	2.35	3.69	4.80	0.09	1.16	1.29	0.32	0.87	0.88	0
4	2	34.7	838	1	5.6	20	2	5.3	11	0.10	1.05	3.71	0.01	0.18	0.52	0.15	0.53	0.97	0	
7	-	13	35.6	100	-	-	-	2	4.5	54	3.19	7.11	14.59	0.33	0.96	2.22	0.35	0.52	0.69	0
	0	1	38.6	1045	2	9.5	20	2	12.4	51	2.36	34.95	137.05	0.04	0.41	0.89	0.61	0.90	1.00	31
	1	1	49.3	1055	2	6.4	20	2	9.9	41	2.22	19.81	80.58	0.02	0.47	0.84	0.69	0.96	1.00	42
	2	1	57.4	711	2	6.3	20	2	10.1	47	2.01	24.46	95.56	0.06	0.46	1.27	0.33	0.95	1.00	42
	3	1	114.9	953	2	13.5	20	2	13.1	47	2.23	102.80	284.96	0.03	1.77	3.59	0.36	0.61	1.00	14
4	1	115.1	1017	2	8.2	20	2	10.7	44	2.05	50.00	257.82	0.05	0.15	0.51	0.56	0.82	1.00	8	

と、ほとんど同じである。K = 0, 1, 2, 4で δ の更新回数が0となっていることから、クラス間が十分な距離で分離しており、 $C^t \in C_{All}^*$ であればデータ数の違いの影響は低いと考え

られる。データ6では従来法の結果が良好であった。データ6は3つのクラスタからなり、各クラスタは凸形状ではあるものの、各クラスタの分布状態が異なっている。そのため、提

表 4 実データの実験結果

K	δ の更新回数			Al.5の繰返し数			Al. 2の繰返し数			計算時間(秒)			δ			NMI			
	最小	平均	最大	最小	平均	最大	最小	平均	最大	最小	平均	最大	最小	平均	最大	最小	平均	最大	
ir.	-	5	10.4	52	-	-	-	2	3.6	22	0.04	0.06	0.20	0.61	0.66	0.78	0.51	0.58	0.62
	0	3	13.8	313	2	7.4	20	2	5.9	12	0.09	0.20	0.57	0.55	0.87	1.40	0.57	0.71	0.74
	1	2	32.0	127	2	2.5	4	2	2.7	9	0.08	0.11	0.25	0.80	0.82	0.84	0.74	0.74	0.74
	2	1	7.4	310	20	20.0	20	2	3.5	16	0.42	0.57	0.93	0.70	0.72	0.84	0.66	0.66	0.66
	3	1	12.2	680	1	9.9	20	2	3.7	15	0.02	0.32	0.84	0.18	0.46	0.74	0.19	0.49	0.70
	4	4	103.0	202	1	1.2	2	2	2.1	4	0.02	0.04	0.27	0.27	0.56	3.88	0.00	0.24	0.74
wi.	-	1	3.0	100	-	-	-	2	2.0	6	0.01	0.02	0.46	52.69	52.95	65.81	0.08	0.30	0.31
	0	4	14.7	30	2	2.7	20	2	4.3	18	0.05	0.11	1.54	28.34	62.63	506.70	0.08	0.10	0.41
	1	2	34.5	352	2	6.0	20	2	6.5	20	0.06	0.50	2.59	14.89	101.26	464.56	0.09	0.38	0.42
	2	2	22.7	342	2	4.4	20	2	6.8	17	0.05	0.36	2.48	17.46	83.77	458.79	0.40	0.42	0.42
	3	1	36.7	377	3	10.0	20	2	7.5	18	0.32	1.13	2.49	13.02	76.40	285.55	0.39	0.42	0.44
	4	2	42.0	516	2	11.7	20	2	5.8	17	0.04	1.34	4.59	16.15	102.71	302.23	0.22	0.35	0.44
sa.	-	2	57.5	100	-	-	-	2	3.2	44	0.20	2.63	4.36	16.7	2.1×10^7	2.7×10^8	0.01	0.48	0.57
	0	0	0.0	0	2	2.2	3	2	2.0	2	0.93	1.58	2.19	14.44	14.44	14.44	0.32	0.35	0.36
	1	0	0.0	0	2	3.3	5	2	2.0	2	1.56	3.97	7.75	14.44	14.44	14.44	0.32	0.35	0.38
	2	0	0.0	0	1	1.1	3	2	2.0	2	0.15	0.16	0.46	14.44	14.44	14.44	0.01	0.01	0.02
	3	0	0.0	0	1	1.0	1	2	2.0	2	0.15	0.15	0.18	14.44	14.44	14.44	0.01	0.02	0.02
	4	0	0.0	0	1	1.0	1	2	2.0	2	0.15	0.16	0.18	14.44	14.44	14.44	0.01	0.01	0.02
pe.	-	10	23.3	71	-	-	-	2	5.6	40	2.98	5.05	11.43	123.0	4.7×10^7	3.2×10^8	0.67	0.72	0.76
	0	0	0.0	0	2	2.1	4	2	2.0	2	10.31	14.41	19.31	10.76	10.76	10.76	0.09	0.33	0.35
	1	0	0.0	0	2	10.3	20	2	2.0	2	16.6	289.4	763.6	10.76	10.76	10.76	0.32	0.34	0.39
	2	0	0.0	0	1	1.0	2	2	2.0	2	0.43	0.45	0.85	10.76	10.76	10.76	0.02	0.02	0.03
	3	0	0.0	0	1	1.0	1	2	2.0	2	0.43	0.44	0.45	10.76	10.76	10.76	0.02	0.02	0.03
	4	0	0.0	0	1	1.0	1	2	2.0	2	0.44	0.45	0.46	10.76	10.76	10.76	0.02	0.02	0.03

