The University of Osaka
Institutional Knowledge Archive

Title	教師なし分類学習に関する計量心理学的研究
Author(s)	西田, 豊
Citation	大阪大学, 2013, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/27494
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

博士学位論文

教師なし分類学習に関する計量心理学的研究

西田　豊

大阪大学大学院人間科学研究科

人甲16286

博士学位論文

教師なし分類学習に関する計量心理学的研究

西田　豊

大阪大学大学院人間科学研究科

目次

第 1 章　序論 1

1.1　分類とクラスタリング 1

1.1.1　2 つの分類をめぐる研究 2

1.1.2　データ形式 3

1.1.3　類似度と距離 4

1.2　代表的な 3 つのクラスタリング手法 6

1.2.1　K 平均法 7

1.2.2　階層的クラスタリング 8

1.2.3　混合分布モデルによるクラスタリング 11

1.2.4　3 手法の特徴 14

1.3　本論文の目的と構成 16

第 2 章　数学的準備 18

2.1　ベクトルと行列 18

2.2　ラグランジュ乗数法 23

2.3　正則化法 24

2.4　分類結果の評価指標 26

第 3 章　ファジィクラスタリング 28

3.1　ファジィ集合 28

3.2　ファジィクラスタリング 29

3.3　ファジィ c 平均法 29

3.3.1　べき乗型ファジィ c 平均法 30

3.3.2 エントロピー正則化ファジィ c 平均法 31
3.4 クラスターサイズを考慮したファジィ c 平均法 33

第 4 章 K 平均法における変数選択法の開発 36
4.1 はじめに 36
4.2 提案手法 38
4.2.1 目的関数 38
4.2.2 アルゴリズム 38
4.3 数値実験 40
4.3.1 人工データ生成と解析方法 40
4.3.2 結果と考察 41
4.4 データ解析 42
4.4.1 アイリスデータ 42
4.4.2 動物評定データ 43
4.5 クラスターサイズを考慮した変数選択 K 平均法の開発 45
4.5.1 目的関数 45
4.5.2 アルゴリズム 46
4.5.3 数値実験 48
4.6 まとめ 49

第 5 章 データ行列の最小二乗近似によるファジィクラスタリング法の開発 53
5.1 はじめに 53
5.2 提案手法 54
5.2.1 目的関数 54
5.2.2 アルゴリズム 55
5.3 関連手法 56
5.4 数値実験 56
5.4.1 人工データ生成と解析方法 56
5.4.2 結果と考察 58
5.5 データ解析 59
5.5.1 データと解析方法 59

5.5.2 結果と考察 60
5.6 まとめ 63

第 6 章 次元縮約をともなうファジィ K 平均法の開発 64
6.1 はじめに 64
6.2 次元縮約 K 平均法 65
6.3 提案手法 67
6.3.1 目的関数 67
6.3.2 アルゴリズム 68
6.4 数値実験 68
6.4.1 数値実験 1 69
6.4.2 数値実験 2 70
6.5 データ解析 72
6.5.1 アイリスデータ 72
6.5.2 動物評定データ 74
6.6 クラスターサイズを考慮した次元縮約ファジィ K 平均法の改良 77
6.6.1 目的関数 77
6.6.2 アルゴリズム 78
6.6.3 データ解析 79
6.7 まとめ 79

第 7 章 総合考察 82
7.1 総括と提案手法の位置づけ 82
7.1.1 変数に対するパラメータのファジィ化 83
7.1.2 個体に対するパラメータのファジィ化 84
7.2 認知モデルとしての教師なし分類学習手法 85
7.2.1 認知機能へのモデルベースアプローチ 86
7.2.2 データ解析と認知機能の目的 86
7.2.3 教師なし概念生成の認知モデルとしてのクラスタリング 88
7.2.4 提案手法の概念形成モデルとしての適用可能性 90
7.3 計量心理学における教師なし分類学習の展望 91

7.3.1 分類境界の非線形化 91
7.3.2 制約付きクラスタリング 92
7.3.3 クラスタリングをともなった同時解析 92
7.3.4 クラスター数の決定法 93

要約 94

引用文献 99

謝辞 107

第 1 章

序論

本論文の主題は，計量心理学の立場から，分類を行うデータ解析法を考えることである．本章では分類の目的と方法について論じ，分類を目的とした従来のデータ解析法を俯瞰した上で，本論文の目的を導きだし，その概要を述べる．

1.1　分類とクラスタリング

分類とは物や現象を種類や性質などの基準により，似たもの同士をまとめ，似ていないものを区別することである．我々の周りには物や現象があふれており，対象を把握するためには少数の群に分けて整理することが必要である．人間は日常生活においても，あらゆる科学においても分類を行い，利用している．たとえば，図書館では発行形態や内容に基づいて図書が分類されているし，スーパーでは売り場ごとに商品が分類されている．科学分野では，古くは，博物学において自然物が分類され，知識が体系化されてきた．

分類は対象を把握する道具としてだけ存在するのではなく，我々が行っている外界の認識そのものにも深く関係する．ソシュールは命名すること，すなわち連続体を非連続体に区切ることによって，外界が認識されると考えた (de Saussure, 1916)．この連続体を非連続体に区切るという作業は分類に他ならない．たとえば，我々が知覚する色は，光の波長とともに連続的に変化すると考えられるが，特定の波長から得られる類似した視覚経験を分類し，「赤」，「青」といった名前をつけて区別している．

1.1.1　2 つの分類をめぐる研究

分類という操作はその特徴から 2 種類に分けることができる．我々が手にするデータには，外的基準が与えられている場合とそうでない場合がある．ここで，外的基準とは分類対象の個体がどの群に所属するかについての情報を指す．外的基準が与えられている分類は，あらかじめ分類すべき群が既知であるため教師あり分類と呼ばれる．それに対して，外的基準が与えられていない分類は，分類する群が未知であるため教師なし分類と呼ばれる．また，分類のための基準を構成すること，すなわちデータから知識を獲得することを分類学習という．

分類に関する研究は，様々な分野にまたがってなされてきた．生物学においては，多様な生物を分類することを目的とする分類学が，1 つの学問分野として成立している．統計学と心理学の分野ではそれぞれ異なる目的を持って，教師あり分類や教師なし分類について研究されてきた．データ解析法の開発を目的とする統計学や，その 1 分野である計量心理学では，教師あり分類は判別と呼ばれ，教師なし分類はクラスタリングと呼ばれている．判別分析は所属する群が既知のデータを用いて，群判別を行う規則を構成する分析手法の総称である (McLachlan, 1992; 佐藤, 2009)．判別分析は 2 つのステップから成り立つ．はじめに，どの群に所属しているかが既知のデータを用いて判別関数を求める．次に判別関数を用いて判別対象の個体を適切な群へ分類する．クラスタリングは，クラスター分析とも呼ばれ，所属する群が未知のデータを分類する解析法の総称である (Gan, Ma, & Wu, 2007; 宮本, 1999; 佐藤, 2009)．外的変数を用いないため，個体間の類似度もしくは非類似度に基づいて分類を行う．

人間の認知的プロセスの解明を目的とする認知心理学においては，「概念」は古くからあるテーマの 1 つで，概念がどのように形成されるかというトピックは現在でも盛んに研究されている．概念形成は，無限に存在する個体を，少数の抽象化された知識として整理するという点で分類といえる．概念形成の過程も，教師の「あり」と「なし」によって区別され，それぞれ，教師あり概念学習 (Kruschke, 1992; Nosofsky, 1986) や教師なし概念学習 (Fried & Holyoak, 1984; Pothos & Chater, 2002) と呼ばれる．ある概念を獲得するとき，事例と一緒にその事例がどの概念に属しているかというラベルとともに提示される場合がある．このような状況は教師あり概念学習である．何者か全く分からない未知の事例集合を前にしたとき，その事例を少数のグループに分類する．このような状況は教師なし

概念学習である．

計量心理学はもちろんであるが，実証研究が主な認知心理学においても，数理的アプローチが取られることがある．両者の違いはモデルがデータに特化しているか否かといえる．認知心理学におけるモデルは，認知プロセスを仮定し，どのようにデータが発生したかを記述できるのもで無くてはならない．すなわち，実証研究から得られた知見が反映された記述妥当性の高いモデルを構築することが求められる．それに対して，計量心理学ではどのような現象から得られたデータでも解析できる方法を開発することが目的である．したがって，計量心理学におけるモデルは，認知心理学におけるモデルよりも，比較的単純で汎用性があるものでなければならない．ただし，認知的なプロセスを組み込んだデータ解析法も存在し (例えば Nakatani, 1972; Shiina, 1986)，厳密に区別することは難しい．本論文では，分類学習のうちデータ構造の探索を目的とする教師なし分類学習，すなわちクラスタリングを計量心理学の立場から扱う．

1.1.2　データ形式

クラスタリングの対象となるデータの形式に次の 2 つがある (齋藤・宿久, 2006)．1 つは個体 × 変数の行列で表現される多変量データで，個体が持つ変数の値によって構成される (表 1)．もう 1 つは個体 × 個体の行列で表現される関連性データで，個体間の類似度（非類似度）によって構成される (表 2)．

データ形式によって，適用できる解析法が異なる．多変量データを対象としたクラスタリング法には K 平均法 (MacQueen, 1967) や混合分布を用いたクラスタリング (McLachlan & Basford, 1987) などがある．類似性データを対象としたクラスタリング法には階層的クラスタリングや ADCLUS (ADditive CLUStering: Shepard & Arabie, 1979)，および，その拡張手法である GENNCLUS(DeSarbo, 1982) や INDCLUS(Carroll & Arabie, 1983) などがある．多変量データから個体間の類似度（非類似度）を何らかの指標により定義し，関連性データを構成することは可能であるが，関連性データから多変量データを構成することはできない．したがって，観測されたものが関連性データであるならば，必然的に解析手法に制限が生じる．本論文で扱うデータ形式は，個体 × 変数の行列で表現される多変量データである．

表 1　多変量データの形式

	変			数		
個	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1m}
	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2m}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	x_{i1}	x_{i2}	$\cdots$	x_{ij}	$\cdots$	x_{im}
体	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	x_{n1}	x_{n2}	$\cdots$	x_{nj}	$\cdots$	x_{nm}

表 2　類似性データの形式

	個			体		
個	s_{11}	s_{12}	$\cdots$	s_{1n}	$\cdots$	s_{1n}
	s_{21}	s_{22}	$\cdots$	s_{2n}	$\cdots$	s_{2n}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	s_{i1}	s_{i2}	$\cdots$	s_{ii}	$\cdots$	s_{in}
体	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	s_{n1}	s_{n2}	$\cdots$	s_{ni}	$\cdots$	s_{nn}

1.1.3　類似度と距離

クラスタリングにおいては，類似度が重要な役割を果たす．類似度とは 2 つの個体間で定義される数量で，値が大きいほど似ていると仮定される．非類似度はその逆で，値が小さいほど似ていると仮定される．クラスタリングは外的変数を用いないため，類似度をもとに分類を行う．宮本 (1999) ではクラスタリングを「個体の集まりをいくつかのクラス（部分集合）に分割し，それぞれのクラスの中では個体どうしの類似度が大きく，異なるクラスについては類似度が小さくなるようにすることである.」と規定している．また，Tversky (1977) は類似性について「対象を分類し，概念を形成し，般化を行うための組織化原理としての役割を果たす[*1]」と述べており，心理学においても非常に重要な概念である．

非類似性は距離モデルによって表現されることが多い．個体間の非類似性は，定義された距離によって判断される．すなわち，図 1 で示されるような個体 i, j に対応する $\mathbf{x}_i, \mathbf{x}_j$ 間の距離を $d(\mathbf{x}_i, \mathbf{x}_j)$ で表すと，個体 i, j 間の非類似性を $d(\mathbf{x}_i, \mathbf{x}_j)$ と定義する．

[*1] 原文: It serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations.

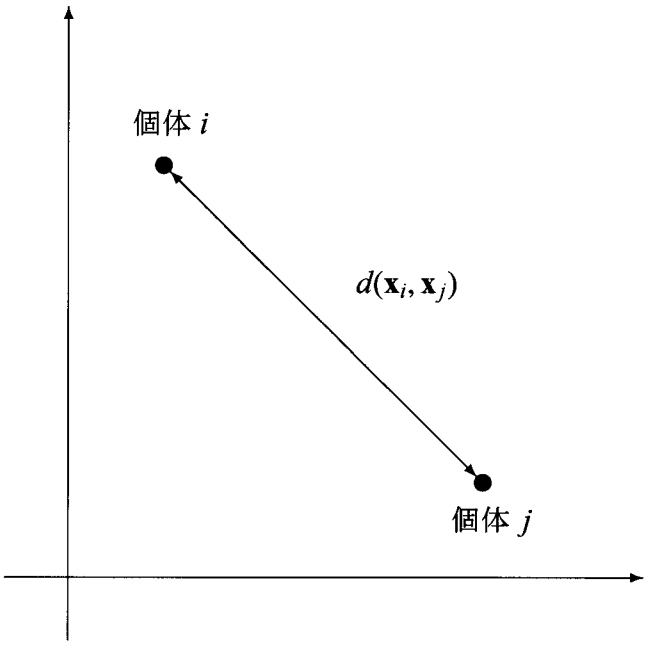

図 1　個体間の距離（2 次元空間）

ここでの距離とは次の条件を満たす関数のことである．

$$d(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \qquad \text{(非負性)} \tag{1.1}$$

$$d(\mathbf{x}_i, \mathbf{x}_i) = 0 \qquad \text{(同一性)} \tag{1.2}$$

$$d(\mathbf{x}_i, \mathbf{x}_j) = d(\mathbf{x}_j, \mathbf{x}_i) \qquad \text{(対称性)} \tag{1.3}$$

$$d(\mathbf{x}_i, \mathbf{x}_l) + d(\mathbf{x}_l, \mathbf{x}_j) \geq d(\mathbf{x}_i, \mathbf{x}_j) \ \text{(三角不等式)} \tag{1.4}$$

すなわち，(1.1) は距離は常に 0 以上である非負性を，(1.2) は自分自身との距離は 0 である同一性を，(1.3) は $\mathbf{x}_i$ から $\mathbf{x}_j$ への距離と $\mathbf{x}_j$ から $\mathbf{x}_i$ への距離が等しい対称性を，そして (1.4) は三角形の 3 辺の長さに関する条件である三角不等式を示している．

以上の距離の公理を満たすものにミンコフスキーのパワー距離

$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^{m} |\mathbf{x}_{ik} - \mathbf{x}_{jk}|^{\alpha} \right)^{1/\alpha} \tag{1.5}$$

がある．一般によく用いられるのは，$\alpha = 2$ としたユークリッド距離

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2} \tag{1.6}$$

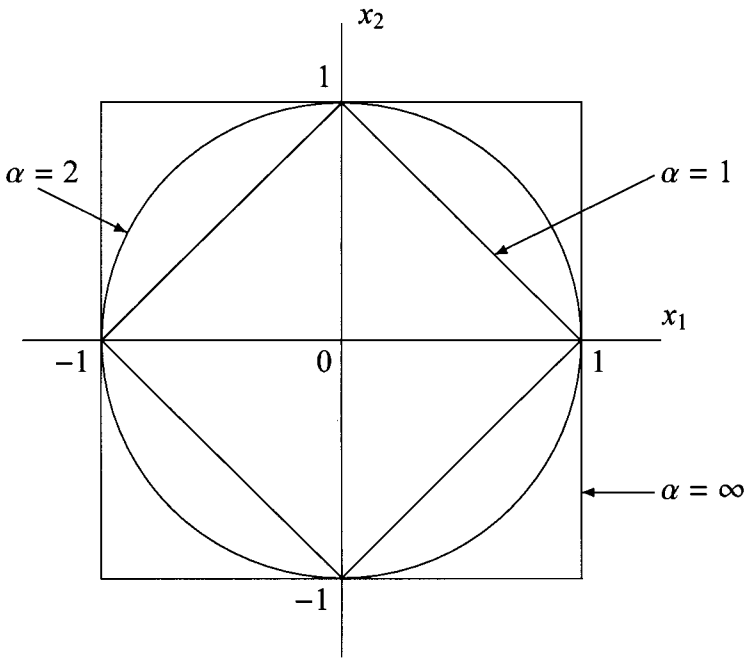

図 2　ミンコフスキー距離の基準円

であり，ユークリッド距離の 2 乗

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{m} (x_{ik} - x_{jk})^2 \tag{1.7}$$

を非類似度として用いることも多い．$\alpha = 1$ としたマンハッタン距離（市街距離）

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{m} |x_{ik} - x_{jk}| \tag{1.8}$$

が用いられることもある．$\alpha = \infty$ とした場合はチェビシェフ距離と呼ばれ，

$$d(\mathbf{x}_i, \mathbf{x}_j) = \max_{k=1,2,\ldots,m} |x_{ik} - x_{jk}| \tag{1.9}$$

と表される．

原点から等距離にある点を結んだ曲面を基準曲面と呼び，ユークリッド距離の場合，超球を描く．各距離概念における 2 次元での関係を図 2 に示す．

1.2　代表的な 3 つのクラスタリング手法

クラスタリングの方法は，大きく階層的クラスタリングと非階層的クラスタリングの 2 つに分けられる．階層的クラスタリングさらに併合的方法と分割的方法に分類されるが，

研究場面において，よく用いられているのは併合的方法である．

併合的階層クラスタリングは，似ている個体を逐次的にクラスターにまとめ上げるという作業を行う．最終的に全個体は 1 つのクラスターにまとめられる．クラスター形成過程は階層的になっており，樹形図を描くことによって結果を視覚的に確認できる．樹形図を任意の結合レベルでカットすることにより任意のクラスター数を得ることができる．

非階層的クラスタリングは，定義された基準を最小化，もしくは最大化するような個体の分割を見いだす方法である．最もよく用いられる方法は K 平均法である．ただし，K 平均法はある特定の方法というよりも，ある概念に基づくクラスタリング方法の族と考えられる (宮本, 1999).