案法における R^2 上の距離基準では、適切に評価できなかつたと考えられる。データ 6 のようにガウス分布形状となっている場合は外れ値が悪影響を与えることが多く、実際に提案法における $K = 0$ のクラスタリング結果は全て (2,100,198) という分類であった。 $K = 1$ のときのみ、半分くらいの試行において NMI 値が 1 となっており、データの分布に対する最適な K の検討が必要と考える。データ 7 は 4 つのクラスタからなるため提案法の NMI 値が 1 となる割合は小さくなっているものの、従来法の NMI 値が一度も 1 となっていないことから、クラスタ数が増えてもクラスタ間が十分分離しているデータに対しては、ある程度の精度を期待できると考えられる。ただし、 K の値が増加すると、計算時間が増加するのに対して、NMI 値が 1 となる回数が減少している。十分な距離で分離していないクラスタ対が影響を与えていると考えられ、図 8 の外側の二つのクラスタが該当すると推測する。データ 5 に関

しても同様の事が考えられる。

次に、表 4 に実データの実験結果を示す。

表 4 より、iris は $K = 0, 1, 2$ で、従来法より提案法の NMI 値の平均が高くなっているが、 $K = 3, 4$ で低くなっている。完全に分離していないため、 K の増加に対して性能は低下するが、データ 6 のように極端に悪影響を与える外れ値は含んでいないと考えられる。wine に関しては、従来法の δ の更新回数の平均よりも、提案法の更新回数の平均の方が大きくなっている。また、 $K = 0$ のときのみ、 δ の平均が小さくなっているため、クラスタ間距離のみで分割する基準が適切ではないと考えられる。実際にデータ間の最大距離は 1402 とかなり大きい。 $K \geq 1$ で、従来法より提案法の NMI 値の平均が高くなっているものの、図 17 より、 δ の値はかなり広範囲で NMI 値が最大値に近い 0.4 となっているため、 $K \geq 1$ における従来法に対する提案法の優位性には疑問が残る。satimage と pendigit に関しては、提案法

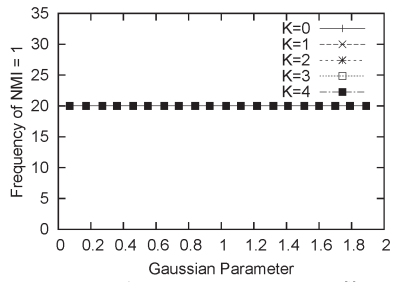


図 20 データ 1 の NMI=1 の数

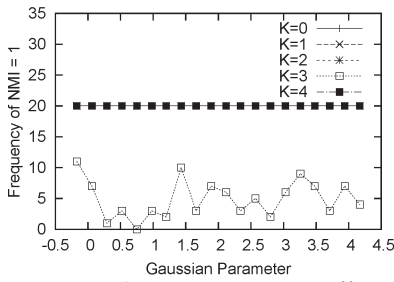


図 21 データ 2 の NMI=1 の数

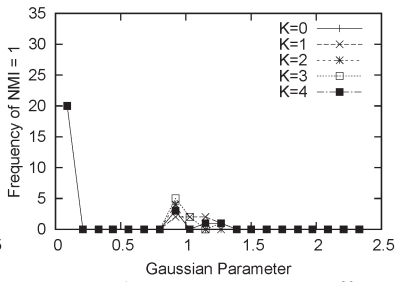


図 22 データ 3 の NMI=1 の数

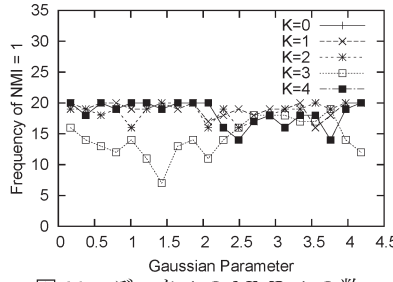


図 23 データ 4 の NMI=1 の数

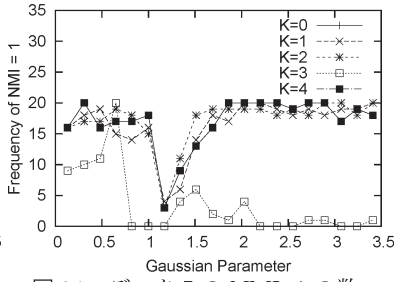


図 24 データ 5 の NMI=1 の数

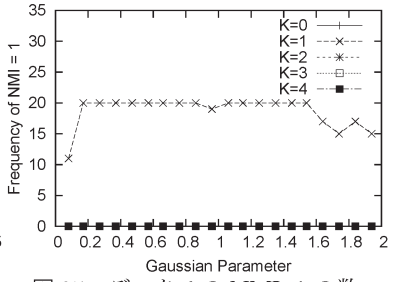


図 25 データ 6 の NMI=1 の数

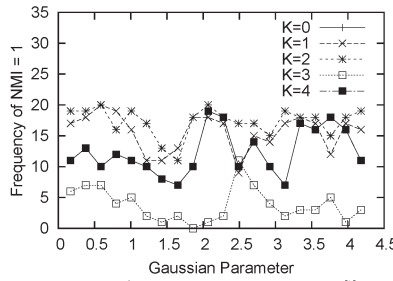


図 26 データ 7 の NMI=1 の数

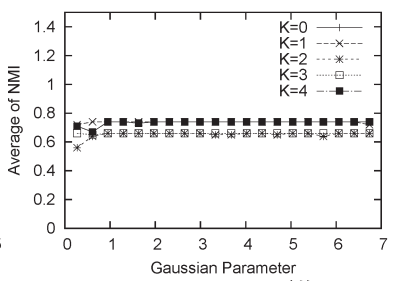


図 27 iris の NMI 平均

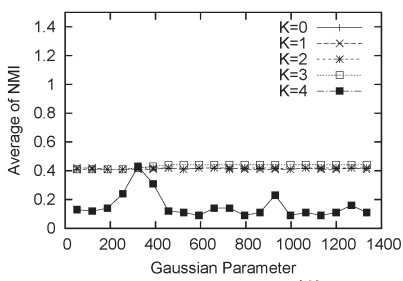


図 28 wine の NMI 平均

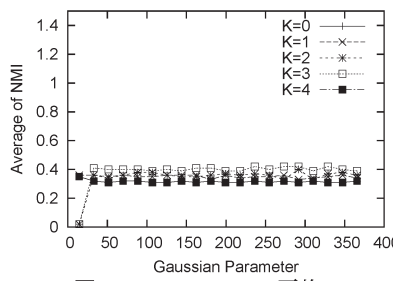


図 29 sat の NMI 平均

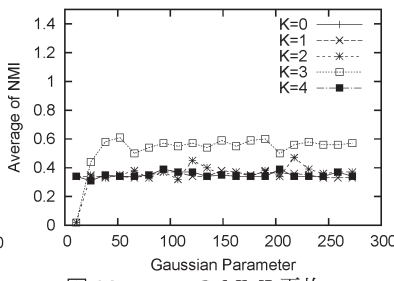


図 30 pen の NMI 平均

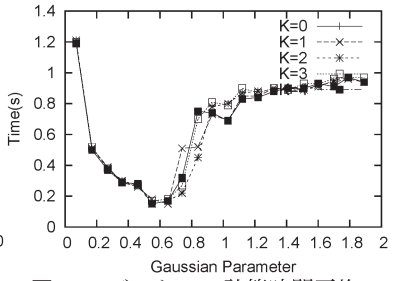


図 31 データ 1 の計算時間平均

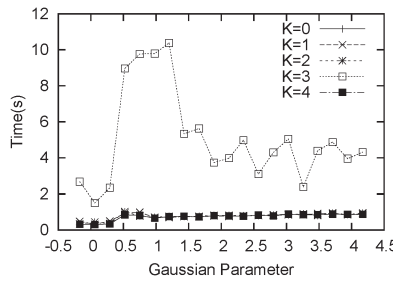


図 32 データ 2 の計算時間平均

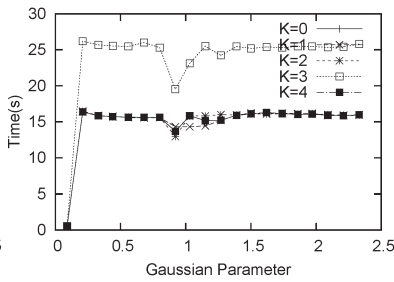


図 33 データ 3 の計算時間平均

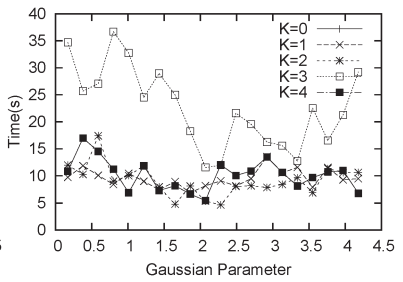


図 34 データ 4 の計算時間平均

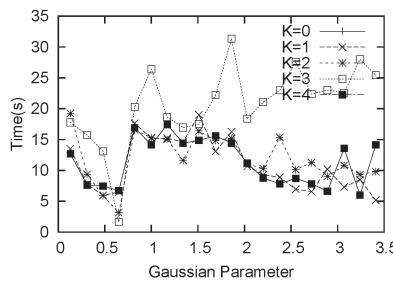


図 35 データ 5 の計算時間平均

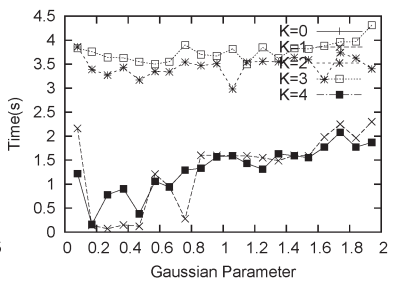


図 36 データ 6 の計算時間平均

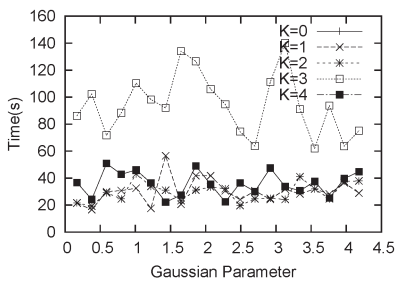


図 37 データ 7 の計算時間平均

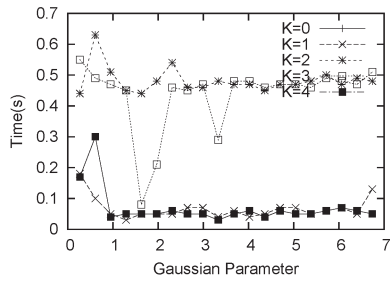


図 38 iris の計算時間平均

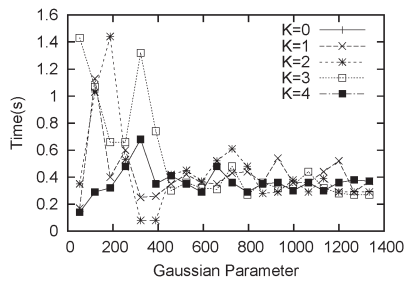


図 39 wine の計算時間平均

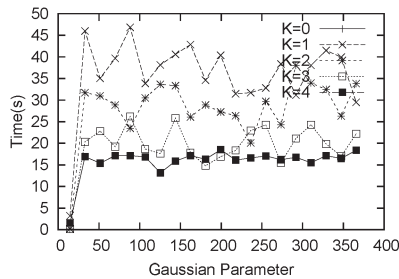


図 40 sat の計算時間平均

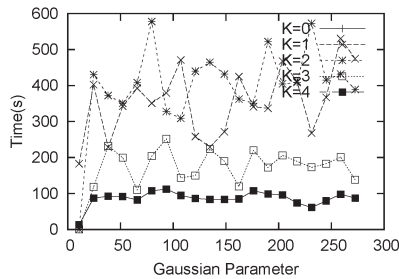


図 41 pen. の計算時間平均