クラスタリング手法は数多く提案されているが，ソフトウェアへの実装や，使い勝手から実際の研究でよく用いられているのは，K 平均法，階層的クラスタリング，混合分布モデルによるクラスタリングの 3 つである．本節では各手法を概観し，その特徴をまとめる．

1.2.1　K 平均法

K 平均法は非階層的手法であり，最もよく用いられるクラスタリング手法の一つである．データ行列を $N \times P$ の行列 $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]'$ とする．$\mathbf{x}_i$ は個体 i の P 次元ベクトルである．あらかじめ分割されるべきクラスター数 K が与えられているとき，K 平均法は以下の関数

$$F(\mathbf{U}, \bar{\mathbf{X}}) = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 \tag{1.10}$$

を最小にすることを目的とする．ただし，最小化に際して以下の制約を設ける．

$$u_{ik} \in \{0, 1\} \tag{1.11}$$

$$\sum_{k=1}^{K} u_{ik} = 1 \tag{1.12}$$

ここで $\mathbf{U} = (u_{ik})$ はメンバーシップと呼ばれる $N \times K$ のパラメータ行列である．メンバーシップ値は制約 (1.11) にあるように，0 か 1 の値のみを取る．個体 i がクラスター k に所属するときのみ $u_{ik} = 1$，所属しないとき $u_{ik} = 0$ という二値で表現される．

$\bar{\mathbf{X}} = [\bar{\mathbf{x}}_1, ..., \bar{\mathbf{x}}_K]'$ はセントロイドと呼ばれる $K \times P$ のパラメータ行列であり，$\bar{\mathbf{x}}_k$ はクラスター k の平均ベクトルを表す．目的関数 (1.10) の値を最小にすることは，クラスター内偏差平方和を最小にするようなクラスター割り当て $\mathbf{U}$ とクラスター平均 $\bar{\mathbf{X}}$ を求めていることを意味する．

メンバーシップ行列 $\mathbf{U}$ の最適解は

$$u_{ik} = \begin{cases} 1, \text{ if } \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 \leq \|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2. \\ 0, \text{ otherwise.} \end{cases} \tag{1.13}$$

で得られる．

また，最適なセントロイド行列 $\bar{\mathbf{X}}$ は

$$\bar{\mathbf{x}}_k = \frac{\sum_{i=1}^{N} u_{ik}\mathbf{x}_i}{\sum_{i=1}^{N} u_{ik}} \tag{1.14}$$

となる．

K 平均法全体のアルゴリズムは以下のようになる．

0. 初期化: パラメータを初期化する．クラスター数 K を与える．
1. **U**-step: $\bar{\mathbf{X}}$ を所与とし，(1.13) を用いて $\mathbf{U}$ を更新する．
2. $\bar{\mathbf{X}}$-step: $\mathbf{U}$ を所与とし，(1.14) を用いて $\bar{\mathbf{X}}$ を更新する．
3. 収束判定: (1.10) の値が収束していれば終了．そうでなければ 1 へ戻る．

1.2.2　階層的クラスタリング

階層的クラスタリング法では，定義された距離が近い個体から順番に 1 組ずつ結合し，小さなクラスターを形成することからはじめ，最終的にはすべての個体を含む 1 つのクラスターを形成する．分析者は任意の結合段階をクラスタリングの結果として用いることができる．データの集合を $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ とし，その個体間の非類似度を $d(\mathbf{x}_i, \mathbf{x}_j)$ とする．また，$\mathcal{G} = \{G_1, ..., G_K\}$ をクラスターの集合とし，クラスター間の非類似度を $D(G, G')$ とする．

階層的クラスタリング法のアルゴリズムは次のステップからなる．

1. すべての個体間の非類似度 $d(\mathbf{x}_i, \mathbf{x}_j)$ を計算する．
   各個体を個体を 1 つ含むクラスターとして考えるため，クラスター間の非類似度は $D(G_i, G_j) = d(\mathbf{x}_i, \mathbf{x}_j)$ と定義される．クラスター数を $K = N$ とする．
2. 距離が最短のクラスター対を結合する．
   クラスター間の非類似度が最小，すなわち $D(G_l, G_m) = \min D(G_i, G_j)$ となる 2 つのクラスター G_l と G_m を探す．G_l と G_m を $\mathcal{G}$ から取り除き，$G' = G_l \cup G_m$ を $\mathcal{G}$ に加える．クラスター数を 1 つ減らし $K = K - 1$ とする．
3. すべての $G_k \in \mathcal{G}, G_k \neq G'$ についてクラスター間の非類似度 $D(G', G_k)$ を再計算する．
4. $K = 1$ になるまで 2. と 3. を繰り返す．

階層的クラスタリングにおいて，非類似度の計算法はアルゴリズムとは独立したものであり，様々な方法が考えられている．分析者は以下に示すいずれかの非類似度の定義を選択することができる．しかし，個体間の非類似度をもとに定義されたクラスター間の非類似度は，非類似度更新のための計算量が多く扱いにくい．したがって，結合後のクラスター間距離を結合前のクラスター間距離により定式化した更新式を用いる方が効率が良い．これらの更新式は定義式より導かれる．

1. 最短距離法
   クラスター間で最も非類似度が小さい個体間の非類似度

$$D(G_l, G_m) = \min_{\mathbf{x} \in G_l, \mathbf{y} \in G_m} d(\mathbf{x}, \mathbf{y}) \tag{1.15}$$

   によってクラスター間非類似度を定義する．距離の更新式は

$$D(G_k, G_l \cup G_m) = \frac{1}{2} D(G_k, G_l) + \frac{1}{2} D(G_k, G_m) - \frac{1}{2} |D(G_k, G_l) - D(G_k, G_m)| \tag{1.16}$$

$$= \min\{D(G_k, G_l), D(G_k, G_m)\} \tag{1.17}$$

   で与えられる．
2. 最長距離法
   クラスター間で最も非類似度が大きい個体間の非類似度

$$D(G_l, G_m) = \max_{\mathbf{x} \in G_l, \mathbf{y} \in G_m} d(\mathbf{x}, \mathbf{y}) \tag{1.18}$$

によってクラスター間非類似度を定義する．非類似度の更新式は

$$D(G_k, G_l \cup G_m) = \frac{1}{2}D(G_k, G_l) + \frac{1}{2}D(G_k, G_m) + \frac{1}{2}|D(G_k, G_l) - D(G_k, G_m)| \quad (1.19)$$
$$= \max\{D(G_k, G_l), D(G_k, G_m)\} \quad (1.20)$$

で与えられる．

3. 群間平均法

クラスター間の個体同士すべての非類似度の平均

$$D(G_l, G_m) = \frac{1}{n_l n_m} \sum_{\mathbf{x} \in G_l, \mathbf{y} \in G_m} d(\mathbf{x}, \mathbf{y}) \quad (1.21)$$

によってクラスター間非類似度を定義する．ここで，n_l, n_m はそれぞれクラスター G_l, G_m に含まれる個体の個数である．非類似度の更新式は

$$D(G_k, G_l \cup G_m) = \frac{n_l}{n_l n_m} D(G_k, G_l) + \frac{n_m}{n_l n_m} D(G_k, G_m) \quad (1.22)$$

で与えられる．

4. 重心法

クラスター間非類似度をクラスター重心間の距離の 2 乗

$$D(G_l, G_m) = \|\mu(G_l) - \mu(G_m)\|^2 \quad (1.23)$$

で定義する．非類似度としてユークリッド距離を用いる．ここで，$\mu(G)$ は

$$\mu(G) = \frac{1}{n} \sum_{\mathbf{x} \in G} \mathbf{x} \quad (1.24)$$

で与えられるクラスター G の重心である．非類似度の更新式は

$$D(G_k, G_l \cup G_m) = \frac{n_l}{n_l + n_m} D(G_k, G_l) + \frac{n_m}{n_l + n_m} D(C_k, C_m) - \frac{n_l n_m}{(n_l + n_m)^2} D(G_l, G_m) \quad (1.25)$$

で与えられる．

5. Ward 法

Ward 法は凝集型最適化法として提案された (Ward, 1963)．クラスター G_l と G_m を結合したとき，増加する SSE

$$\Delta SSE(G_l, G_m) = SSE(G_l \cup G_m) - SSE(G_l) - SSE(G_m) \quad (1.26)$$

が最小になるような 2 つのクラスターを結合する．非類似度としてユークリッド距離を用いる．ここで SSE はクラスター G の重心 $\mu(G)$ と各個体 $\mathbf{x}$ との距離の 2 乗和を示し，

$$SSE(G) = \sum_{\mathbf{x} \in G} \|\mathbf{x} - \mu(G)\|^2 \tag{1.27}$$

と定義される．

クラスター間非類似度を重み付きクラスター重心間の距離の 2 乗

$$D(G_l, G_m) = \frac{n_l n_m}{n_l + n_m} \|\mu(G_l) - \mu(G_m)\|^2 \tag{1.28}$$

で定義する．非類似度の更新式は

$$\begin{aligned} &D(G_k, G_l \cup G_m) = \\ &\frac{n_k + n_l}{n_k + n_l + n_m} D(G_k, G_l) + \frac{n_k + n_m}{n_k + n_l + n_m} D(G_k, G_m) - \frac{n_k}{n_k + n_l + n_m} D(G_l, G_m) \end{aligned} \tag{1.29}$$

で与えられる．

また G_k, G_l, G_m を 3 つのクラスターとしたとき，G_k と $G_l \cup G_m$ 間の非類似度は Lance-Williams 更新式 (Lance & Williams, 1967)

$$\begin{aligned} &D(G_k, G_l \cup G_m) = \\ &\alpha_l D(G_k, G_l) + \alpha_m D(C_k, C_m) - \beta D(G_l, G_m) + \gamma |D(G_k, G_l) - D(G_k, G_m)| \end{aligned} \tag{1.30}$$

により統一的に記述することができる．ここで，α_l, α_m, β, γ は実数のパラメータで，表 3 のように定めると各更新式と一致する．

1.2.3　混合分布モデルによるクラスタリング

各クラスターと正規分布を対応させ，データが複数の正規分布から発生していると考えるモデルである．このモデルを混合正規分布と呼び，単純に正規分布をクラスター数 K だけ足し合わせたものとなる．

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x} | \mu_k, \mathbf{\Sigma}_k) \tag{1.31}$$

π_k は分布の混合比率を示すパラメータであり，$0 \leq \pi_k \leq 1$ かつ $\sum_k \pi_k = 1$ を満たす．ここで，2 値確率変数 $\mathbf{z} = \{z_k\}$ を導入する．ただし z_k は観測されない潜在変数であり，

表 3　Lance-Williams 更新式のパラメータ値

	α_l	α_m	β	γ
最長距離法	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
最長距離法	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
群間平均法	$\frac{n_l}{n_l+n_m}$	$\frac{n_m}{n_l+n_m}$	0	0
重心法	$\frac{n_l}{n_l+n_m}$	$\frac{n_m}{n_l+n_m}$	$-\frac{n_l n_m}{(n_l+n_m)^2}$	0
Ward 法	$\frac{n_k+n_l}{n_k+n_l+n_m}$	$\frac{n_k+n_m}{n_k+n_l+n_m}$	$-\frac{n_k}{n_k+n_l+n_m}$	0

$z_k \in \{0,1\}$ かつ $\sum_k z_k = 1$ を満たす．$\mathbf{z}$ の周辺分布は

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k} \tag{1.32}$$

と書ける．$\mathbf{z}$ が与えられた下での $\mathbf{x}$ の条件付き分布は

$$p(\mathbf{z}|\mathbf{x}) = \prod_{k=1}^{K} N(\mathbf{x}|\mu_k, \mathbf{\Sigma}_k)^{z_k} \tag{1.33}$$

となる．従って，$\mathbf{x}$ と $\mathbf{z}$ の同時分布は $p(\mathbf{x},\mathbf{z}) = p(\mathbf{z})p(\mathbf{z}|\mathbf{x})$ で与えられる．ここで，$p(\mathbf{x},\mathbf{z})$ を周辺化し $\mathbf{z}$ を消去すると

$$p(\mathbf{x}) = \sum_{z} p(\mathbf{x},\mathbf{z}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x}|\mu_k, \mathbf{\Sigma}_k) \tag{1.34}$$

となる．これは混合正規分布と同じとなる．すなわち，データが与えられていれば，個体 $\mathbf{x}_i$ に対応する潜在変数 $\mathbf{z}_i$ が存在する．

潜在変数を持つモデルの最尤解を求める方法として EM アルゴリズムがある (Dempster, Laird, & Rubin, 1977; McLachlan & Basford, 1987)．EM アルゴリズムは尤度関数を最大にするようなパラメータを E ステップと M ステップと呼ばれる 2 つのパラメータ更新手続きを繰り返し行う．

E ステップでは $\mathbf{x}$ が与えられた下での $\mathbf{z}$ の条件付き確率 $q(z_k) = p(z_k = 1|\mathbf{x})$ を

$$q(z_{ik}) = \frac{\pi_k N(\mathbf{x}_i|\mu_k, \mathbf{\Sigma}_k)}{\sum_{l}^{K} \pi_l N(\mathbf{x}_i|\mu_l, \mathbf{\Sigma}_l)} \tag{1.35}$$

によって計算する．

M ステップではモデルパラメータを計算する．平均 μ_k は

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N} q(z_{ik})\mathbf{x}_i, \tag{1.36}$$

分散共分散 $\mathbf{\Sigma}_k$ は

$$\mathbf{\Sigma}_k = \frac{1}{N_k} \sum_{i=1}^{N} q(z_{ik})(\mathbf{x}_i - \mu_k)(\mathbf{x}_i - \mu_k)', \tag{1.37}$$

混合比率 π_k は

$$\pi_k = \frac{N_k}{N}, \tag{1.38}$$

ただし，

$$N_k = \sum_{i=1}^{N} q(z_{ik}) \tag{1.39}$$

によって与えられる．以下に混合正規分布における EM アルゴリズムを示す．

0. 初期化：平均 μ_k，分散共分散 $\mathbf{\Sigma}_k$，混合比率 π_k を初期化し，対数尤度の初期値を計算する．
1. E ステップ：$\mathbf{x}$ が与えられた下での $\mathbf{z}$ の条件付き確率 $q(z_k) = p(z_k = 1|\mathbf{x})$ を (1.35) を用いて計算する．
2. M ステップ：モデルパラメータ，平均 μ_k，分散共分散 $\mathbf{\Sigma}_k$，混合比率 π_k を (1.36)，(1.37)，(1.38) を用いて計算する．
3. 収束確認：対数尤度を計算し，収束判定を行う．収束していなければ 1. へ戻る．

混合分布モデルによるクラスタリングは，個体 $\mathbf{x}_i$ がどのクラスターに所属するかを示す変数 $\mathbf{z}_i$ が欠損していると考え，EM アルゴリズムを用いてパラメータを推定する分析方法であるといえる．

$q(z_k)$ の値を用いることによって，どの個体がどのクラスターに所属しているかを解釈することができる．$q(z_k)$ は確率であるため 0 以上 1 以下の値をとる．そのため，K 平均法のように個体はただ 1 つのクラスターに所属するのではなく，所属する程度としての値が得られる．ただし，対数尤度には多くの局所解が存在し，EM アルゴリズムを用いて必ずしも大域解に収束するとは限らない．

1.2.4　3 手法の特徴

クラスター数について

K 平均法と混合分布モデルによるクラスタリングは非階層的クラスタリングであり，分析を実行する前に，クラスターの数を指定しなくてはならない．それに対して非階層的クラスタリングは事前にクラスター数を決めなくとも分析を実行できる．非階層的クラスタリングの場合，分析者は樹形図などを考慮して，分析後にクラスター数を決めれば良い．最もよく用いられる方法は，結合距離が極端に変化する段階を採用することである (齋藤・宿久, 2006)．K 平均法で用いられるクラスター数決定法としてギャップ統計量 (Tibshirani, Walther, & Hastie, 2001)，Cubic Clustering Criterion (CCC: Sarle, 1983) などがあるが，定番のものはない．混合分布モデルによるクラスタリングは確率モデルであるので，情報量規準 AIC (Akaike, 1974) や BIC (Schwarz, 1978) を用いてクラスター数を決定することが多い．しかし，AIC がクラスター数決定に有効であるかどうかは議論の余地がある (宮本, 2009).

初期値と局所解について

K 平均法は，データとして得られた個体を何らかの基準を用いて分割することを目的とする．これは組み合わせ最適化問題と呼ばれるものであり，大域解を求めることは難しい．したがって，得られた解は局所解であり，初期値の与え方に依存し，良い解に収束しないことも多い．複数の初期値を用いて，良い解を求める方法がとられる．

混合分布モデルによるクラスタリングは K 平均法と比べ，収束するまでの反復回数と 1 回の反復にかかる計算量が多い．そのため，混合分布モデルによるクラスタリングを実行するに当たっては良い初期値を選択する必要がある．この初期値の決定法に K 平均法が用いられることがある．はじめに K 平均法を適用し，そこで得られたパラメータ等を EM アルゴリズムの初期値として用いる．すなわち，EM アルゴリズムの平均の初期値には K

平均法で推定したクラスター平均を，分散共分散の初期値には K 平均法で計算したクラスター内分散を，そして，混合比率の初期値には全個体のうち K 平均法で分類したクラスターに所属する個体の比率をそれぞれ用いる．

階層的クラスタリングは目的関数を最小化するパラメータを求める方法ではないため，初期値や局所解は無く，同じ距離の定義を用いれば常に同じ結果を出力してくれる．ただし，異なる距離の定義を用いれば結果が大きく変わることもある．

K 平均法と混合分布モデルによるクラスタリングの関係

一見，全く異なる手法のように見える K 平均法と混合分布モデルによるクラスタリングであるが，両者のアルゴリズムを比較すると非常に似ていることが分かる (Bishop, 2006). 混合する正規分布の各要素の共分散行列が $\epsilon\mathbf{I}$ であるような混合正規分布モデルを考える．ϵ は全ての要素に共通の分散であり，$\mathbf{I}$ は単位行列である．

$$p(\mathbf{x}|\mu_k, \mathbf{\Sigma}_k) = \frac{1}{(2\pi\epsilon)^{D/2}} \exp\left\{-\frac{1}{2\epsilon}\|\mathbf{x} - \mu_k\|^2\right\} \tag{1.40}$$

$\mathbf{x}$ が与えられた下での $\mathbf{z}$ の条件付き確率 $q(z_k) = p(z_k = 1|\mathbf{x})$ は

$$q(z_{ik}) = \frac{\pi_k \exp\left\{\|\mathbf{x}_i - \mu_k\|^2/2\epsilon\right\}}{\sum_{l}^{K} \pi_l \exp\left\{\|\mathbf{x}_i - \mu_l\|^2/2\epsilon\right\}} \tag{1.41}$$

となる．ここで，$\|\mathbf{x}_i - \mu_k\|^2$ が最小になる k を k^* とおく．$\epsilon \to 0$ の極限を考えると分母での k^* に対応する項は最も遅く 0 に近づき，個体 $\mathbf{x}_i$ に関する事後確率は 1 に収束する $q(z_{ik^*})$ 以外はすべて 0 に収束する．したがって，極限においては K 平均法と同様にクリスプな分類がなされる．

K 平均法は共分散をパラメータとして持たず，平均のみを推定する手法である．K 平均法や混合分布モデルを拡張した手法として，共分散を持ちクリスプな分類を行う混合正規分布モデルに関する研究に楕円 K 平均法が (Sung & Poggio, 1998)，確率モデルによるアプローチでない共分散をパラメータとして持つ K 平均法として K-L 情報量正則化ファジィ c 平均法がある (宮岸・市橋・本多, 2001).