の NMI 値は全体的に小さかった。 δ の更新回数が全て 0 となっており、データ 6 のような外れ値に分割された後、 δ の更新が発生しなかったためと考えられる。実際に、クラスタリング結果を調べたところ、pen digit は、ほとんどが (2,2,2,2,2,2,2,5,78,703), satimage は、ほとんどが (2,2,2,8,51,415) というように、一つのクラスタ数が極端に大きくなり、2 に近い値のクラスタ数がいくつか見られる偏りのある分類結果となっていた。また、図 18 と図 19 より、かなり広範囲な δ の値に関して、satimage の NMI 値は 0.6 近辺、pen digit の NMI 値は 0.75 近辺となっているため、初期クラスタへの依存度が低ければ、従来法のやや曖昧な停

止基準の方が有効であると考えられる。ただし、従来法の δ の平均と最大の値がかなり大きくなっており、pen digit に関しては 4 個程度が $10^7 \sim 10^8$ のオーダーであるものの、satimage に関しては半数程度が $10^7 \sim 10^8$ のオーダーとなっており、停止基準はやや不安定である。

5-5 初期値の変化に対する精度と計算時間

提案法の探索方向の有効性を確認するため、A1. 5 の入力である初期 δ の値に対する精度と計算時間の変化を調べた。表 1 と表 2 より、各データの初期 δ と距離の最大値を用いて $d\delta = (\text{距離の最大値} - \text{初期}\delta)/20$ とし、初期 δ から始めて逐次 $\delta = \delta + d\delta$ と加算しながら $\delta > \text{距離の最大値}$ となるまで 20 回の δ の更新を行った。各 δ に対して、20 回の A1. 5 の試行を行い、人工データに対しては、NMI 値が 1 となった回数と計算時間の平均、実データに対しては、NMI 値と計算時間の平均を調べた。結果を図 20～図 41 に示す。図 20～図 26 より、人工データに関して、データ 3 以外のデータで NMI が 1 となった回数は表 3 の結果とほぼ同等となっている。データ 3 のみ初期 δ の依存度が高いと考えられる。計算の詳細を確認すると、ほとんどの試行で A1. 4 で正解クラスタへの分類には成功しているものの、A1. 3 で δ の推定に失敗していた。いずれも、A1. 3 の 6 行目の 3 番目の停止条件を満たして計算を終了していたため、強制終了ではない。本来、式 (28) と式 (29) は 2 つのクラスタに所属する全データに対して評価しなければならない。データ数がほぼ同じクラスタに対しては、各クラスタから他のクラスタまでの最短距離に対応するデータである x_a と x_b のみの評価で、ある程度 δ を推定できている。 x_a と x_b は式 (45) より R^D 上の距離計算となるため、データ 3 に関しては、2 つのクラスタのデータ数が異なる影響が顕著に表れたと考えられる。図 27～図 30 より、実データの NMI の平均に関しては、iris と wine は表 4 の結果とほぼ同等である。satimage と pen digit は表 4 の従来法ほどではないものの、NMI の平均がやや向上しており、初期 δ の値は δ_{\min} よりも、ある程度大きい方がよいと考えられる。計算時間に関しては、図 40 より、satimage で $K=1$ のとき平均的に計算量が多くなっている。また、図 41 より、pen digit は全体的に計算量が多くなっているが、データ数が多いため妥当であ

り、逆に表4と異なり、提案法による計算が適切に実行されたとも判断できる。

6. 考察

表3より、提案法では $K=0$ でデータ6以外の人工データは、高い割合で正解クラスタへの分類に成功しており、特にクラスタ数が2の場合は、完全に正解クラスタを得ることに成功している。この理由について考察する。まず、クラスタ数が2の場合で、 $K=0, X=C_a^t \cup C_b^t$ とし、式(55)の仮定が成立するとき、Al.4は必ず正解クラスタへの分類に成功することを示す。

$$\begin{aligned} d_{\min}(C_a^t, C_b^t) &\geq d_{\min}(C_a, C_b), \forall C_a \cup C_b = X, \\ d_{\min}(C_a^t - Y, Y) &< d_{\min}(C_a^t - Y, C_b^t), \forall Y \subset C_a^t, \\ d_{\min}(C_b^t - Y, Y) &< d_{\min}(C_b^t - Y, C_a^t), \forall Y \subset C_b^t, \\ \{C_a, C_b\}, \{C_a^t, C_b^t\} &\in C_{Al1}^*. \end{aligned} \quad (55)$$

式(55)は任意の $2^{|X|}$ 個の2対の C_a と C_b の組合せに対する仮定に見えるため、かなり厳しい制約とも取れるが、図2や図3のようなデータ分布で $\{C_a, C_b\} \in C_{Al1}^*$ を仮定しているため、実際に分離可能なクラスタに対して設定する条件として妥当であると考えられる。まず、以下の場合分けを行う。

- (1) $C_a \subseteq C_a^t, C_b^t \subseteq C_b$.
- (2) $C_b \subseteq C_b^t, C_a^t \subseteq C_a$.
- (3) $C_a \not\subseteq C_a^t, C_b \not\subseteq C_b^t$.

(1)の場合については、式(55)の仮定より、 $Y = C_b - C_b^t$ を満たす Y に対して式(56)が成立する。

$$d_{\min}(C_a \cup Y, C_b - Y) = d_{\min}(C_a^t, C_b^t). \quad (56)$$

また、 $\forall Y \subset C_b^t$ に関して式(57)が成立する。

$$d_{\min}(C_a^t \cup Y, C_b^t - Y) < d_{\min}(C_a^t, C_b^t). \quad (57)$$

同様に、 $\forall Y \subset C_a$ に関して式(58)が成立する。

$$d_{\min}(C_a - Y, C_b \cup Y) < d_{\min}(C_a^t, C_b^t). \quad (58)$$

式(56)、式(57)、式(58)より、式(50)における X_b が、 $X_b = C_b - C_b^t$ となることを示せばよい。式(55)の2行目と3行目の仮定より、 $x \notin X_b, \exists x \in C_a^t$ かつ $x' \in X_b, \forall x' \in C_b^t$ となることはない。したがって、 $X_b \subseteq C_b - C_b^t$ または $C_b - C_b^t \subseteq X_b$ であるが、 $X_b \subset C_b - C_b^t$ または $C_b - C_b^t \subset X_b$ は式(55)の1行目の仮

定に矛盾する。したがって、Al.4は、 C_a^t と C_b^t に更新する。Al.5におけるAl.4の呼び出し回数は、もしAl.3が $\delta \in \Delta$ の推定に成功し、Al.1によるクラスタ更新に変化がなければ、最終的に C_a と C_b が変化しない終了段階の呼び出しも含めて2回となる。表3ではデータ1とデータ3が(1)の場合に該当すると考えられ、50回の試行におけるAl.4の全呼び出し回数が2となっている。(2)に関しても、 a と b の添字を入れ換えて考えると(1)の場合と同様のことがいえる。(3)に関しては、まず、 $C_a^t = C_{a,1}^t \cup C_{a,2}^t, C_{a,1}^t \cap C_{a,2}^t = \emptyset, C_b^t = C_{b,1}^t \cup C_{b,2}^t, C_{b,1}^t \cap C_{b,2}^t = \emptyset$ とし、 $C_a = C_{a,1}^t \cup C_{b,1}^t, C_b = C_{a,2}^t \cup C_{b,2}^t$ とする。式(58)の d_a と d_b に関して、 $d_b > d_a$ であったと仮定すると、式(55)の仮定と式(58)の定義より、 $C_{a,2}^t \subseteq X_b$ 、または、 $C_{b,2}^t \subseteq X_b$ となる。一般性を失わず、 $C_{a,2}^t \subseteq X_b$ とし $Y = X_b - C_{a,2}^t$ とすると、Al.4の更新により、式(59)の C_a', C_b' を得る。

$$\begin{aligned} C_a' &= C_{a,1}^t \cup C_{b,1}^t \cup C_{a,2}^t \cup Y \\ &= C_a^t \cup C_{b,1}^t \cup Y, \\ C_b' &= C_{b,2}^t - Y. \end{aligned} \quad (59)$$