1.3　本論文の目的と構成

ここまで，分類を目的としたクラスタリング手法についてまとめた．混合分布モデルによるクラスタリングは確率モデルであるため，得られるメンバーシップパラメータは所属確率として得ることができるが，K 平均法や階層的クラスター分析ではメンバーシップパラメータは 2 値として表現される．メンバーシップの値が連続値で与えられるクラスタリングをファジィクラスタリングもしくはソフトクラスタリングと呼び，2 値で与えられるクラスタリングをクリスプクラスタリングもしくはハードクラスタリングと呼ぶ．

3 章で詳しく紹介するように，K 平均法にファジィ理論を取り入れ，個体がどれか 1 つだけのクラスターに所属するのではなく，どの程度所属するかを表現できるように拡張された手法がある (Bezdek, 1980; Miyamoto & Mukaidono, 1997)．宮本 (2009) はファジィクラスタリングの有用性について 4 つの観点，1. 理論の本質記述，2. 異種モデル結合，3. 技法の理論解析，4. 理論解析結果の応用，から考察している．このように，ファジィ概念を用いたクラスタリング技法を発展させることは，理論的にも応用的にも重要であり，進められるべきである．特に，探索的に用いられるクラスタリング法はパラメータがファジィな連続値で与えられる方が，柔軟な解釈を行うことができると考えられる．

本論文では，K 平均法を基本的考え方とし，新手法の開発と拡張を行う．より具体的には，パラメータをファジィ化することによって結果の解釈可能性の拡張を試みる．本論文では 3 つの研究において，新手法の開発と従来法の拡張を行っている．ただし，ファジィ理論は従来のようにメンバーシップパラメータだけに適用されるのではなく，変数に関する係数パラメータにも適用される．第 4 章，第 5 章，第 6 章が本論文の主となる研究で，パラメータをファジィ化したクラスタリング法の開発を行う．

第 4 章では，K 平均法においては潜在的に固定化されている変数に対する係数パラメータを顕在化した定式化を行い，ファジィ理論を援用することにより，変数に対する係数パラメータを連続値として推定可能なアルゴリズムを開発した．この方法では変数に対する係数パラメータを解釈することにより，変数選択を可能にしたり，クラスターの特徴を容易に理解できる．

第 5 章では，行列表現による K 平均法の定式化を用いることによって，従来行われてきたファジィ K 平均法とは異なるファジィクラスタリング法の目的関数を提案し，パラ

メータを推定する交互最適化アルゴリズムを開発した．この方法は主成分分析と目的関数が一致し，制約条件のみが異なるため，ファジィクラスタリングの特徴と，主成分分析の特徴を併せ持った中間的手法といえ，両手法と同様な解釈が可能となる．

第 6 章では，次元縮約とクラスタリングを単一の目的関数の最小化によって達成する次元縮約 K 平均法を拡張し，ファジィなメンバーシップパラメータを得られるような目的関数を提案し，そのパラメータを推定する交互最適化アルゴリズムを開発した．クラスターが明確に分離していない場合，ファジィなメンバーシップを用いることで，個体がクラスター所属する程度が連続値になり，より適切に解釈可能になる．

第 7 章では，第 4 章，第 5 章，第 6 章をふまえて，提案された手法についてクラスタリング手法もしくは多変量解析法としての位置づけを行い，総合的考察を行う．また，教師なし分類手法としてのクラスタリングを，心理的なモデルとしてとらえることの可能性を考察する．最後に，計量心理学におけるクラスタリング法の展望を述べる．

以上，本節では第 4 章から第 7 章で示される研究・考察の概要について述べたが，続く，第 2 章，第 3 章では後続する章のための準備を行う．第 2 章では線形代数と最適化についての数理的基礎とクラスタリング結果の評価指標を概説する．第 3 章ではファジィ概念を導入した種々の K 平均法を概観する．

第 2 章

数学的準備

本章は，本論文で必要となる線形代数，数値最適化，および解析結果の評価指標に関する数学的基礎をまとめ，後に続く章の準備を行う．はじめに，行列・ベクトルの表記法，基本演算，基本的定義を行い，逆行列，固有値分解，特異値分解についてまとめる．次に，パラメータ推定に必要となる 2 つの方法を概説する．1 つは，制約条件のもとで関数の極値を求める方法であるラグランジュ乗数法であり，もう 1 つは，不良設定問題を解くための正則化法である．最後に，クラスタリングの分類結果の指標となる Adjusted Rand Index を紹介する．

2.1　ベクトルと行列

本論文では，行列には $\mathbf{A}$ のように正立で大文字のボールド体の記号を，ベクトルには $\mathbf{a}$ のように正立で小文字のボールド体の記号を，スカラーには a のように斜体で小文字の記号を用いる．

ベクトルと行列とその転置

数 (スカラー) を縦に並べたものを列ベクトル，横に並べたものを行ベクトルという．n 個のスカラー，$a_1, a_2, ..., a_n$ を要素とする列ベクトル，m 個のスカラー，$b_1, b_2, ..., b_m$ を要

素とする行ベクトルはそれぞれ

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{2.1}$$

$$\mathbf{b} = \begin{bmatrix} b_1 & b_2 & \dots & b_m \end{bmatrix} \tag{2.2}$$

のように表せる．

nm 個のスカラーを n 行 m 列の表のようにまとめたものを行列という．

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \tag{2.3}$$

$\mathbf{A}$ は a_{ij} を要素とする $n \times m$ 行列である．行列の行と列を入れ替えることを転置と呼び，元の行列の右肩にプライム ($'$) をつけて表現する．行列 $\mathbf{A}$ を転置したものは m 行 n 列の行列

$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix} \tag{2.4}$$

である．

ベクトルを用いて行列を表すこともある．m 個の $n \times 1$ のベクトル $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$，もしくは n 個の $m \times 1$ のベクトル $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \dots, \tilde{\mathbf{a}}_n$ を用いると $\mathbf{A}$ は

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{a}}_1' \\ \tilde{\mathbf{a}}_2' \\ \vdots \\ \tilde{\mathbf{a}}_n' \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_2 & \dots & \tilde{\mathbf{a}}_n \end{bmatrix}' \tag{2.5}$$

と表せる．

零行列，対角行列，単位行列

要素がすべて 0 の $n \times 1$ ベクトルを零ベクトルといい，

$$\mathbf{0}_n = [0, \dots, 0]' \tag{2.6}$$

で表す．要素がすべて 0 の $n \times m$ 行列を零行列といい，

$$ {}_n\mathbf{O}_m = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \tag{2.7} $$

で表す．

行数と列数が等しい行列 $\mathbf{A} = (a_{ij})$ を正方行列と呼ぶ．正方行列 $\mathbf{A}$ の要素のうち $i = j$ である要素 a_{ii} を対角要素，$i \neq j$ である要素 a_{ij} を非対角要素と呼ぶ．非対角要素がすべて 0 の正方行列を対角行列と呼ぶ．対角要素が $d_1, d_2, ..., d_n$ である対角行列

$$ \mathbf{D} = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix} \tag{2.8} $$

と表せる．その対角要素のみを用いて

$$ \mathbf{D} = diag(d_1, d_2, ..., d_n) \tag{2.9} $$

と表記することもある．

特に，対角要素がすべて 1 である $n \times n$ 対角行列を単位行列と呼び，$\mathbf{I}_n$ と表記する．すなわち

$$ \mathbf{I}_n = diag(1, 1, ..., 1) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{2.10} $$

である．

行列の階数

m 個の $n \times 1$ のベクトル $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m$ の一次結合について

$$ c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + ... + c_m\mathbf{a}_m = \mathbf{0}_n \tag{2.11} $$

が成り立つのが $c_1 = ... = c_m = 0$ に限られるとき $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m$ は一次独立であるという．$c_1, ..., c_m$ のうち少なくとも 1 つが 0 でない場合において (2.11) が成り立つとき $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m$ は一次従属であるという．

$n \times m$ の行列 $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m]$ の m 個の列ベクトル $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m$ の中で一次独立なベクトルの最大個数を $\mathbf{A}$ の階数 (rank) と呼び，rank $\mathbf{A}$ と書く．

逆行列

$n \times n$ の正方行列 $\mathbf{A}$ に対し

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n \tag{2.12}$$

となる $n \times n$ の正方行列 $\mathbf{B}$ が存在するとき，この $\mathbf{B}$ を $\mathbf{A}$ の逆行列といい $\mathbf{A}^{-1}$ と書く．

正方行列 $\mathbf{A}$ の逆行列 $\mathbf{A}^{-1}$ が存在するとき，$\mathbf{A}$ は正則または非特異であるという．正則でない行列を特異であるという．

ノルムと距離

$n \times m$ の行列 $\mathbf{A} = (a_{ij})$ の要素の平方和を

$$\|\mathbf{A}\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2 = \sum_{j=1}^{m}\sum_{i=1}^{n} a_{ij}^2 \tag{2.13}$$

と書き，$\| \bullet \|^2$ で表す．$\| \bullet \|^2$ の平方根

$$\|\mathbf{A}\| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2} = \sqrt{\sum_{j=1}^{m}\sum_{i=1}^{n} a_{ij}^2} \tag{2.14}$$

を行列 $\mathbf{A}$ のノルムと呼ぶ．$n \times m$ の行列 $\mathbf{A} = (a_{ij})$ と $\mathbf{B} = (b_{ij})$ の対応する要素の差の二乗和は

$$\|\mathbf{A} - \mathbf{B}\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} (a_{ij} - b_{ij})^2 \tag{2.15}$$

である．

今度はベクトルの場合を考える．$n \times 1$ のベクトル $\mathbf{a} = [a_1, ..., a_n]'$ でも行列の場合と同様に，そのノルムは

$$\|\mathbf{a}\| = \|\mathbf{a}'\| = \sqrt{\mathbf{a}'\mathbf{a}} = \sqrt{\sum_{i=1}^{n} a_i^2} \tag{2.16}$$

と定義される．$\mathbf{a}$ と $n \times 1$ のベクトル $\mathbf{b} = [b_1, ..., b_n]'$ の差ベクトル $\mathbf{a} - \mathbf{b}$ のノルム

$$\|\mathbf{a} - \mathbf{b}\| = \sqrt{(\mathbf{a} - \mathbf{b})'(\mathbf{a} - \mathbf{b})} = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{2.17}$$

をベクトル間の距離と呼ぶ．

固有値と固有ベクトル

正方行列 $\mathbf{A}$ に対し，定数 δ および $\mathbf{0}$ でないベクトル $\mathbf{v}$ があり，

$$\mathbf{Av} = \delta\mathbf{v} \tag{2.18}$$

が成り立つとき，δ を $\mathbf{A}$ の固有値といい，$\mathbf{v}$ を固有値 δ に対応する固有ベクトルという．

対称行列の固有値分解

階数 r の $n \times n$ の対称行列 $\mathbf{A}$ は

$$\mathbf{A} = \mathbf{V\Delta V}' \tag{2.19}$$

と分解できる．ただし，$\mathbf{\Delta}$ は固有値を対角要素に持つ対角行列:

$$\mathbf{\Delta} = \mathrm{diag}(\delta_1, ..., \delta_r),\ \delta_1 \geq ... \geq \delta_r,\ \delta_j \neq 0,\ (j = 1, ...r) \tag{2.20}$$

であり，$\mathbf{V}$ は固有値に対応した固有ベクトルをその順に並べた直交行列:

$$\mathbf{V}'\mathbf{V} = \mathbf{I}_r \tag{2.21}$$

である．

特異値と特異ベクトル

階数 r の $n \times m$ の行列 $\mathbf{B}$ に対し，$\mathbf{BB}'$ あるいは $\mathbf{B}'\mathbf{B}$ の 0 より大きな固有値 $\delta_j (j = 1, ...r)$ の正の平方根

$$\lambda_j = \sqrt{\delta_j} \tag{2.22}$$

を $\mathbf{B}$ の特異値といい，対応する固有ベクトルを特異ベクトルという．

特異値分解

階数 r の $n \times m$ の行列 $\mathbf{B}$ は

$$\mathbf{B} = \mathbf{K\Lambda L}' \tag{2.23}$$

と分解できる．ただし，$\mathbf{\Lambda}$ は特異値を対角要素に持つ対角行列:

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_r),\ \lambda_1 \geq ... \geq \lambda_r > 0 \tag{2.24}$$

であり，$\mathbf{K}$，$\mathbf{L}$ は特異値に対応した特異ベクトルをその順に並べた直交行列:

$$\mathbf{K}'\mathbf{K} = \mathbf{L}'\mathbf{L} = \mathbf{I}_r \tag{2.25}$$

である．

証明は，ten Berge (1993)，高根 (1995) を参照のこと．

特異値分解による行列の最小二乗近似

階数 r の $n \times m$ の行列 $\mathbf{B}$ とし，その特異値分解が $\mathbf{B} = \mathbf{K}\boldsymbol{\Lambda}\mathbf{L}'$ で与えられているとする．また，上位 p 個の特異値を対角要素に持つ $p \times p$ の対角行列を $\boldsymbol{\Lambda}_p$，特異値に対応する左特異ベクトルおよび右特異ベクトルの行列をそれぞれ，$\mathbf{K}_p$，$\mathbf{L}_p$ と表す．このとき，$n \times m$ の行列 $\mathbf{A}$ が $\mathbf{B}$ よりも低階数であるという制約条件

$$\operatorname{rank} \mathbf{A} \leq p \leq r = \operatorname{rank} \mathbf{B} \tag{2.26}$$

のもとで

$$f(\mathbf{A}) = \|\mathbf{B} - \mathbf{A}\|^2 \tag{2.27}$$

を最小にする $\mathbf{A}$ は

$$\mathbf{A} = \mathbf{K}_p\boldsymbol{\Lambda}_p\mathbf{L}_p' \tag{2.28}$$

で与えられる (Eckart & Young, 1936).

2.2　ラグランジュ乗数法

1 つ以上の制約条件の下で関数の極値を求める方法である．ラグランジュ乗数と呼ばれるパラメータ λ を導入するためラグランジュ乗数法という．推定したい P 個のパラメータを $\boldsymbol{\theta} = (\theta_p)$ とし，制約条件 $g(\boldsymbol{\theta}) = 0$ のもとで関数 $f(\boldsymbol{\theta})$ の極値を求めることを考える．制約条件 $g = 0$ は $P - 1$ 次元の曲面を示す．この曲面上で関数 f が極値を取る点では関数 f の等値面が曲面 $g = 0$ に接している．両者の法線ベクトル ∇f，∇g は平行でなければならないため，ある定数 λ が存在して

$$\nabla f + \lambda \nabla g = 0 \tag{2.29}$$

が成り立つ．λ はラグランジュ乗数と呼ばれるパラメータである．

ここで，次式で定義されるラグランジュ関数 L

$$L(\boldsymbol{\theta}, \lambda) = f(\boldsymbol{\theta}) + \lambda g(\boldsymbol{\theta}) \tag{2.30}$$

を導入する．ラグランジュ関数 L を θ について偏微分すると

$$\frac{\partial L}{\partial \theta_p} = \frac{\partial f}{\partial \theta_p} + \lambda \frac{\partial g}{\partial \theta_p} \tag{2.31}$$

となり，極値の条件が $\nabla_{\boldsymbol{\theta}} L = \nabla f + \lambda \nabla g = 0$ になることが分かる．ラグランジュ関数 L を λ について偏微分すると

$$\frac{\partial L}{\partial \lambda} = g(\boldsymbol{\theta}) \tag{2.32}$$

となり，これを 0 と置いたものは制約条件そのものである．

すなわち，制約条件 $g(\boldsymbol{\theta}) = 0$ のもとで関数 $f(\boldsymbol{\theta})$ の極値を求めるためにはラグランジュ関数を定義し，$\boldsymbol{\theta}$ と λ の両方に対する極値を求めればよい．推定したいパラメータ数が P 個であれば，$P+1$ 個の方程式が得られる．未知数は $\theta_1, ..., \theta_P, \lambda$ の $P+1$ 個であるからこれを解けば解が求まる．

制約条件が複数ある場合も同様に求まる．たとえば C 個の制約条件 $g_c = 0 (c = 1, ..., C)$ のもとで関数 f の極値を求めることを考える．ラグランジュ関数 L を

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = f(\boldsymbol{\theta}) + \sum_{c=1}^{C} \lambda_c g_c(\boldsymbol{\theta}) \tag{2.33}$$

と置く．L を θ，λ について偏微分すると

$$\frac{\partial L}{\partial \theta_p} = \frac{\partial f}{\partial \theta_p} + \lambda \frac{\partial g}{\partial \theta_p} \tag{2.34}$$

$$\frac{\partial L}{\partial \lambda_c} = g_c(\boldsymbol{\theta}) \tag{2.35}$$

である．以上から $P+C$ 個の方程式が得られる．未知数は $\theta_1, ..., \theta_P, \lambda_1, ..., \lambda_C$ の $P+C$ 個であるからこれを解けば解が求まる．

2.3　正則化法

我々が扱う問題には良設定問題と不良設定問題という区別を行うことがある．Hadamard (1923) は次の 3 条件が満たされている数学的問題を良設定問題と定義した．

- 解が存在すること
- 解が一意であること
- 解はデータに対して連続的に依存すること

3 つ目の条件は「解が安定していること」と表現されることもある．良設定でない問題は不良設定問題と呼ばれる．このような問題に対しては正則化というアプローチが用いられる．よく用いられる正則化は Tikhonov の正則化法であり様々な分野で用いられている (Tikhonov & Arsenin, 1977).

たとえば，我々の視覚について考えてみよう．視覚処理が行っていることは 2 次元である網膜像から 3 次元である外界の構造を推定することである．Marr (1982) は初期視覚の目的を「網膜に投影された 2 次元画像データから 3 次元世界の可視表面の幾何学的構造を推測復元すること」と定義している．しかしながら 2 次元データから 3 次元構造を復元する問題は，解が一意に定まらず，不良設定問題である (Poggio, Torre, & Koch, 1985). Poggio et al. (1985) は視覚の不良設定問題を，正則化法を用いて解けることを示した．

統計学の分野ではパラメータ縮小推定法として正則化が用いられる．回帰分析を例にとるとその正則化を用いた目的関数は以下のようになる．

$$\|\mathbf{y} - \mathbf{XB}\|^2 + \lambda P(\mathbf{B}) \tag{2.36}$$

ここで第 1 項は通常の回帰分析における最小二乗基準であり，第 2 項はパラメータ β に対する正則化項を示す．正則化項の $P(\bullet)$ は様々な関数をとることができる．代表的なものに次の 2 つがある．正則化項にパラメータの 2 乗和：

$$P(\mathbf{B}) = \|\mathbf{B}\|^2 = \sum_{j=1}^{m} \beta_j^2 \tag{2.37}$$

をとったときリッジ回帰 (Hoerl & Kennard, 1970) と呼ばれる．リッジ回帰は独立変数間の相関が高く，多重強線性の問題がある場合でも安定した解が得られる．絶対値の和：

$$P(\mathbf{B}) = \|\mathbf{B}\|_1 = \sum_{j=1}^{m} |\beta_j| \tag{2.38}$$

のとき LASSO(Tibshirani, 1996) と呼ばれる．LASSO ではいくつかの推定値が 0 になり，スパースな解が得られる．0 であるパラメータに対応する変数は予測に影響を及ぼさないことを意味し，変数選択を行っていると解釈できる．

正則化された目的関数を最小化することは，正則化されていない最小 2 乗基準を制約条件：

$$\sum_{j=1}^{m} |\beta_j|^q \leq \eta \tag{2.39}$$

のもとで最小化することと等しい．ここで $q = 2$ のときリッジ回帰，$q = 1$ のとき LASSO となる．ただし η はラグランジュ乗数法によって決まる定数である．

2.4　分類結果の評価指標

本論文で用いる分類結果を評価するための指標を導入する．Adjusted Rand Index (ARI: Hubert & Arabie, 1985) は 2 つのクラスタリング結果を比較するための指標で，分類がまったく同じになれば上限の 1 をとる．与えられたデータ行列 $\mathbf{X}$ が 2 つの異なる方法，ここでは $\mathcal{P}$ と $\mathcal{Q}$，によってクラスタリングされたとする．$\mathcal{P}$ と $\mathcal{Q}$ はそれぞれ R 個と C 個のクラスターを持っている．各クラスターに分類された共起頻度 t は表 4 のようにあらわされる．このとき ARI は

$$ARI = \frac{\binom{N}{2}(a+d) - [(a+b)(a+c) + (c+d)(b+d)]}{\binom{N}{2}^2 - [(a+b)(a+c) + (c+d)(b+d)]}$$

によって計算される．ただし a, b, c, d はそれぞれ

$$a = \frac{\sum_{r=1}^{R}\sum_{c=1}^{C} t_{rc}^2 - N}{2},$$

$$b = \frac{\sum_{r=1}^{R} t_{r+}^2 - \sum_{r=1}^{R}\sum_{c=1}^{C} t_{rc}^2}{2},$$

$$c = \frac{\sum_{c=1}^{C} t_{+c}^2 - \sum_{r=1}^{R}\sum_{c=1}^{C} t_{rc}^2}{2},$$

$$b = \frac{\sum_{r=1}^{R}\sum_{c=1}^{C} t_{rc}^2 + N^2 - \sum_{r=1}^{R} t_{r+}^2 - \sum_{c=1}^{C} t_{+c}^2}{2}$$

である．

表 4　2 つの分割結果 $\mathcal{P}$ and $\mathcal{Q}$ において，各クラスターに分類された共起頻度

分割結果		$\mathcal{Q}$				
	クラスター	q_1	q_2	$\cdots$	q_C	合計
	p_1	t_{11}	t_{12}	$\cdots$	t_{1C}	t_{1+}
	p_2	t_{21}	t_{22}	$\cdots$	t_{2C}	t_{2+}
$\mathcal{P}$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	p_R	t_{R1}	t_{R2}	$\cdots$	t_{RC}	t_{R+}
合計		t_{+1}	t_{+2}	$\cdots$	t_{+C}	$t_{++} = N$

第 3 章

ファジィクラスタリング

本章では，本論文の複数の章において言及されるファジィ c 平均法について記述する．はじめに，ファジィ集合を定義し，概念を簡単に説明する．その後，ファジィ集合の概念を用いたファジィクラスタリングについて概観する．最後に，K 平均法の制約条件を緩めたファジィ c 平均法の異なる定式化による，2 種類のバリエーションとクラスターサイズを考慮した拡張手法を紹介する．

3.1　ファジィ集合

ファジィ集合は曖昧さを数値として表現することのできる集合論として提唱された (Zadeh, 1965)．通常の集合論においては，要素が集合に属するか否かははっきりと判断できる．しかしながら，日常生活においては，ある要素が集合に属するか属さないかが分からない，曖昧なものも存在する．ファジィ集合ではこのような曖昧さを取り扱うために要素が集合へ属する程度を用いて表現する．

全体集合 X のファジィ部分集合 A とは，メンバーシップ関数：

$$h_A : X \to [0, 1] \tag{3.1}$$

によって特性づけられた集合である．任意の要素 $x \in X$ に対して，$h_A(x)$ は x がファジィ部分集合 A に属する程度を表す．

$h_A(x)$ の値が 1 であれば完全に x が A に属することを意味し，0 であれば完全に x が A に属さないことを意味し，$0 < h_A(x) < 1$ の値であればその値の程度だけ x が A に属することを意味する．通常の集合論の特性関数の値域は，0 か 1，つまり，所属するかしない

かの 2 値であるため，メンバーシップ関数は通常の集合論における特性関数を連続値へと拡張したものと考えることができる．

ファジィ集合にも通常の集合と同様に演算などが定義されるが，ファジィクラスタリングを考えるにあたってはファジィ集合の詳細は必要無いので割愛する．

3.2　ファジィクラスタリング

通常のクラスタリングでは，個体の集合を互いに排他的部分集合に分割することを目標とする．ある個体がクラスターに属するか否かが明確な状態である．しかしながら，実際のデータに関しては，クラスターへの所属が明確でない状態もある．また，強制的に分類したとしても，有益な情報をデータから引き出しているとは言えない．

そこで，クラスターへの所属が明確でない状態を，そのままを数値で表現することを考える．ファジィ部分集合の概念を用いて，個体がクラスターにどの程度属するかを表現する方法がファジィクラスタリングである．K 平均法のメンバーシップにファジィネスが導入された手法が次節で紹介するファジィ c 平均法である．[*1]

3.3　ファジィ c 平均法

ここでは本論文で言及される 2 つの方法：べき乗型ファジィ c 平均法とエントロピー正則化ファジィ c 平均法を紹介する．ファジィ c 平均法は K 平均法においてクラスターをファジィ部分集合で表現したものであり，個体が複数のクラスターに所属することを許した方法である (宮本, 1999; Miyamoto, Ichihashi, & Honda, 2008; 佐藤, 2009).

K 平均法の目的関数は

$$F(\mathbf{U}, \bar{\mathbf{X}}) = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 \tag{3.2}$$

のように書くことができた．その制約は

$$u_{ik} \in \{0, 1\} \tag{3.3}$$

$$\sum_{k=1}^{K} u_{ik} = 1 \tag{3.4}$$

[*1] Bezdek が命名したと思われるファジィ c 平均法という呼称が用いるが，なぜファジィ K 平均法でないのかは不明である．c や K はクラスター数を示すパラメータで本質的な違いはない．

であった．ファジィ c 平均法では，この u_{ik} が 0 か 1 の値のみ取り得るという制約を緩め，0 から 1 の値を取り得るようにする．すなわち

$$u_{ik} \in [0, 1] \tag{3.5}$$

$$\sum_{k=1}^{K} u_{ik} = 1 \tag{3.6}$$

と制約を置く．しかし，このように u_{ik} に関する制約を 2 値から連続値へと緩和しただけではファジィメンバーシップを得ることはできない (宮本, 1999)．K 平均法の目的関数は u_{ik} に関して線形であり，制約も線形であるため，この問題は線形計画となり，最適解は制約条件が形成する多角形の端点である．端点はクリスプな解を与えるので，ファジィメンバーシップを得られない．従って，ファジィネスを導入するためには目的関数を変更する必要がある．

3.3.1　べき乗型ファジィ c 平均法

ファジィパラメータをメンバーシップパラメータにべき乗することによりファジィメンバーシップを得るファジィ c 平均法について述べる．この方法は Dunn (1973) が提唱し，Bezdek (1980) が一般化を行った．べき乗型ファジィ c 平均法の目的関数は

$$F_{eFCM}(\bar{\mathbf{X}}, \mathbf{U}) = \sum_{i=1}^{N} \sum_{k=1}^{K} (u_{ik})^{\alpha} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 \tag{3.7}$$

と表される．ここで α は $\alpha > 1$ を満たすファジィパラメータである．

メンバーシップパラメータ $\mathbf{U}$ の最適解はラグランジュ乗数法を用いて求めることができる．制約条件 (3.6) に関して，ラグランジュ乗数を λ とすると，ラグランジュ関数は次のように書ける．

$$L(\mathbf{U}, \lambda) = F_{eFCM} + \lambda(\sum_{k=1}^{K} u_{ik} - 1) \tag{3.8}$$

$L(\mathbf{U}, \lambda)$ を u_{il} で偏微分すると，

$$\frac{\partial L(\mathbf{U}, \lambda)}{\partial u_{il}} = \alpha (u_{il})^{\alpha - 1} \|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2 + \lambda \tag{3.9}$$

となる．$\partial L(\mathbf{U}, \lambda)/\partial u_{il} = 0$ を解いて

$$u_{il} = \left(\frac{-\lambda}{\alpha \|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2} \right)^{1/(\alpha - 1)} \tag{3.10}$$

を得る．(3.6) を用いると，

$$\sum_{l=1}^{K} u_{il} = \sum_{l=1}^{K} \left(\frac{-\lambda}{\alpha}\right)^{1/(\alpha-1)} \left(\frac{1}{\|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2}\right)^{1/(\alpha-1)} \tag{3.11}$$

$$= \left(\frac{-\lambda}{\alpha}\right)^{1/(\alpha-1)} \left[\sum_{l=1}^{K} \left(\frac{1}{\|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2}\right)^{1/(\alpha-1)}\right] = 1 \tag{3.12}$$

となり，移行して，

$$\left(\frac{-\lambda}{\alpha}\right)^{1/(\alpha-1)} = \left[\sum_{l=1}^{K} \left(\frac{1}{\|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2}\right)^{1/(\alpha-1)}\right]^{-1} \tag{3.13}$$

を得る．これを (3.10) に代入すると，

$$u_{ik} = \left[\sum_{l=1}^{K} \left(\frac{\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2}{\|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2}\right)^{\frac{1}{\alpha-1}}\right]^{-1} \tag{3.14}$$

が得られる．

また，クラスター中心 $\bar{\mathbf{X}}$ の最適解は F_{eFCM} を $\bar{\mathbf{x}}_k$ について偏微分すれば

$$\bar{\mathbf{x}}_k = \frac{\sum_{i=1}^{N} (u_{ik})^{\alpha} \mathbf{x}_i}{\sum_{i=1}^{N} (u_{ik})^{\alpha}} \tag{3.15}$$

を得る．

べき乗型ファジィ c 平均法のアルゴリズムは以下のようになる．

0. 初期化: パラメータ $\mathbf{U}$，$\bar{\mathbf{X}}$ を初期化する．ファジィパラメータ α とクラスター数 K を与える．
1. $\mathbf{U}$-step: $\bar{\mathbf{X}}$ を所与とし，(3.14) を用いて $\mathbf{U}$ を更新する．
2. $\bar{\mathbf{X}}$-step: $\mathbf{U}$ を所与とし，(3.15) を用いて $\bar{\mathbf{X}}$ を更新する．
3. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

3.3.2　エントロピー正則化ファジィ c 平均法

エントロピー関数による正則化を利用し，K 平均法をファジィ拡張したものがエントロピー正則化ファジィ c 平均法である (Miyamoto & Mukaidono, 1997)．エントロピー正則

化ファジィ c 平均法の目的関数は

$$F_{rFCM}(\bar{\mathbf{X}}, \mathbf{U}) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 + \lambda^{-1}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\log u_{ik} \tag{3.16}$$

とあらわされる．ここで λ は $\lambda > 0$ を満たす正則化パラメータである．

メンバーシップパラメータ $\mathbf{U}$ の最適解はラグランジュ乗数法を用いて求めることができる．制約条件 (3.6) に関して，ν をラグランジュ乗数とすると，ラグランジュ関数は次のように書くことができる．

$$L(\mathbf{U}, \nu) = F_{rFCM} + \nu(\sum_{k=1}^{K} u_{ik} - 1) \tag{3.17}$$

$L(\mathbf{U}, \nu)$ を u_{ik} で偏微分すると，

$$\frac{\partial L(\mathbf{U}, \nu)}{\partial u_{ik}} = \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 + \lambda^{-1}(1 + \log u_{ik}) + \nu \tag{3.18}$$

となる．$\partial L(\mathbf{U}, \nu)/\partial u_{ik} = 0$ を解いて

$$u_{ik} = \exp(-\lambda\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 - \lambda\nu - 1) \tag{3.19}$$

を得る．(3.6) を用いると

$$\exp(-1 - \lambda\nu) = \left\{\sum_{k=1}^{K}\exp(-\lambda\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2)\right\}^{-1} \tag{3.20}$$

となるので，これを (3.19) に代入すると

$$u_{ik} = \frac{\exp(-\lambda\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2)}{\sum_{l=1}^{K}\exp(-\lambda\|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2)} \tag{3.21}$$

を得る．

また，最適なクラスター中心パラメータ $\bar{\mathbf{X}}$ は F_{rFCM} を $\bar{\mathbf{x}}_k$ について偏微分すれば

$$\bar{\mathbf{x}}_k = \frac{\sum_{i=1}^{N} u_{ik}\mathbf{x}_i}{\sum_{i=1}^{N} u_{ik}} \tag{3.22}$$

より得られる．

エントロピー正則化ファジィ c 平均法のアルゴリズムは以下のようになる．

0. 初期化: パラメータ $\mathbf{U}$, $\bar{\mathbf{X}}$ を初期化する．正則化パラメータ λ とクラスター数 K を与える．
1. $\mathbf{U}$-step: $\bar{\mathbf{X}}$ を所与とし，(3.21) を用いて $\mathbf{U}$ を更新する．
2. $\bar{\mathbf{X}}$-step: $\mathbf{U}$ を所与とし，(3.22) を用いて $\bar{\mathbf{X}}$ を更新する．
3. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

3.4　クラスターサイズを考慮したファジィ c 平均法

K 平均法はクラスターに含まれる個体の数が等しくなりやすいことが知られている．そのため，クラスターに含まれる個体の数が大きく異なるように分類されるべき場合でも，個体の数が等しくなるように分類されてしまうことがある．このような問題に対し (Miyamoto et al., 2008) はクラスターサイズ調整パラメータ α を導入して解決を試みている．この方法はクラスターに含まれる個体の数とクラスターの領域が比例するという考えに基づいている．

クラスターサイズを考慮したファジィ c 平均法の目的関数は

$$F_{rFCMA}(\mathbf{U}, \bar{\mathbf{X}}, \boldsymbol{\alpha}) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 + \lambda^{-1}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\log\frac{u_{ik}}{\alpha_k} \tag{3.23}$$

で表される．各パラメータには以下の制約が課される．

$$\sum_{k=1}^{K} u_{ik} = 1 \ \text{ and } \ 0 \le u_{ik} \le 1, \tag{3.24}$$

$$\sum_{k=1}^{K} \alpha_k = 1 \ \text{ and } \ 0 \le \alpha_k \le 1. \tag{3.25}$$

メンバーシップパラメータ $\mathbf{U}$ の最適解はエントロピー正則化ファジィ c 平均法と同様にラグランジュ乗数法を用いると

$$u_{ik} = \frac{\alpha_k \exp(-\lambda\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2)}{\sum_{l=1}^{K} \alpha_l \exp(-\lambda\|\mathbf{x}_i - \bar{\mathbf{x}}_l\|^2)} \tag{3.26}$$

を得る．

クラスター中心パラメータ $\bar{\mathbf{X}}$ もエントロピー正則化ファジィ c 平均法と同様で

$$\bar{\mathbf{x}}_k = \frac{\sum_{i=1}^{N} u_{ik}\mathbf{x}_i}{\sum_{i=1}^{N} u_{ik}} \tag{3.27}$$

となる．

クラスターサイズ調整パラメータ $\boldsymbol{\alpha}$ はラグランジュ乗数法を用いて最適解を導出する．ラグランジュ乗数を ν とすると，ラグランジュ関数は

$$L(\boldsymbol{\alpha}, \nu) = F_{rFCMA} + \nu(\sum_{k=1}^{K} \alpha_k - 1) \tag{3.28}$$

と書くことができる．

$L(\boldsymbol{\alpha}, \nu)$ を α_k で偏微分すると，

$$\frac{\partial L(\boldsymbol{\alpha}, \nu)}{\partial \alpha_k} = -\lambda^{-1} \sum_{i=1}^{N} \frac{u_{ik}}{\alpha_k} + \nu \tag{3.29}$$

となる．$\partial L(\boldsymbol{\alpha}, \nu)/\alpha_k = 0$ を解いて

$$\alpha_k \nu = \lambda^{-1} \sum_{i=1}^{N} u_{ik} \tag{3.30}$$

を得る．(3.25) を用いると

$$\nu = \lambda^{-1} \sum_{i=1}^{N} \sum_{l=1}^{K} u_{il} \tag{3.31}$$

となるので，これと $\sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik} = N$ を (3.30) に代入すると

$$\alpha_k = \frac{1}{N} \sum_{i=1}^{N} u_{ik} \tag{3.32}$$

を得る．

アルゴリズム全体は以下のようになる．

0. 初期化: パラメータ $\mathbf{U}$, $\bar{\mathbf{X}}$, $\boldsymbol{\alpha}$ を初期化する．正則化パラメータ λ とクラスター数 K を与える．
1. $\mathbf{U}$-step: $\bar{\mathbf{X}}$ と $\boldsymbol{\alpha}$ を所与とし，(3.26) を用いて $\mathbf{U}$ を更新する．
2. $\bar{\mathbf{X}}$-step: $\mathbf{U}$ と $\boldsymbol{\alpha}$ を所与とし，(3.27) を用いて $\bar{\mathbf{X}}$ を更新する．
3. $\boldsymbol{\alpha}$-step: $\mathbf{U}$ と $\bar{\mathbf{X}}$ を所与とし，(3.32) を用いて $\boldsymbol{\alpha}$ を更新する．
4. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

第 4 章

K 平均法における変数選択法の開発

4.1　はじめに

多くの場合，データにはクラスター構造に関係する本質的な変数と，クラスター構造には関係ない変数が混在している．K 平均法はそのようなデータセットを扱うことはできるが，どの変数が本質的で，どの変数が不要であるかということまで判別することはできない．この問題に対処する方法は 2 つ有る．一つは主成分分析などの次元縮約法を用いて，データ構造を抽出する方法である．しかしながら，このアプローチは元々の変数ではなく，合成された成分をもとに結果を解釈しなければならないため，使いにくい面もある．もう一つの方法は変数選択を行うことである．このアプローチは変数の重要性がパラメータの数値として表現されるため解釈が容易である．

回帰分析，判別分析などの外的変数が存在する分析法において，変数選択は活発に研究されている．しかし，クラスタリング分野においてはさほど研究が進んでいない．本章では K 平均法におけるデータ構造に関係の無い変数を検出する方法を提案する．