式(59)より、次の更新は(1)の場合に帰着できるため、結局(3)の場合のAl.5におけるAl.4の呼び出し回数は、(1)の場合と同様Al.3が $\delta \in \Delta$ の推定に成功し、Al.1によるクラスタ更新に変化がないと仮定すると、最後の変化なしにおける呼び出しも含めて3回となる。表3より、データ2がこの場合に該当すると考えられる。 $C^t \in C_{Al1}^*$ であれば、 $\delta \in \Delta$ が存在するため、Al.3の $\delta \in \Delta$ の推定可能性の問題となるが、2つのクラスタのデータ数が同じであるデータ1とデータ2に関しては、距離の分布における中央値を探索方向の切り替えに利用し、最小値と最大値からなる区間を探索区間とした戦略が有効であったと推測できる。実際のデータのAl.5の分類過程における(1)の場合の典型的な例として、データ1で初期 δ を表1の初期 δ としたときの分類過程を図42と図43に示し、(3)の場合の典型的な例として、データ2で初期 δ を表1の初期 δ としたときの分類過程を図44、図45、図46に示す。また、(3)のもう一つの例として、データ1で初期 δ を表1の最大値としたときの分類過程を図47、図48、図49に示す。図42はデータ1に対す

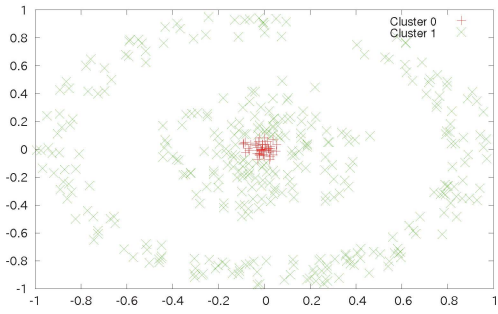


図 42 データ 1 の 1 回目の Algorithm 1 実行直後

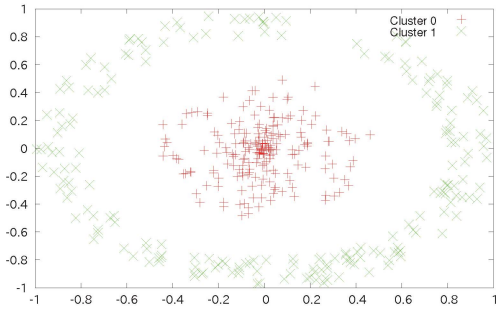


図 43 データ 1 の 1 回目の Algorithm 4 実行直後

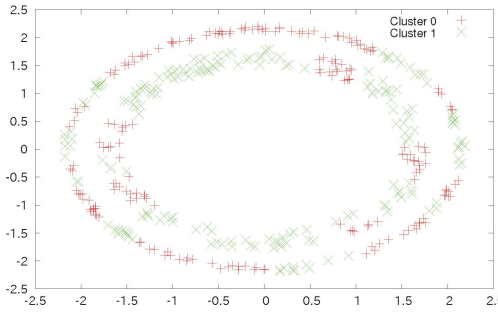


図 44 データ 2 の 1 回目の Algorithm 1 実行直後

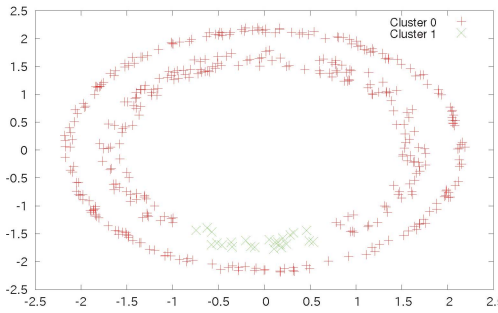


図 45 データ 2 の 1 回目の Algorithm 4 実行直後

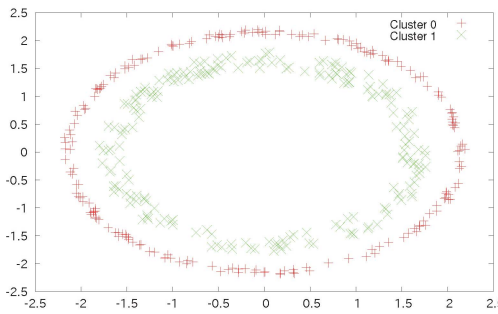


図 46 データ 2 の 2 回目の Algorithm 4 実行直後

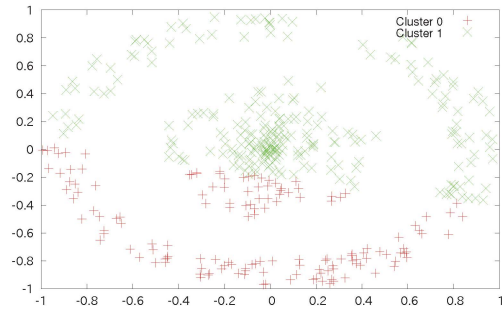


図 47 データ 1 で初期 δ を最大値としたときの 1 回目

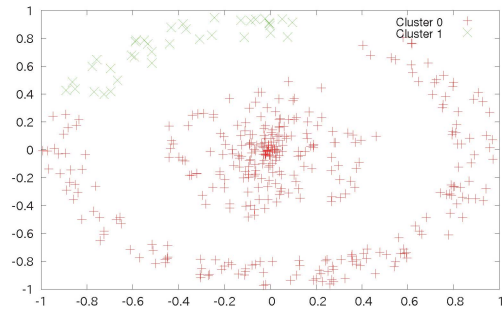


図 48 データ 1 で初期 δ を最大値としたときの 2 回目

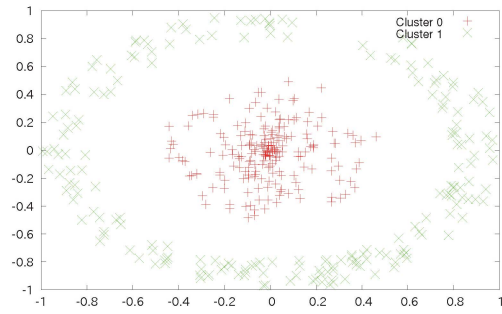


図 49 データ 1 で初期 δ を最大値としたときの 3 回目