データ行列を $N \times P$ の $\mathbf{X} = (x_{ij})$ とする．ここで x_{ij} は i 番目の個体 $(i = 1, ..., N)$ の j 番目の変数 $(j = 1, ..., P)$ を意味する要素である．$\mathbf{U} = (u_{ik})$ はメンバーシップ行列を示し，u_{ik} は i 番目の個体がクラスター $k(= 1, ..., K)$ に所属する程度を表している．そして，$\bar{\mathbf{X}} = (\bar{x}_{kj})$ は $K \times P$ のセントロイド行列を表し，クラスターの重心を意味する．

K 平均法の目的は以下の目的関数

$$F(\mathbf{U}, \bar{\mathbf{X}}) = \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ik} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2. \tag{4.1}$$

を最小にすることであった．K 平均法はセントロイドと個体間の距離を計算する際，すべての変数を等しく扱うため，ある変数が他の変数よりも重要であるような場合を扱うことはできなかった．そのため本研究では変数の重要性すなわち重みパラメータ w_j を導入し，変数 j について個体 i とセントロイド k 間の距離を

$$d_{ikj} = w_j(x_{ij} - \bar{x}_{kj})^2 \tag{4.2}$$

と計算する．

重みパラメータの導入によって，データ構造に関係の無い変数を特定することが可能になる．重みパラメータの推定のために，変数選択法の 1 つである LASSO 法 (Tibshirani, 1996) と同様に正則化法を用いるが，本研究では正則化項としてエントロピー項を用いる．すなわち，重みパラメータのエントロピーを K 平均法の目的関数に付け加える．

エントロピーを正則化項として用いる方法は，ファジィメンバーシップを得るためにメンバーシップパラメータのエントロピーを正則化項として K 平均法の目的関数に付け加えるファジィ c 平均法 (Miyamoto & Mukaidono, 1997) として開発された．本研究で導入する重みパラメータは，ファジィ c 平均法におけるメンバーシップパラメータと同様の制約を持つため，重みパラメータに対しても同様のアルゴリズムを適用することができる．重みパラメータの値が 0 に近い値になるということは，そのパラメータに対応する変数はクラスタリングには必要ない変数であると言うことを意味する．

提案方法のすべての重みパラメータが等しい値になった場合，通常の K 平均法と等しくなる．すなわち，通常の K 平均法では重みパラメータが等しく固定されており，提案方法では正則化法によって [0, 1] のファジィな値をとることができるよう拡張していると考えることができる．Huang, Ng, Rong, & Li (2005) は本章で提案する手法と似たアプローチをとり，W-K 平均法を提案している．彼らは Bezdek (1980) が提案したべき乗型ファジィ c 平均法を応用することによって重みパラメータを導入している．

続く節では，提案する方法の目的関数を示し，開発したパラメータ推定アルゴリズムを提示する．さらに提案手法の有用性を人工データと実データの解析を通して検証する．

4.2 提案手法

4.2.1 目的関数

本研究で提案する方法の目的関数は

$$F(\mathbf{U}, \bar{\mathbf{X}}, \mathbf{w}) = \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{P} u_{ik} w_j (x_{ij} - \bar{x}_{kj})^2 + \lambda^{-1} \sum_{j=1}^{P} w_j \log w_j, \tag{4.3}$$

と定義される．ここで $\mathbf{w}$ は w_j を要素として含む $P \times 1$ の変数に対する重みパラメータベクトルである．$\lambda(> 0)$ は正則化パラメータである．右辺第 1 項はクラスター内二乗和を示しており，第 2 項は重みパラメータに対する正則化項である．

メンバーシップパラメータ $\mathbf{U}$ と 重みパラメータ $\mathbf{w}$ には以下の制約がある．

$$\sum_{k=1}^{K} u_{ik} = 1,\ u_{ik} \in \{0, 1\}. \tag{4.4}$$

この制約は個体はただ 1 つだけのクラスターに所属することを意味する．

$$\sum_{j=1}^{P} w_j = 1,\ 0 \leq w_j \leq 1. \tag{4.5}$$

この制約は重みの総量 1 であり，各変数に分配されることを意味する．もしある変数がデータ構造に関係ない場合，対応する重みパラメータが 0 に近い値となって検出することが可能になる．

4.2.2 アルゴリズム

パラメータ $(\mathbf{U}, \bar{\mathbf{X}}, \mathbf{w})$ を推定するために，目的関数 (4.3) を最小化する反復アルゴリズムを開発した．ただし，λ は事前に与えられているとする．各パラメータを求める 3 つのステップを目的関数の値が収束するまで反復を繰り返す．

メンバーシップ $\mathbf{U}$ の最適解は，w_j の部分を除いて通常の K 平均法とよく似ている．重み付き距離の 2 乗が最も小さくなるクラスターへ個体が割り当てられる．

$$u_{ik} = \begin{cases} 1, \text{ if} \displaystyle\sum_{j=1}^{P} w_j (x_{ij} - \bar{x}_{kj})^2 \leq \sum_{j=1}^{P} w_j (x_{ij} - \bar{x}_{lj})^2 \\ 0, \text{ otherwise.} \end{cases} \tag{4.6}$$

セントロイド $\bar{\mathbf{X}}$ の最適解は，通常の K 平均法と同じである．クラスターに所属する個体の平均を求めれば良い．

$$\bar{x}_{kj} = \frac{\sum_{i=1}^{N} u_{ik} x_{ij}}{\sum_{i=1}^{N} u_{ik}} \tag{4.7}$$

重み係数 $\mathbf{w}$ の最適解は，ラグランジュ乗数法を用いた制約付き最適化問題を解くことによって得られる．(4.3) は

$$D_j = \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ik}(x_{ij} - \bar{x}_{kj})^2. \tag{4.8}$$

を用いて

$$F(\mathbf{U}, \bar{\mathbf{X}}, \mathbf{w}) = \sum_{j=1}^{P} w_j D_j + \lambda^{-1} \sum_{j=1}^{P} w_j \log w_j, \tag{4.9}$$

と書き換えることができる．ラグランジュ乗数を α とすると，ラグランジュ関数は

$$\psi(\mathbf{w}, \alpha) = \sum_{j=1}^{P} w_j D_j + \lambda^{-1} \sum_{j=1}^{P} w_j \log w_j + \alpha(\sum_{j=1}^{P} w_j - 1). \tag{4.10}$$

ように書くことができる．ラグランジュ関数 $\psi(\mathbf{w}, \alpha)$ を w_j と α について偏微分すると，

$$\frac{\partial \psi(\mathbf{w}, \alpha)}{\partial w_j} = D_j + \lambda^{-1}(1 + \log w_j) + \alpha \tag{4.11}$$

$$\frac{\partial \psi(\mathbf{w}, \alpha)}{\partial \alpha} = \sum_{j=1}^{P} w_j - 1. \tag{4.12}$$

を得る．$\partial \psi(\mathbf{w}, \alpha)/\partial w_j = 0$ を解いて

$$w_j = \exp(-\lambda D_j) \times \exp(-1 - \lambda\alpha) \tag{4.13}$$

が得られ，$\partial \psi(\mathbf{w}, \alpha)/\partial \alpha = 0$ を解いて

$$\sum_{j=1}^{P} w_j = 1. \tag{4.14}$$

が得られる．(4.13) を (4.14) へ代入することによって

$$\exp(-1-\lambda\alpha) = \left\{\sum_{j=1}^{P}\exp(-\lambda D_j)\right\}^{-1} \tag{4.15}$$

となる．したがって，(4.15) を (4.13) へ代入すると

$$\begin{aligned} w_j &= \frac{\exp(-\lambda D_j)}{\sum_{m=1}^{P}\exp(-\lambda D_m)} \\ &= \frac{\exp(-\lambda\sum_{k=1}^{K}\sum_{i=1}^{N}u_{ik}(x_{ij}-\bar{x}_{kj})^2)}{\sum_{m=1}^{P}\exp(-\lambda\sum_{k=1}^{K}\sum_{i=1}^{N}u_{ik}(x_{im}-\bar{x}_{km})^2)}. \end{aligned} \tag{4.16}$$

が得られる．

アルゴリズム全体は次のようになる．

0. 初期化: パラメータを初期化する．クラスター数 K，および，正則化パラメータ λ を与える．
1. **U**-step: $\bar{\mathbf{X}}$ と $\mathbf{w}$ を所与とし，(4.6) を用いて $\mathbf{U}$ を更新する．
2. $\bar{\mathbf{X}}$-step: $\mathbf{U}$ と $\mathbf{w}$ を所与とし，(4.7) を用いて $\bar{\mathbf{X}}$ を更新する．
3. **w**-step: $\mathbf{U}$ と $\bar{\mathbf{X}}$ を所与とし，(4.16) を用いて $\mathbf{w}$ を更新する．
4. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

4.3　数値実験

提案手法のクラスタリング精度と局所解の頻度を検討するために数値実験を行った．

4.3.1　人工データ生成と解析方法

想定するデータはクラスター構造を持つ部分 ($\mathbf{X}^{(t)}$) と持たない部分 ($\mathbf{X}^{(e)}$) から構成される．形式的には

$$\mathbf{X} = \left[\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_p^{(t)}, \mathbf{x}_1^{(e)}, \ldots, \mathbf{x}_q^{(e)}\right]$$

のように示される．ここで p はクラスター構造を持つ変数の数であり，q はクラスター構造を持たない変数の数である．

クラスターサイズ n_k は [30, 70] の整数一様分布から発生させた．$\mathbf{X}^{(t)}$ は平均ベクトル μ_k，共分散行列 $v(\theta)\mathbf{I}_P$ の P 変量正規分布から生成した．平均ベクトル μ_k は [-10, 10] の実数一様分布から抽出し，各変数の分散 $v(\theta)$ はクラスター間分離度 $\theta = 0.9$，群間平方和 $SSB = \sum_{k=1}^{K} n_k(\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}})^2$ とし，

$$v(\theta) = \frac{SSB \times (1-\theta)}{N\theta}$$

によって決定した．$\mathbf{X}^{(e)}$ は [-10, 10] の実数一様分布から発生させた．全変数の数を $p + q = 10$ とし，$p = 8, 5, 2$ の 3 条件を設けた．クラスター数は $K = 3$ とした．各条件で以上の手続きを 100 回反復し計 300 個のデータを得た．

各条件で発生させたデータを用いて，クラスター数は $K = 3$ と指定して提案手法と通常の K 平均法を適用した．正則化パラメータ λ はそれぞれの条件で $\lambda = 0.0001, 0.0005, 0.001$ とした．初期値を 50 セット用意し，目的関数が最も小さくなったものを解として採用した．

4.3.2　結果と考察

表 5 にクラスタリング精度の指標である Adjusted Rand Index (ARI: Hubert & Arabie, 1985) と局所解の頻度を示す．ARI は 2 つの分割結果間の類似性の指標であり，分割結果が完全に一致したとき上限の 1 をとる．ARI の値は，クラスター構造を持つ変数が少なくなる（p の値が小さくなる）と，提案手法，通常の K 平均法の両者とも減少する．しかし，提案手法は通常の K 平均法ほど分類精度が下がらず，高い精度でクラスタリングできることが示された．局所解の頻度は，クラスター構造を持つ変数が多いとき，提案手法は通常の K 平均法よりも多い．しかし，クラスター構造を持つ変数が少なくなると両手法の差はほとんど無い．局所解が多いことから，提案手法の実用に当たっては複数の初期値のセットを用いる必要がある．以上から，提案手法は局所解が多いにもかかわらず，通常の K 平均法よりもよい ARI の値を示しており，有用なものとして受け入れられる．

表 5　各条件での ARI の値と局所解の頻度

		p=8			p=5			p=2		
		1st	Med	3rd	1st	Med	3rd	1st	Med	3rd
ARI	提案手法	.93	.96	.98	.90	.92	.96	.56	.79	.93
	K 平均法	.92	.96	.98	.60	.91	.94	.02	.33	.46
局所解	提案手法	9	25	37	25.8	37.5	45	38	45	48
	K 平均法	1	4.5	10	3	22.5	41	40.8	46	48

4.4　データ解析

本節では，提案手法の有用性を 2 つのデータセットの解析を通して示す．

4.4.1　アイリス データ

データセットの 1 つ目はアイリスデータセット (Fisher, 1936) を用いる．通常の K 平均法，W-K 平均法 (Huang et al., 2005) と提案手法の比較を行い，変数選択の観点から考察を行う．このデータ行列は 150 個体，4 変数（がく片の長さと幅, 花弁の長さと幅）で構成される．各個体は，セトーサ，バーシクル，バージニカという 3 種類のアヤメから 50 個体ずつサンプルされたものである．つまり，150 個体は 50 個体ずつ 3 群に分類されることが既知である．

方法

アイリスデータの個体は 3 群に分類されることが既知であるので，本データ解析においても，クラスター数を $K = 3$ と設定する．チューニングパラメータである提案手法における λ と W-K 平均法での β はそれぞれ $\lambda = .2, .1, .05$ と $\beta = 2, 3, 4$ のように設定した．初期値を 50 セット用意し，目的関数が最も小さくなったものを解として採用した．

表 6　ARI の値と各変数に対する重み係数の推定値

	提案手法 (λ)			W-K 平均法 (β)			K 平均法
λ/β	.2	.1	.05	2	3	4	
ARI	.89	.89	.89	.87	.89	.89	.73
がく片の長さ	.001	.030	.103	.082	.158	.188	.250
がく片の幅	.094	.226	.275	.191	.238	.247	.250
花弁の長さ	.005	.064	.160	.100	.178	.203	.250
花弁の幅	.900	.680	.462	.627	.425	.362	.250

結果と考察

表 6 は ARI の値と各変数ごとの推定された係数を示している．提案手法の観点から見ると，通常の K 平均法では重み係数が変数の数の逆数に固定されていると見なすことができる．すなわち，アイリスデータセットの場合は，すべての変数に等しく $1/4 = 0.25$ の係数がかかっていると解釈される．ARI に関して提案手法と W-K 平均法はほとんど同じ値を取っており，同程度の分類性能であることが分かる．また，両手法とも通常の K 平均法よりも高い値を示しており，通常の K 平均法よりよい分類ができることが見て取れる．

係数に関して提案手法は W-K 平均法はよりスパースな値を取っている．この性質は，不必要な変数を判別しやすくなるため，変数選択の観点から見ると好ましいものである．

4.4.2　動物評定データ

2 つ目のデータセットは動物特徴評定データ (豊田, 2008) である．ここでは，提案手法の挙動を詳細に検討する．動物特徴評定データセットには 30 種の動物（表 8）に関して 13 の特徴（表 7）について評定されており，その値はその特徴を有するか否かの 2 値で表現されている．

方法

正則化パラメータを $\lambda = 1$ とし，クラスター数 K を 2 から 5 に変化させて提案手法を適用した．初期値を 50 セット用意し，目的関数が最も小さくなったものを解として採用

表 7　各変数に対する係数の推定値．各列で値が大きい上位 K 個まで網掛けを行った．

クラスター数：K	2	3	4	5
小	0.01	0.01	0.02	0.02
中	0.00	0.00	0.00	0.00
大	0.00	0.00	0.00	0.00
2 本足	0.12	0.10	0.08	0.07
4 本足	**0.32**	**0.24**	**0.20**	**0.18**
毛	0.12	0.10	0.08	0.08
蹄	0.00	**0.24**	**0.20**	**0.18**
タテガミ	0.01	0.01	0.01	0.01
羽	**0.32**	**0.24**	**0.20**	**0.18**
狩猟	0.00	0.00	0.01	0.01
走る	0.00	0.00	0.00	0.00
飛ぶ	0.04	0.03	**0.20**	**0.18**
泳ぐ	0.04	0.03	0.02	**0.09**

した．

結果と考察

推定されたクラスタリング結果と係数を表 7，表 8 に示す．クラスター数が $K = 2$ のとき，ほ乳類と鳥類の 2 クラスターに分類された．このとき，「4 本足」と「羽」の係数の値が特に大きくなっている．すなわちほ乳類と鳥類に分類するために，「4 本足」と「羽」の情報を他の変数よりも重視していると解釈できる．

クラスター数が $K = 3$ のとき，蹄有りほ乳類，蹄無しほ乳類，鳥類の 3 クラスターに分類された．「4 本足」と「羽」の他に「蹄」の係数の値が大きくなっている．クラスターを 3 つにするに当たり「蹄」の変数を考慮しほ乳類を細分化したと考えられる．

クラスター数が $K = 4$ のとき，蹄有りほ乳類，蹄無しほ乳類，飛ぶ鳥類，飛ばない鳥類の 4 クラスターに分類された．変数を見ると新たに「飛ぶ」の係数に大きな値が割り振ら

れている．

クラスター数が $K = 5$ のとき，蹄有りほ乳類，蹄無しほ乳類，飛ぶ鳥類，飛ばない鳥類，水鳥の 5 クラスターに分類された．「泳ぐ」に対する係数がわずかではあるが，大きくなっている．

以上のように，係数の値が大きな変数はクラスターを特徴付けるものであり，分類を行う上で重要な変数であると解釈できる．逆に 0 かほとんど 0 と推定されているような係数に対する変数は，分類にはさほど有用な情報を持っていないと分かる．

4.5　クラスターサイズを考慮した変数選択 K 平均法の開発

K 平均法はクラスターに含まれるサンプルの数が等しくなりやすいことが知られている．そのため，クラスターに含まれるサンプルの数が大きく異なるように分類されるべき場合でも，サンプルの数が等しくなるように分類されてしまうことがある．このような問題に対し Miyamoto et al. (2008) はクラスターサイズ調整パラメータ $\boldsymbol{\alpha}$ を導入して解決を試みている．この方法は，クラスターに含まれるサンプルの数と，クラスターの領域が比例するという考えに基づいている．

本章で提案した変数選択をともなった K 平均法も通常の K 平均法同様のサンプルサイズの問題を持つため，Miyamoto et al. (2008) の方法を用いて解決を試みる．すなわち，クラスターサイズ調整機能を有する変数選択をともなった K 平均法の目的関数とそのアルゴリズムを提案する．

クラスターサイズ調整機能を持つ方法には，メンバーシップ値がクリスプとファジィの 2 つが提案されているが，本章は変数に関するパラメータのファジィ化に焦点を当てているため，本節ではクリスプな手法を採用する．しかしながら，クラスターサイズ調整機能は，エントロピー正則化ファジィ c 平均法を基礎とした拡張手法である．したがって，西田 (2011) で提案された，エントロピー正則化による，変数選択をともなったファジィ K 平均法を基にクラスターサイズ調整機能を導入する．

4.5.1　目的関数

データを $N \times P$ の行列 $\mathbf{X} = (x_{ij})$ とする．N と P はそれぞれ個体と変数の数である．クラスター数を K で表し，n_k はクラスター k に含まれる個体数とする．提案手法の目的関

数は

$$F(\mathbf{U}, \bar{\mathbf{X}}, \mathbf{w}, \boldsymbol{\alpha}) = \sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{P} u_{ik} w_j (x_{ij} - \bar{x}_{kj})^2 + \lambda^{-1}\sum_{j=1}^{P} w_j \log w_j - \phi^{-1}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\log\alpha_k \tag{4.17}$$

とあらわせる．

$\mathbf{U} = (u_{ik})$ は個体 i がクラスター k に含まれるか否かを示す $N \times K$ のメンバーシップ行列であり，個体 i がクラスター k に所属する場合は 1，所属しない場合は 0 となる．$\bar{\mathbf{X}} = (x_{kj})$ はクラスター k のセントロイドを示す $K \times P$ の平均行列である．$\mathbf{w}$ は w_j を要素として含む $P \times 1$ の変数に対する重みパラメータベクトルである．$\boldsymbol{\alpha}$ は α_k を要素として含む，クラスター k のクラスターサイズを示すベクトルである．$\lambda > 0$，$\phi > 0$ は正則化パラメータである．

この目的関数に課される制約は 3 つある．

$$\sum_{k=1}^{K} u_{ik} = 1 \text{ and } u_{ik} \in \{0, 1\}, \tag{4.18}$$

$$\sum_{j=1}^{P} w_j = 1 \text{ and } 0 \le w_j \le 1, \tag{4.19}$$

$$\sum_{k=1}^{K} \alpha_k = 1 \text{ and } 0 \le \alpha_k \le 1. \tag{4.20}$$

制約 (4.18) は，個体はいずれかのクラスターに含まれ，かつ 1 つのみのクラスターに含まれることを意味し，(4.19) は変数への重み付けの総和は 1 であり，各変数へ配分することを意味し，(4.20) はクラスターサイズは全体を 1 とした比率として表現されることを示す．

4.5.2　アルゴリズム

目的関数 (4.17) の関数を最小にするパラメータ $\mathbf{U}$，$\bar{\mathbf{X}}$，$\mathbf{w}$，$\boldsymbol{\alpha}$ を求める 4 つのステップを反復する交互最適化アルゴリズムを用いる．それぞれのステップでは以下のように最適解が求まる．

所属度 $\mathbf{U}$ は

$$u_{ik} = \begin{cases} 1, & k = \arg\min \sum_{j=1}^{P} w_j (x_{ij} - \bar{x}_{kj})^2 - \phi^{-1} \log \alpha_k. \\ 0, & \text{otherwise.} \end{cases} \tag{4.21}$$

によって得られる．

セントロイド行列 $\bar{\mathbf{X}}$ は，

$$\bar{x}_{kj} = \frac{\sum_{i=1}^{N} u_{ik} x_{ij}}{\sum_{i=1}^{N} u_{ik}}. \tag{4.22}$$

によって得られる．

重み係数 $\mathbf{w}$ は，

$$w_j = \frac{\exp(-\lambda \sum_{k}^{K} \sum_{i}^{N} u_{ik} (x_{ij} - \bar{x}_{kj})^2)}{\sum_{j=1}^{P} \exp(-\lambda \sum_{k}^{K} \sum_{i}^{N} u_{ik} (x_{ij} - \bar{x}_{kj})^2)}. \tag{4.23}$$

によって得られる．

クラスターサイズ調整パラメータ $\boldsymbol{\alpha}$ は

$$\alpha_k = \frac{n_k}{N}. \tag{4.24}$$

により得られる．

アルゴリズム全体は以下のようになる．

0. 初期化: パラメータを初期化する．クラスター数 K，および，正則化パラメータ λ, ϕ を与える．
1. **U**-step: $\bar{\mathbf{X}}$，$\mathbf{w}$，および，$\boldsymbol{\alpha}$ を所与とし，(4.21) を用いて $\mathbf{U}$ を更新する．
2. $\bar{\mathbf{X}}$-step: $\mathbf{U}$，$\mathbf{w}$，および，$\boldsymbol{\alpha}$ を所与とし，(4.22) を用いて $\bar{\mathbf{X}}$ を更新する．
3. **w**-step: $\mathbf{U}$，$\bar{\mathbf{X}}$，および，$\boldsymbol{\alpha}$ を所与とし，(4.23) を用いて $\mathbf{w}$ を更新する．
4. $\boldsymbol{\alpha}$-step: $\mathbf{U}$，$\bar{\mathbf{X}}$，および，$\mathbf{w}$ を所与とし，(4.24) を用いて $\boldsymbol{\alpha}$ を更新する．
5. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

4.5.3　数値実験

提案手法の有効性を確認するため人工データによる，数値実験を行う．

人工データ生成と解析方法

クラスターサイズが偏った 2 クラスターを想定する．ここでもクラスター構造を持つ部分 $\mathbf{X}^{(t)}$ と，クラスター構造を持たない部分 $\mathbf{X}^{(e)}$ を考える．$\mathbf{X}^{(t)}$ の部分において，クラスター A は平均 (0,0)，分散 (0.5,0.5) を持つ 2 変量正規分布から 200 個体を発生させ，クラスター B は平均 (1.5,0)，分散 (0.2,0.2) を持つ 2 変量正規分布から 100 個体を発生させた．$\mathbf{X}^{(e)}$ として，区間 $[-1,2]$ の連続一様分布から発生させたクラスター構造を持たない 2 変数分をクラスター構造を持つ $\mathbf{X}^{(t)}$ に付け加えた．形式的には

$$\mathbf{X} = \left[\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \mathbf{x}_1^{(e)}, \mathbf{x}_2^{(e)}\right]$$

と書ける．4 変数のうち 2 変数はクラスター構造を持ち，残りの 2 変数はクラスター構造を持たないデータとなっている．従って，解析対象は 300 個体，4 変数からなるデータセットである．分析対象となるデータの散布図を図 3 に示す．

K 平均法，変数選択をともなった K 平均法，サイズ調整機能付き変数選択 K 平均法の 3 手法を発生させたデータセットに適用した．変数選択をともなった K 平均法のチューニングパラメータは $\lambda = 0.01$，サイズ調整機能付き変数選択 K 平均法のチューニングパラメータは $\lambda = 0.01, \phi = 10$ と設定した．

ARI を用いて手法の精度を評価する．ARI は上限 1 を取る指標で 1 に近ければ近いほどクラスタリング性能が良いことを意味する．

結果と考察

各手法の ARI の値を表 9 に示す．サイズ調整機能付き変数選択 K 平均法が最も高い値を示しており，かつ上限の 1 にも近く，良い精度でクラスタリングできていることを示している．図 4，5，6 はクラスタリング結果を色で表現した散布図である（ただし，クラスター構造に関係する 2 変数のみを取り出している）．

通常の K 平均法では，変数の重要性や，クラスターサイズが考慮されないため，小さいクラスターの領域が過大になっている．変数選択をともなった K 平均法では，表 9 に示

されているように，クラスター構造を持つ第 1 変数と第 2 変数への係数の値 w が大きくなっており，変数の重要性を考慮した結果となっている．通常の K 平均法よりも改善されているが，クラスターサイズが考慮されないため，まだ適切とは言えないクラスター領域となっている．

今回提案したサイズ調整機能付き変数選択 K 平均法は，変数の重要性や，クラスターサイズが考慮されるため，真のクラスター構造をほぼ特定できた．変数への重み係数 w は変数選択 K 平均法とサイズ調整変数選択 K 平均法では同じ傾向の値を示した．真のクラスターサイズの比は 1:2 であるが，その推定値であるクラスターサイズ調整パラメータ α の比は 1:1.88 となっており，良い結果であった（表 9）．

4.6　まとめ

正則化法を用いた変数選択をともなった K 平均法を提案し，パラメータを推定するための反復アルゴリズムを開発した．数値実験から提案手法は通常の K 平均法よりもよい成績を示すことが分かった．しかしながら，提案手法は局所解が多いことも示された．そのため，提案手法を用いる際には，通常の K 平均法で用いるよりも多くの初期値のセットを用いて，解を求めることが求められる．

アイリスデータの解析結果から，提案手法は K 平均法よりもよい成績であり，類似手法である W-K 平均法と同程度の分類性能を持つことが示された．係数の推定では提案手法は W-K 平均法よりもスパースな値を求めることが可能であり，変数選択を目的とした場合，望ましい結果を得られることが示された．動物評定データの解析を通じて係数の値が大きい変数がクラスターを特徴付けていることが示された．このことは係数の値が大きい変数のみに注目して解釈を行えばよいことを意味しており，通常の K 平均法を用いた場合よりも結果の解釈が容易になる．

変数選択を実行する K 平均法は K 平均法を基礎としているため，K 平均法が持つ問題点をそのまま引き継いでしまう．そのため，クラスターに含まれる個体の数が等しくなりやすい．そこでクラスターサイズ調整パラメータを導入し問題の解決を試みた．数値実験の結果，クラスターサイズに偏りがあるデータセットに対しては，通常の K 平均法や変数選択を実行する K 平均法よりも良い成績を示すことが明らかとなった．

提案手法の適用に当たっては，正則化パラメータを決定する必要があるが，決定的な方

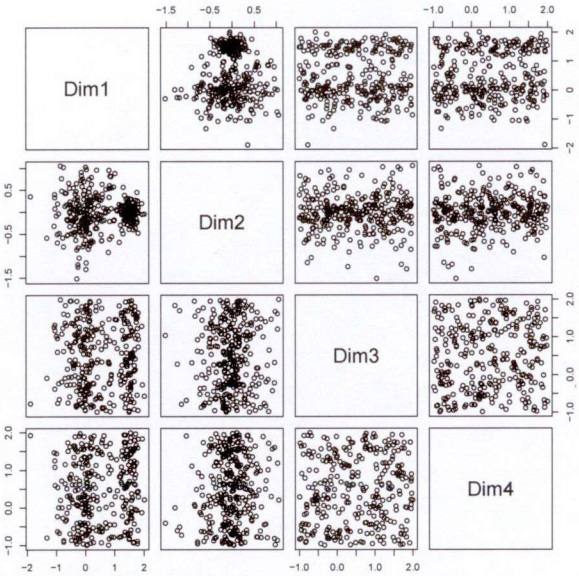

図 3　人工データ

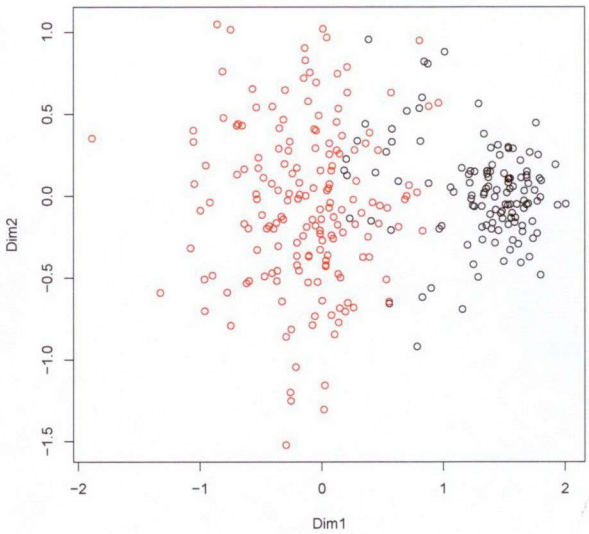

図 4　K 平均法

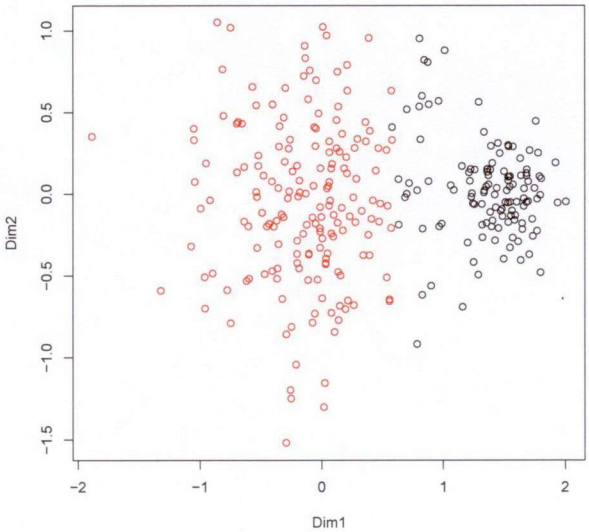

図 5　変数選択 K 平均法

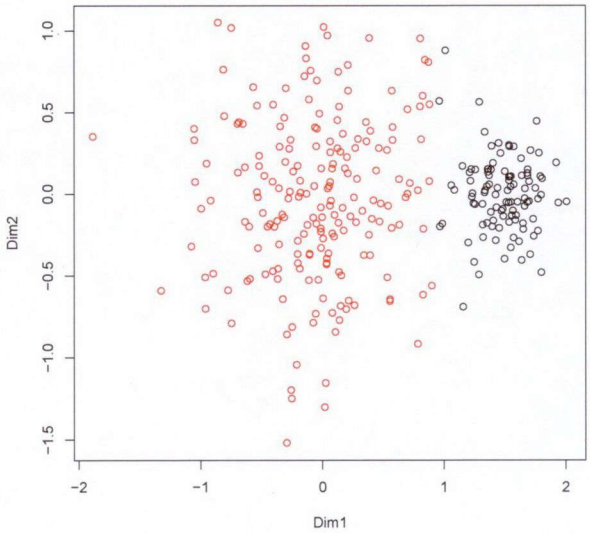

図 6　サイズ調整変数選択 K 平均法

法はまだない．そのため，分析者がパラメータの値を変化させて，分析結果を見ながら決定する必要がある．

表 8　クラスター数 $K = 2, 3, 4, 5$ でのクラスタリング結果

クラスター数：K	2		3			4				5				
	A	B	A	B	C	A	B	C	D	A	B	C	D	E
イヌ	1		1			1				1				
オオカミ	1		1			1				1				
キツネ	1		1			1				1				
クマ	1		1			1				1				
サル	1		1			1				1				
タヌキ	1		1			1				1				
チーター	1		1			1				1				
トラ	1		1			1				1				
ネコ	1		1			1				1				
ハイエナ	1		1			1				1				
ヒョウ	1		1			1				1				
ライオン	1		1			1				1				
イノシシ	1			1			1				1			
ウシ	1			1			1				1			
ウマ	1			1			1				1			
シカ	1			1			1				1			
シマウマ	1			1			1				1			
ゾウ	1			1			1				1			
ブタ	1			1			1				1			
ペンギン		1			1			1				1		
ダチョウ		1			1			1				1		
ニワトリ		1			1			1				1		
カナリヤ		1			1				1				1	
スズメ		1			1				1				1	
タカ		1			1				1				1	
フクロウ		1			1				1				1	
ワシ		1			1				1				1	
ハト		1			1				1				1	
アヒル		1			1				1					1
ガチョウ		1			1				1					1

表 9　各手法の ARI 値とパラメータの推定値

	K 平均法	変数選択 K 平均法	サイズ調整変数選択 K 平均法
ARI	0.66	0.70	0.95
$w1$		0.446	0.430
$w2$		0.409	0.420
$w3$		0.073	0.076
$w4$		0.072	0.074
$\alpha 1$			0.347
$\alpha 2$			0.653

第 5 章

データ行列の最小二乗近似による ファジィクラスタリング法の開発

5.1　はじめに

K 平均法 (MacQueen, 1967) はクラスタリング手法のうち，最もよく用いられる方法の 1 つである．その目的関数は以下のように書くことができる．

$$F(\mathbf{U}, \bar{\mathbf{X}}) = \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ik} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 \tag{5.1}$$

ここで，$\mathbf{X}$ はデータ，$\mathbf{U}$ は所属度を示すメンバーシップ，$\bar{\mathbf{X}}$ はクラスタセントロイドである．通常の K 平均法において，個体がクラスターに所属するか否かは，2 値のメンバーシップパラメータによって表現され，あいまいさは無い．

ファジィ c 平均法はクラスターへの所属度にあいまいさを認め，0 以上 1 以下の連続値をとるメンバーシップパラメータによって所属度の度合いを表現しようとする方法である．ファジィパラメータ α を導入した，べき乗型ファジィ c 平均法 (Bezdek, 1980) や，正則化法を用いたエントロピー正則化ファジィ c 平均法 (Miyamoto & Mukaidono, 1997) は，K 平均法の目的関数に手を加え，それぞれ

$$F(\mathbf{U}, \bar{\mathbf{X}}) = \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ik}^{\alpha} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 \tag{5.2}$$

$$F(\mathbf{U}, \bar{\mathbf{X}}) = \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ik} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 + \lambda^{-1} \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ik} \log u_{ik} \tag{5.3}$$

のように目的関数を定義し，ファジィメンバーシップの推定を可能としていた．目的関数を変形させなければならないのは，元々の K 平均法の目的関数も，制約条件も線形のため，最適解が制約条件の端点として与えられるため，メンバーシップパラメータがファジィになり得ないためである．

K 平均法の目的関数は行列ベースで表記できる (e.g. 足立, 2011)．本章では目的関数を

$$F(\mathbf{U}, \bar{\mathbf{X}}) = \|\mathbf{X} - \mathbf{U}\bar{\mathbf{X}}\|^2 \tag{5.4}$$

のように行列を用いて書き，これを元に議論を進める．提案手法は，行列表記の目的関数を用いて，メンバーシップパラメータをリパラメトライズすることにより，目的関数にファジィパラメータや正則化項を加えることなく，直接ファジィなメンバーシップを推定する．すなわち，データ行列を 2 つのパラメータ行列の積によって近似する，ファジィクラスタリング法を提案する．ファジィ c 平均法ではファジィネスを調節するチューニングパラメータを事前に設定しなくてはならないが，提案手法には，そのようなチューニングパラメータは存在せず，分析者が決める必要がない．

5.2　提案手法

5.2.1　目的関数

$\mathbf{X}$ を n(個体) $\times$ m(変数) のデータ行列とする．データ行列を n(個体) $\times$ c(クラスター) のメンバーシップ行列 $\mathbf{F}$ と m(変数) $\times$ c(クラスター) の係数行列 $\mathbf{A}$ の積によって近似することを考え，目的関数

$$F(\mathbf{F}, \mathbf{A}) = \|\mathbf{X} - \mathbf{F}\mathbf{A}'\|^2 \tag{5.5}$$

を最小化することを目的とする．ここで $\mathbf{F} = (f_{ik})$ は次の制約を満たす．

$$\sum_{k=1}^{c} f_{ik} = 1,\ f_{ik} \in [0, 1]. \tag{5.6}$$

すなわち，個体 i がクラスター k に所属する程度は，0 以上 1 以下の連続値によって与えられ，所属度の合計は 1 となる．

5.2.2　アルゴリズム

パラメータ $\mathbf{A}$ を推定するステップと，$\mathbf{F}$ を推定するステップを繰り返す交互最小二乗法を用いる．

A-ステップ

$\mathbf{A}$ には制約がないため，その推定値は回帰問題を解くことによって，次のように得られる．

$$\mathbf{A}' = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{X} \tag{5.7}$$