る Al. 1 による初期クラスタリング結果を表しており、図 43 は Al. 4 により C_a と C_b の更新が終了した状態である。図 44 はデータ 2 に対する Al. 1 による初期クラスタリング結果を表しており、図 45 は Al. 4 により C_a と C_b の 1 回目の更新が終了した状態である。図 45 より、 $C_{a,1}^t \cup C_{b,1}^t$ と $C_{a,2}^t \cup C_{b,2}^t$ の状態から (1) の $C_{a,1}^t \cup C_{a,2}^t \cup C_{b,1}^t \cup Y$ と $C_{b,2}^t - Y$ の状態に遷移していることが見て取れる。最終的に、図 46 が 2 回目の更新が終了した状態であり、 δ の更新に成功できれば、正解クラスタが得られる状態となっている。データ 1 で初期 δ をデータ間距離の最大値とした場合の図 47, 図 48, 図 49 においても上記のデータ 2 と同じ過程で分類が行われていることを確認できる。

次に、クラスタ数が 2 以上の複数のクラスタに関して、表 3 より、外れ値の影響があるデータ 6 を除いて、クラスタ数が 3 のデータ

4 とデータ 5, および, クラスタ数が 4 のデータ 7 は, いくつかの試行で正解クラスタに分類することに失敗している。クラスタ間が十分な距離で分離しているにもかかわらず失敗する場合は, 複数クラスタが一つのクラスタとなってしまう, Al. 4 が機能しない状態だと考えられる。以下, 簡単のため, クラスタ数が 3 の場合で考える。3 つの正解クラスタを $C_1^t = C_{1,1}^t \cup C_{1,2}^t \cup C_{1,3}^t, C_2^t = C_{2,1}^t \cup C_{2,2}^t \cup C_{2,3}^t, C_3^t = C_{3,1}^t \cup C_{3,2}^t \cup C_{3,3}^t$ と 3 分割し, 現在得られているクラスタ C_1, C_2, C_3 が式 (60) の状態であると仮定する。

$$\begin{aligned} C_1 &= C_{1,1}^t \cup C_{2,1}^t \cup C_{3,1}^t, \\ C_2 &= C_{1,2}^t \cup C_{2,2}^t \cup C_{3,2}^t, \\ C_3 &= C_{1,3}^t \cup C_{2,3}^t \cup C_{3,3}^t. \end{aligned} \quad (60)$$

このとき, 二つのクラスタが一つに結合してしまう例を以下に示す。まず, C_1 と C_2 が更新対象に選択されたらと仮定する。2 つのクラスタの (3) の場合と同様, 式 (58) より, d_1 と d_2 に関して $d_1 > d_2$ となり, $C_{2,1} \cup C_{3,1} \subseteq X_1$ であったとすると, $Y = X_1 - (C_{2,1} \cup C_{3,1})$ として, Al. 4 による更新 C'_1, C'_2 は (3) と同じ考え方により, 式 (61) となる。

$$\begin{aligned} C'_1 &= C_{1,1}^t - Y, \\ C'_2 &= C_{1,2}^t \cup Y \cup C_{2,1}^t \cup C_{2,2}^t \cup C_{3,1}^t \cup C_{3,2}^t. \end{aligned} \quad (61)$$