F-ステップ

$\mathbf{F}$ の推定では，目的関数を

$$\sum_{i=1}^{N} \|\mathbf{x}'_i - \mathbf{f}'_i\mathbf{A}'\|^2 \tag{5.8}$$

のように書き換え，個体 i ごとに $\mathbf{f}_i$ を求める．

制約 (5.6) を満たすため，$\mathbf{F} = (f_{ik})$ を

$$f_{ik} = \frac{g_{ik}^2}{\sum_{l=1}^{c} g_{il}^2} \tag{5.9}$$

のようにリパラメトライズし，$\|\mathbf{x}'_i - \mathbf{f}'_i\mathbf{A}'\|^2$ を最小にする $\mathbf{f}_i$ を求める．

アルゴリズム全体は以下のようになる．

0. 初期化: パラメータを初期化する．クラスター数 c を与える．
1. **A**-step: $\mathbf{F}$ を所与とし，(5.7) を用いて $\mathbf{A}$ を更新する．
2. **F**-step: $\mathbf{A}$ を所与とし，$\|\mathbf{x}'_i - \mathbf{f}'_i\mathbf{A}'\|^2$ を最小にする $\mathbf{f}_i$ を求める．
3. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

表 10　提案手法と主成分分析の制約条件

制約条件	提案手法	主成分分析
$\mathrm{rank}\,\mathbf{FA}' \le p \le r = \mathrm{rank}\,\mathbf{X}$	×	○
$\sum_{k=1}^{c} f_{ik} = 1,\ f_{ik} \in [0,1]$	○	×

5.3　関連手法

本提案手法は K 平均法を直接ファジィ拡張を行った手法であるが，主成分分析とも関連がある．両手法とも目的関数は

$$F(\mathbf{F},\mathbf{A}) = \|\mathbf{X} - \mathbf{FA}'\|^2$$

と表される．両手法の定式化はデータ行列を 2 つのパラメータ行列の積によって近似するという点において全く同じである．

表 10 に示すように，両手法の異なる点は制約条件である．提案手法はクラスタリング手法としての制約のみが課されている．すなわち，$\mathbf{F}$ の行和は 1 になり，f_{ik} は 0 以上 1 以下の値を取るという制約で，主成分分析には無いものである．

それに対して，主成分分析ではデータ行列の低階数近似が目的であるから，2 つのパラメータ行列の積はデータ行列よりも階数が小さいという制約が課される．この制約は提案手法には課されない．パラメータ $\mathbf{F}$ はクラスタリング手法ではメンバーシップと呼ばれ，主成分分析では成分得点と呼ばれる．

5.4　数値実験

提案手法の挙動を確認し，パラメータが解釈可能か検討を行う．

5.4.1　人工データ生成と解析方法

3 つのクラスターを想定し，ユニークな平均と分散共分散

$$\mu_1 = (0,2),\ \mu_2 = (-2,-2),\ \mu_3 = (2,-2)$$

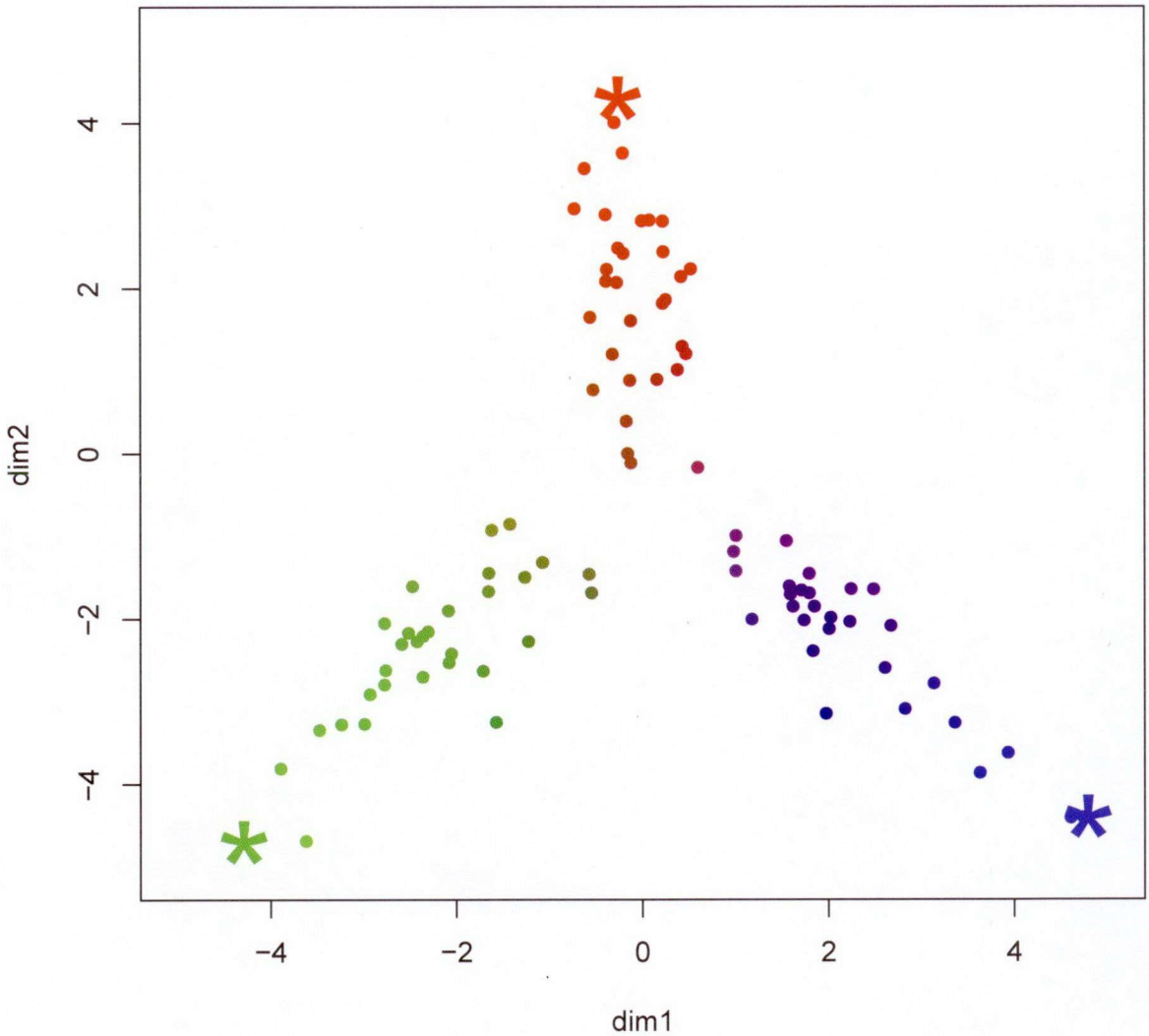

図 7　提案手法によって推定されたメンバーシップ値を用いて色分けを行った人工データの散布図

$$\Sigma_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 1 \end{bmatrix},\ \Sigma_2 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix},\ \Sigma_3 = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

を持つ 3 つの 2 変量正規分布から各 30 の個体を発生させた（図 7，8）．すなわち，分析対象となるデータセットは 90(個体) × 2(変数) のデータ行列である．

クラスター数 $c = 3$ として提案手法を適用した．また比較対象としてべき乗型ファジィ c 平均法も同じ人工データに対して適用した．このとき，ファジィパラメータは $\alpha = 3$ と設定した．

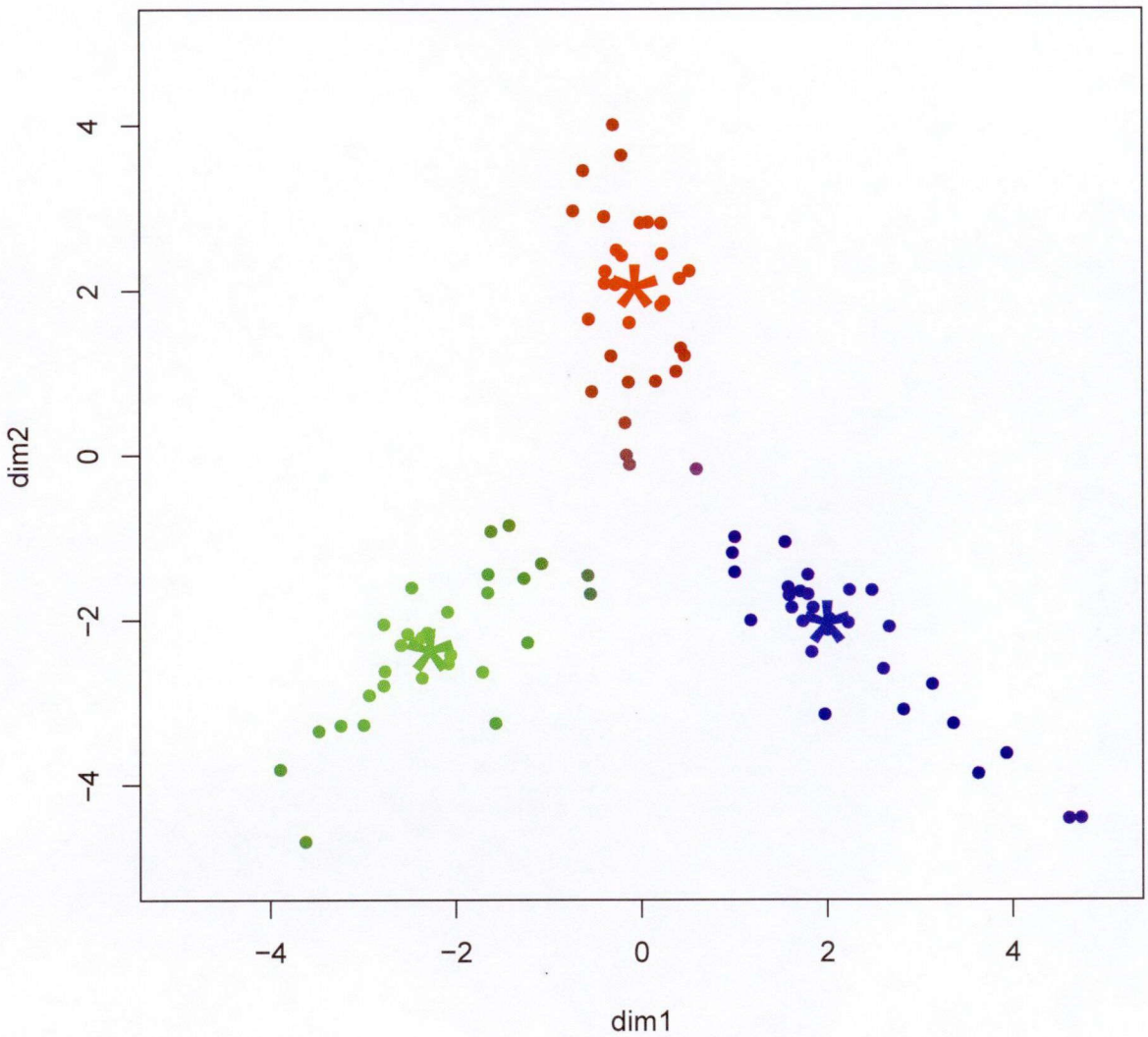

図 8　ファジィ c 平均法によって推定されたメンバーシップ値を用いて色分けを行った人工データの散布図

5.4.2　結果と考察

提案手法によって推定されたメンバーシップパラメータ $\mathbf{F}$ の値を赤，緑，青の 3 色に変換して散布図の個体に色づけを行ったものを図 7 に示す．色が明るいほど，特定のクラスターに対するメンバーシップ値が高いことを示しており，色が暗いほど，複数のクラスターにメンバーシップ値が分配されていることを示す．この図においては，原点付近に位置する個体ほど混色されて色が暗くなっており，個体の所属が曖昧になっていることがわ

かる．一方，布置の周辺に位置する個体ほど色が明るくなっており，特定のクラスターのみに所属することを示している．

べき乗型ファジィ c 平均法によって推定されたメンバーシップパラメータ $\mathbf{U}$ の値を赤，緑，青の 3 色に変換して散布図の個体に色づけを行ったものを図 8 に示す．この図においても，提案手法と同様に原点付近に位置する個体ほど色が暗くなっており，個体の所属が曖昧になっていることを示す．しかし，布置の周辺に位置する個体も色が暗くなっており，所属度が低くなっている．

両者の違いはもう一つのパラメータ，提案手法における係数行列 $\mathbf{A}$ とべき乗型ファジィ c 平均法におけるセントロイド $\bar{\mathbf{X}}$ から考察できる．図 7，8 において，提案手法で推定された $\mathbf{A}$ とべき乗型ファジィ c 平均法で推定された $\bar{\mathbf{X}}$ を「*」によってプロットした．

べき乗型ファジィ c 平均法において，セントロイド $\bar{\mathbf{X}}$ はメンバーシップによって重みづけられたクラスター平均であり，セントロイドに近い個体ほど大きなメンバーシップ値を取る．そのため，クラスターが隣接する箇所ではなくとも，所属クラスター中心から離れていれば，所属度が下がってしまうことがある．

一方，提案手法においては係数行列 $\mathbf{A}$ はクラスター中心ではなく，クラスターに所属する個体の分布方向を示す．そのため，個体のメンバーシップ値は，自身から最も近いクラスターを除いた他のクラスターから離れていればいるほど，所属クラスターへの所属度が高くなる．

提案手法はべき乗型ファジィ c 平均法とは異なるメンバーシップ値を与えるが，このような結果の方が解釈しやすい場合もある．したがって，提案手法はファジィクラスタリング法として，有用であると考えられる．

5.5　データ解析

提案手法と主成分分析を比較しながら，その類似点と相違点について検討を行う．

5.5.1　データと解析方法

成績データ (田中・垂水, 1995) に対して，提案手法を適用する．このデータは高校生 50 人の 6 科目の試験成績である．変数は国語，英語，社会，数学 I，数学 II，理科の 6 科目となっている．変数ごとに標準化を行い，クラスター数 $c = 2, 3$ として提案手法を適用し

表 11　主成分分析によって得られた主成分負荷量

	成分 1 (負の総合)	成分 2 (文理)
国語	-0.54	-0.58
英語	-0.62	-0.45
社会	-0.59	-0.51
数学 I	-0.80	0.36
数学 II	-0.85	0.28
理科	-0.73	0.51

た．初期値を 50 セット用意し，目的関数が最も小さくなったものを解として採用した．

同様のデータに対し主成分分析を適用したところ，固有値が 2.919, 1.263, 0.672, 0.585, 0.363, 0.198 のように得られた．値が 1 以上であることと値の減衰を考慮して第 2 主成分までを採用した．なお，解の回転は行っていない．

5.5.2　結果と考察

はじめに，主成分分析の結果から解釈を行う．表 11 に示す成分負荷量を見ると，第 1 主成分負荷量は全ての変数で負の大きな絶対値となっている．したがって，第 1 主成分は総合能力と解釈できる．第 2 主成分負荷量は理系科目に対して正の大きな絶対値，文系科目に対して負の大きな絶対値を持つ．したがって，第 2 主成分は文系と理系を示す軸であると解釈できる．次からは，提案手法の結果を主成分分析の結果と比較しながら解釈を行っていく．

クラスター数 $c = 2$ のとき

提案手法によって推定されたパラメータ **A** の値を表 12 に示す．クラスター 1 に関する係数は全て正の値，クラスター 2 に関する係数は全て負の値となっており，それぞれ正の総合能力，負の総合能力を示していると考えられる．したがって，個体は，総合能力が高いクラスターと低いクラスターに分類されていると考えられ，パラメータ **F** は個体がクラスターに所属する程度と解釈できる．

表 12　提案手法で推定されたパラメータ **A** の値 (c = 2)

	C 1 (正の総合)	C 2 (負の総合)
国語	1.17	-1.09
英語	1.36	-1.26
社会	1.28	-1.19
数学 I	1.74	-1.62
数学 II	1.85	-1.72
理科	1.59	-1.48

表 13　提案手法 (c = 2) によって得られた **F** と主成分分析によって得られた主成分得点との相関係数

	成分 1 (負の総合)	成分 2 (文理)
$\mathbf{f}_1$ (正の総合)	-1.00	0.00
$\mathbf{f}_2$ (負の総合)	1.00	0.00

提案手法を主成分分析のように解釈するならば，**A** はクラスターと変数の関連の強さを，**F** はクラスターに関する個体得点を示していると考えられる．提案手法には **F** の行和が 1 という制約があるため，c = 2 の場合，クラスター 1 とクラスター 2 は真逆の特性を持つと考えられる．すなわち，実質的には 1 次元の特性を表と裏から見ているのと同じである．この性質は，**F** が相関係数において第 1 主成分得点と一致することからも示唆される（表 13）．

クラスター数 c = 3 のとき

提案手法によって推定されパラメータ **A** の値を，表 14 に示す．クラスター 1 に関する係数は，全ての科目で負の大きな値を示しており総合能力を示していると考えられ，総合能力が低いクラスターであると解釈できる．クラスター 2 に関する係数は，理系科目に大きな値，文系科目に小さな値を示しており，理系科目が得意なクラスターであると考えら

表 14　提案手法で推定されたパラメータ **A** の値 (c = 3)

	C 1 (負の総合)	C 2 (理系)	C 3 (文理)
国語	-2.74	0.28	2.73
英語	-2.80	0.88	2.16
社会	-2.79	0.62	2.42
数学 I	-2.03	3.50	-1.46
数学 II	-2.33	3.47	-1.10
理科	-1.55	3.66	-2.17

表 15　提案手法 (c = 3) によって得られた **F** と主成分分析によって得られた主成分得点との相関係数

		成分 1 (負の総合)	成分 2 (文理)
$\mathbf{f}_1$	(負の総合)	0.92	0.40
$\mathbf{f}_2$	(理系)	-0.87	0.49
$\mathbf{f}_3$	(文理)	-0.13	-0.99

れる．クラスター 3 に関する係数は，文系科目に正の大きな値，理系科目に負の大きな値を示しており，文系科目が得意で理系科目が苦手なクラスターであると解釈できる．

表 15 に，提案手法によって得られたパラメータ **F** と，主成分分析によって得られた成分得点との相関係数を示す．$\mathbf{f}_1$ は特に成分得点 1 と相関が高い．$\mathbf{f}_2$ は成分得点 1 と相関が高いが，成分得点 2 とも相関がある．$\mathbf{f}_3$ は成分得点 2 と非常に高い相関がある．成分 1 は負の総合能力，成分 2 は文理能力であったため，先ほどのクラスター 1 が負の総合能力，クラスター 2 が理系能力，クラスター 3 が文理能力を示しているという解釈と整合する．

以上より，提案手法のパラメータの解釈は，主成分分析による解釈と整合的であることが示された．

5.6 まとめ

データ行列を 2 つのパラメータ行列の積によって近似する新たな方法を提案した．また，提案手法の目的関数を最小化するアルゴリズムを提案した．提案手法は制約は異なるが，主成分分析と同じ定式化を行っている．そのため，K 平均法と主成分分析の両方の特徴を持った中間的手法と位置づけられる．c の値（クラスター数）は事前に決めなければならいのは K 平均法や主成分分析と同様であるが，その取り得る範囲は K 平均法と同様 $1 \leq c \leq n$ であり，主成分分析とは異なる．

数値実験とデータ解析により，次の 2 点が示された．1 つは，パラメータ $\mathbf{F}$ は 0 から 1 の値を取るという制約が有るため，ファジィクラスタリングにおけるメンバーシップとして解釈できる．もう 1 つは，パラメータ $\mathbf{A}$ は K 平均法と異なりクラスター重心とはならず，主成分分析における負荷量のようなクラスターと変数の関連の強さを示す値となっている．以上より，提案手法はファジィクラスタリング手法であると同時に，変数の数である m 次元のデータ空間に，クラスター数である c 本の軸を構成し，個体を 0 から 1 の間で得点化する手法であると考えることができる．

第 6 章

次元縮約をともなうファジィ K 平均法の開発

6.1　はじめに

個体のクラスタリングを行うにあたって，いくつかの変数はデータに内在するクラスター構造には関わり無いと考えられる場合や，変数の数が多い場合，はじめに主成分分析を行い主成分得点を求め，その主成分得点の分類を行うという方法が取られることがある．しかし，このような 2 段階の手続きは「タンデム」な方法と呼ばれ批判も多い (Arabie & Hubert, 1994; De Soete & Carroll, 1994; Vichi & Kiers, 2001)．主成分分析を用いて次元を縮約する際に，クラスター構造に関する情報が欠落するため，主成分得点を分類しても，真のクラスター構造を推定できない可能性が指摘されている．タンデムな方法を用いることで起こる問題を解決する方法として，次元縮約とクラスタリングを 1 つの基準の最小化によって達成する方法が開発されている (De Soete & Carroll, 1994; Vichi & Kiers, 2001).