次に, C'_2 と C_3 が更新対象に選択されたらと仮定する。上記同様, 式 (58) より, $d_3 > d_2$ となり, $C_{2,3} \cup C_{3,3} \subseteq X_3$ であったとすると, $Z = X_3 - (C_{2,3} \cup C_{3,3})$ として, Al. 4 による更新 C''_2, C'_3 は式 (62) となる。

$$\begin{aligned} C''_2 &= C_{1,2}^t \cup Y \cup Z \cup C_{2,1}^t \cup C_{2,2}^t \cup C_{2,3}^t \\ &\quad \cup C_{3,1}^t \cup C_{3,2}^t \cup C_{3,3}^t, \\ C'_3 &= C_{1,3}^t - Z. \end{aligned} \quad (62)$$

式 (62) より, $C'_2 \cup C'_3 \subseteq C''_2$ となる。式 (62) は, あくまでも一つの計算例であり, 実際には各部分クラスタが結合して二つ以上の正解クラスタが一つのクラスタとして分類されるまでの計算回数はより多くなると考えられる。クラスタ数 k の増加に伴い, 上記の状態が発生する可能性は高まると推測できる。また, Al. 4 による正解クラスタへの分類は成功しているものの, Al. 3 による δ の推定に失敗している

場合もあり, データ 5 がこれに該当する。表 3 では, データ 5 は 50 回中 4 回の試行において NMI が 1 とはならなかったが, その 4 回中 3 回が δ の推定の失敗であった。初期クラスタ割当の影響が少なからずあると推測できる。4 つのクラスタからなるデータ 7 の Al. 4 による典型的な分類失敗例を図 50 に示す。図 50 より, 外側の 2 つのクラスタが一つに結合されているのが見て取れる。

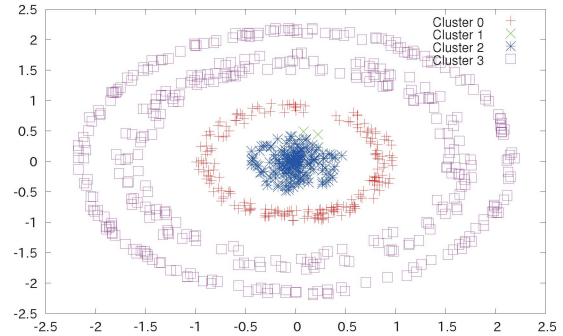


図 50 データ 7 の失敗例

7. むすび

本論文では, 文献 (柳, 2012) の改良として, データ間の距離の分布情報を利用して, カーネル k -平均法のガウスカーネルの適切なパラメータを探索するアルゴリズムを提案した。いくつかの人工データと実データを用いて, 従来法との比較実験を行い, クラスタ間が十分な距離で分離しているデータに関しては提案手法が有効であることを示した。一方で, クラスタが分離していないデータ, または, 外れ値を含むデータに関しては有効性を確認できなかった。今後の課題として, 外れ値にも強い基準の検討がある。

参考文献

- [1] Ben-Hur, A., D. Horn, H.T. Siegelmann, and V. Vapnik (2002) "Support vector clustering," *The Journal of Machine Learning Research*, Vol. 2, pp. 125–137.
- [2] Dhillon, I.S., Y. Guan, and B. Kulis (2004) "Kernel k -means: spectral clustering and normalized cuts," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge dis-*

- covery and data mining*, pp. 551–556, ACM.
- [3] Frank, A. and A. Asuncion (2010) “UCI Machine Learning Repository,” <http://archive.ics.uci.edu/ml>.
- [4] Girolami, M. (2002) “Mercer kernel-based clustering in feature space,” *Neural Networks, IEEE Transactions on*, Vol. 13, No. 3, pp. 780–784.
- [5] Hubert, L. and P. Arabie (1985) “Comparing partitions,” *Journal of classification*, Vol. 2, No. 1, pp. 193–218.
- [6] MacQueen, J. (1967) “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.
- [7] Máté, L. (1989) *Hilbert space methods in science and engineering*: Hilger.
- [8] Muller, K.R., S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf (2001) “An introduction to kernel-based learning algorithms,” *Neural Networks, IEEE Transactions on*, Vol. 12, No. 2, pp. 181–201.
- [9] Ng, A.Y., M.I. Jordan, and Y. Weiss (2002) “On spectral clustering: Analysis and an algorithm,” *Advances in neural information processing systems*, Vol. 2, pp. 849–856.
- [10] Strehl, A. and J. Ghosh (2003) “Cluster ensembles—a knowledge reuse framework for combining multiple partitions,” *The Journal of Machine Learning Research*, Vol. 3, pp. 583–617.
- [11] Vapnik Vladimir, N. (1998) *Statistical learning theory*: Wiley-Interscience.
- [12] Yu, S., L.C. Tranchevent, B. De Moor, and Y. Moreau (2011) “Optimized data fusion for kernel k -means clustering,” *Kernel-based Data Fusion for Machine Learning*, pp. 89–107.
- [13] 柳信一 (2012) 「クラスタ間の距離に基づくカーネル k -平均法」, 『北海道情報大学紀要』, 第24巻, 第1号, pp.47–70, 11月。