本章では，次元縮約とクリスプなクラスタリングを同時に達成する次元縮約 K 平均法 (De Soete & Carroll, 1994) をファジィなメンバーシップパラメータを得られるよう拡張する．メンバーシップをファジィ化することで得られるメリットはいくつかある．通常の K 平均法はメンバーシップパラメータが離散的であるため，局所解が多いことが知られている．パラメータが連続値になることによって，局所解の頻度を抑えることができる (Heiser & Groenen, 1997; Hruschka, 1986)．さらに，クリスプなクラスタリングよりも適合度が高くなると考えられる (Hwang, Dillon, & Takane, 2010)．また，パラメータの変

化が漸進的であるため計算上安定している (McBratney & Moore, 1985). 何より，クラスター間の境界が明確でないときファジィメンバーシップはクリスプメンバーシップよりも実用的である (Bezdek, 1981).

本章ではエントロピー正則化法 (Miyamoto & Mukaidono, 1997) を用いてファジィ化を試みる．K 平均法をファジィ化する方法として Bezdek (1981) によるべき乗型ファジィクラスタリングが提案されているが，本章でエントロピー正則化法によるファジィ化のアプローチを取る理由は，メンバーシップ値の解釈可能性を考慮するためである (Suk & Hwang, 2010). Bezdek (1981) によるべき乗型ファジィクラスタリングもエントロピー正則化ファジィクラスタリングもメンバーシップ値は 0 以上 1 以下で与えられ，その値が大きいほどクラスターに所属する程度が高いと解釈される．しかし，べき乗型ファジィクラスタリングでの所属する程度というのは何を指すのかが不明瞭である (Laviolette, Seaman, Barrett, & Woodall, 1995). それに対し，エントロピー正則化ファジィクラスタリングにおける所属する程度は，個体がクラスターから発生する事後確率として解釈できる (Tran & Wagner, 2000).

本章の構成は以下のようになっている．まず，次節で提案手法のベースとなる次元縮約 K 平均法を紹介する．続く節で，提案手法の目的関数を提示し，パラメータを推定するための交互最適化アルゴリズムを示す．提案手法とそのアルゴリズムの挙動を検討する数値実験を行った後，提案手法の有用性を示すデータ解析結果を示す．

6.2　次元縮約 K 平均法

本節ではクラスラリングと次元縮約を同時に達成する次元縮約 K 平均法を紹介する．まずはじめに本章で使用する，記号を整理しておく．多変量データ行列を N(個体)$\times P$(変数) の行列 $\mathbf{X} = (x_{ij})$ で，個体 i がどの程度クラスター k へ所属するかを示すメンバーシップ行列を P(変数)$\times K$(クラスター) の行列 $\mathbf{U} = (u_{ik})$ で，P 次元の入力空間におけるクラスター k の平均ベクトルを $\bar{\mathbf{x}}_k = 1/N_k \sum_{i=1}^{N} u_{ik}\mathbf{x}_i$ と表す．ただし，$N_k = \sum_{i=1}^{N} u_{ik}$ はクラスター k に所属する個体 i の個数である．

次元縮約 K 平均法の目的関数は

$$F_{REDKM}(\mathbf{C},\mathbf{U}) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|\mathbf{x}_i - \mathbf{c}_k\|^2 \tag{6.1}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 + \sum_{k=1}^{K} N_k\|\bar{\mathbf{x}}_k - \mathbf{c}_k\|^2 \tag{6.2}$$

と表される．この関数を最小にするパラメータ $\mathbf{C}$ と $\mathbf{U}$ を求める．ここで $\mathbf{C}$ は階数が R のセントロイド行列である．すなわち $\mathbf{C}$ に関して

$$\mathrm{rank}(\mathbf{C}) = R \tag{6.3}$$

の制約を課す．これは，セントロイド行列が R 次元に縮約されることを示している．

メンバーシップ行列 $\mathbf{U}$ に関して

$$u_{ik} \in \{0, 1\} \tag{6.4}$$

$$\sum_{k=1}^{K} u_{ik} = 1 \tag{6.5}$$

の制約を課す．すなわち，個体はいずれか 1 つのクラスターのみに所属しなければならない．

目的関数の最適化には $\mathbf{U}$ を最適化するクラスタリングパートと，$\mathbf{C}$ を最適化する次元縮約パートの 2 つのフェイズを繰り返し行うアルゴリズムを用いる．クラスタリングパートでは (6.4)，(6.5) の制約もとで (6.1) を最小にする $\mathbf{U}$ を求める．これは K 平均法のアルゴリズムによって求められる．

$$u_{ik} = \begin{cases} 1 & \text{if } \|\mathbf{x}_i - \mathbf{c}_k\|^2 < \|\mathbf{x}_i - \mathbf{c}_l\|^2 \text{ for } l = 1, ..., K \text{ and } l \neq k, \\ 0 & \text{otherwise.} \end{cases} \tag{6.6}$$

次元縮約パートでは (6.3) の制約のもとで (6.2) を最適にする $\mathbf{C}$ を求める．ここで (6.2) の右辺第 2 項を

$$G(\mathbf{C}) = \|\mathbf{W}^{\frac{1}{2}}(\bar{\mathbf{X}} - \mathbf{C})\|^2 \tag{6.7}$$

とする．ただし，$\mathbf{W}$ は対角要素にクラスターサイズ N_k が並んだ対角行列，

$$\mathbf{W} = \mathrm{diag}(N_1, ..., N_K) \tag{6.8}$$

である．ここで，$\mathbf{U}_R\mathbf{S}_R\mathbf{V}_R'$ を $\mathbf{W}^{\frac{1}{2}}\bar{\mathbf{X}}$ の特異値分解の上位 R 個の特異値と対応する特異ベクトルとすると

$$\hat{\mathbf{C}} = \mathbf{W}^{-\frac{1}{2}}\mathbf{U}_R\mathbf{S}_R\mathbf{V}_R' \tag{6.9}$$

を得る．

次元縮約 K 平均法は個体のクラスタリングと次元縮約を同時に行う手法である (De Soete & Carroll, 1994)．K 平均法などと同様にあらかじめ，個体をいくつのクラスターに分類するか指定しなくてはならない．またいくつの次元に縮約するのかも事前に指定すべきパラメータである．アルゴリズム全体は次のようになる．

0. 初期化: パラメータを初期化する．クラスター数 K，階数 R を与える．
1. **U**-step: **C** を所与とし，(6.6) を用いて **U** を更新する．
2. **C**-step: **U** を所与とし，(6.9) を用いて $\hat{\mathbf{C}}$ を更新する．
3. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

6.3 提案手法

本節では次元縮約 K 平均法をファジィ拡張した手法を提示する．提案手法は次元縮約 K 平均法の目的関数にエントロピー項を加える正則化によって，ファジィネスを導入する (宮本・馬屋原・向殿, 1998).

6.3.1 目的関数

提案手法の目的関数は

$$F_{f\text{-}REDKM}(\mathbf{C},\mathbf{U}) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|\mathbf{x}_i - \mathbf{c}_k\|^2 + \lambda^{-1}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\log u_{ik} \tag{6.10}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 + \sum_{k=1}^{K} N_k\|\bar{\mathbf{x}}_k - \mathbf{c}_k\|^2 + \lambda^{-1}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\log u_{ik} \tag{6.11}$$

とあらわせる．ここで $\lambda > 0$ は正則化パラメータである．この関数を最小にするパラメータ $\mathbf{C}$ と $\mathbf{U}$ を求める．その際，以下の制約を課す．

セントロイド行列 $\mathbf{C}$ の階数に関しての制約は，次元縮約 K 平均法と同様 (6.3) である．

メンバーシップ行列 $\mathbf{U}$ に関しては，パラメータがファジィ化されるため，

$$0 \leq u_{ik} \leq 1 \tag{6.12}$$

$$\sum_{k=1}^{K} u_{ik} = 1 \tag{6.13}$$

の制約が課される．以下，提案手法を次元縮約ファジィ K 平均法と呼ぶ．

6.3.2　アルゴリズム

次元縮約ファジィ K 平均法におけるクラスタリングパートは $\mathbf{U}$ に関して (6.10) の最小化を行う．求めるべき $\mathbf{U}$ はエントロピー正則化ファジィ c 平均法アルゴリズムにより

$$u_{ik} = \frac{\exp(-\lambda\|\mathbf{x}_i - \mathbf{c}_k\|^2)}{\sum_{j=1}^{K} \exp(-\lambda\|\mathbf{x}_i - \mathbf{c}_j\|^2)} \tag{6.14}$$

と得られる．

次元縮約パートでは次元縮約 K 平均法と同様，(6.9) によって $\hat{\mathbf{C}}$ を得る．

アルゴリズム全体は次のようになる．

0. 初期化: パラメータを初期化する．クラスター数 K，階数 R，および，正則化パラメータ λ を与える．
1. **U**-step: $\mathbf{C}$ を所与とし，(6.14) を用いて $\mathbf{U}$ を更新する．
2. **C**-step: $\mathbf{U}$ を所与とし，(6.9) を用いて $\hat{\mathbf{C}}$ を更新する．
3. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

6.4　数値実験

本節では 2 つの数値実験を行い，次元縮約ファジィ K 平均法の挙動を確認する．人工的に発生させたデータの真のクラスターと，次元縮約ファジィ K 平均法によって推定さ

表 16　ARI の値と局所解の頻度の四分位値

	ノイズ変数	Q1	Q2	Q3
ARI	25%	1.00	1.00	1.00
	50%	1.00	1.00	1.00
	75%	0.81	0.90	1.00
局所解	25%	0.00	0.00	0.00
	50%	0.00	0.00	1.00
	75%	0.00	4.00	21.25

れたクラスターの一致度を Adjusted Rand Index (ARI: Hubert & Arabie, 1985) によって評価する．ARI は 2 つのクラスタリング結果がまったく同じになれば上限 1 をとる指標である．なお，次元縮約ファジィ K 平均法は所属度が 0 以上 1 以下で与えられるため，クラスターへの所属度が最も高かったクラスターを所属クラスターとした．

6.4.1　数値実験 1

数値実験 1 ではクラスター構造を形成する変数と，クラスタ構造には関与しないノイズ変数が混在するデータを用いて，ノイズ変数がクラスタリング精度と局所解の頻度にどのように影響するかを検討する．

人工データ生成と解析方法

数値実験 1 で用いた人工データの作成方法を述べる．想定する真のクラスター数は $K = 3$ である．個体の数は各クラスター 10 個の計 30 個である．データの全次元は 16 次元とし，ノイズの次元の比率が全次元に対して 25%，50%，75% の 3 条件を設けた．つまり，25% 条件はクラスター構造に関係する 12 次元と，クラスター構造に関係ない 4 次元からデータが構成された．同様に 50% 条件はクラスター構造に関係する 8 次元，クラスター構造に関係ない 8 次元から，75% 条件はクラスター構造に関係する 4 次元，クラスター構造に関係ない 12 次元から構成された．

クラスター構造に関係するデータ行列に関して，平均は 3 クラスターがそれぞれ，全次元で 0，全次元で 2，全次元で- 2 のベクトルをとり，分散共分散行列は 3 クラスターすべ

てが単位行列の多変量正規分布に従う乱数が用いられた．クラスター構造に関係ないデータ行列に関して，平均は全次元で 0 のベクトル，分散共分散行列は単位行列の多変量正規分布に従う乱数が用いられた．各条件ごとに 100 個のデータセットを作成した．

この真のクラスター構造を持つデータを $\mathbf{X}_{t(L)}$ で表し，ノイズのデータを $\mathbf{X}_{e(J)}$ として表す（ただし，$L + J = 16$）と，数値実験 1 で用いた人工データは真のクラスター構造にかかわるデータ $\mathbf{X}_{t(L)}$ にノイズ $\mathbf{X}_{e(J)}$ を加えたもの

$$\mathbf{X} = [\mathbf{x}_{t(1)}, \ldots, \mathbf{x}_{t(L)}, \mathbf{x}_{e(1)}, \ldots, \mathbf{x}_{e(J)}]$$

と表される．

すべての条件においてクラスター数は $K = 3$，階数は $R = 2$ とした．正則化パラメータは 25%，50%，75% 条件のそれぞれで $\lambda = 0.07, 0.14, 0.19$ として次元縮約ファジィ K 平均法を適用した．

結果と考察

次元縮約ファジィ K 平均法を適用して得られた ARI と局所解の四分位値を表 16 に示す．25% 条件，50% 条件ではほぼ完全に真のクラスターを再現することができている．75% 条件でも高い精度で真のクラスターを特定することができている．クラスター構造に関係ないノイズの次元が増えると，局所解の発生率が増えることが確認された．局所解が多く発生した 75% 条件でもクラスタリング精度は高く保っており，提案手法は有用であると結論づけられる．

6.4.2　数値実験 2

数値実験 2 ではクラスター間分離度がクラスタリング精度と局所解の頻度にどのように影響するかを検討する．

人工データ生成と解析方法

想定するデータはクラスター構造を持つ部分 ($\mathbf{X}^{(t)}$) と持たない部分 ($\mathbf{X}^{(e)}$) から構成される．真のクラスター数は $K = 3$ とする．$\mathbf{X}^{(t)}$ は平均ベクトル μ_k，共分散行列 $v(\theta)\mathbf{I}_P$ の P 変量正規分布から生成した．平均ベクトル μ_k は [-10, 10] の実数一様分布から抽出し，各

表 17　ARI の値と局所解の頻度の四分位値

	クラスター間分離度 θ	Q1	Q2	Q3
	0.99	1.00	1.00	1.00
ARI	0.90	0.74	0.84	0.92
	0.80	0.43	0.54	0.65
	0.99	5.00	8.00	12.25
局所解	0.90	1.75	9.50	31.75
	0.80	29.50	47.50	49.00

変数の分散 $v(\theta)$ はクラスター間分離度 θ，群間平方和 $SSB = \sum_{k=1}^{K} n_k(\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}})^2$ とし，

$$v(\theta) = \frac{SSB \times (1-\theta)}{N\theta} \tag{6.15}$$

によって決定した．$\mathbf{X}^{(e)}$ は [-10, 10] の実数一様分布から発生させた．全変数の数を 10 とし，クラスターを構成する変数を 5，クラスター構造に関係ない変数の数を 5 とした．

クラスター間分離度に 3 つの条件 $\theta = 0.99, 0.9, 0.8$ を設け，各条件で以上の手続きを 100 回反復し計 300 個のデータを得た．各条件で発生させたデータを用いて，クラスター数を $K = 3$，階数を $R = 2$ と指定して提案手法を適用した．正則化パラメータ λ はそれぞれの条件で $\lambda = 0.99, 0.9, 0.8$ とした．初期値を 50 セット用意し，目的関数が最も小さくなったものを解として採用した．

結果と考察

次元縮約ファジィ K 平均法を適用して得られた ARI と局所解の四分位値を Table 17 に示す．クラスターが明確に分離しているときは完全に分類することができ，局所解の発生も少ないことが示された．クラスター間分離度が下がるにつれ，ARI の値が下がり，局所解も多く発生している．$\theta = 0.8$ 条件の局所解の第 3 四分位点は 49 となっており，収束した目的関数の値が全て異なる結果となった．クラスター間分離度が低いときは，局所解が頻発し，高い精度での分類はできないことが明らかになった．

6.5 データ解析

6.5.1 アイリスデータ

本節では，アイリスデータ (Fisher, 1936) を用いて，提案手法の詳細を検討する．このデータ行列は 150 個体，4 変数（がく片の長さと幅, 花弁の長さと幅）で構成される．各個体は，セトーサ，バーシクル，バージニカという 3 種類のアヤメから 50 個体ずつサンプルされたものである．つまり，150 個体は 50 個体ずつ 3 群に分類されることが既知である．

方法

アイリスデータの個体は 3 群に分類されることが既知であるので，本データ解析においても，クラスター数を $K = 3$ と設定する．正則化パラメータを $\lambda = 0.8$，次元数を $R = 2$ として提案手法を適用した．初期値を 50 セット用意し，目的関数が最も小さくなったものを解として採用した．

結果と考察

図 9 に 2 次元空間におけるデータの散布図を示す．各クラスターのセントロイドを「*」で，変数ベクトルを矢印を用いてプロットした．

まず，分類に関して検討する．各個体の色はメンバーシップ値を赤，緑，青の 3 色に変換したものである．したがって，色が鮮やかなほど特定のクラスターへ所属度が高く，中間色であればクラスターへの所属が曖昧であることを意味する．図の右側に位置する青のクラスターは，他の 2 クラスターとは明確に離れており，ほぼすべての個体が青のクラスターに対して非常に高いメンバーシップ値を取っている．赤と緑のクラスターはデータの分布が連続しており，明確な境界はない．赤と緑，それぞれのセントロイドの中間に位置するような個体は，赤と緑の中間色で塗られており，所属の曖昧さが表現されている．赤と緑，それぞれのクラスターに属する個体でも，セントロイドから離れ，周辺に位置する個体は鮮やかな色でプロットされており，特定のクラスターへのメンバーシップ値が高いことを示している．

次にクラスターの特徴について検討する．変数ベクトルを見ると，横軸は，特に花弁の

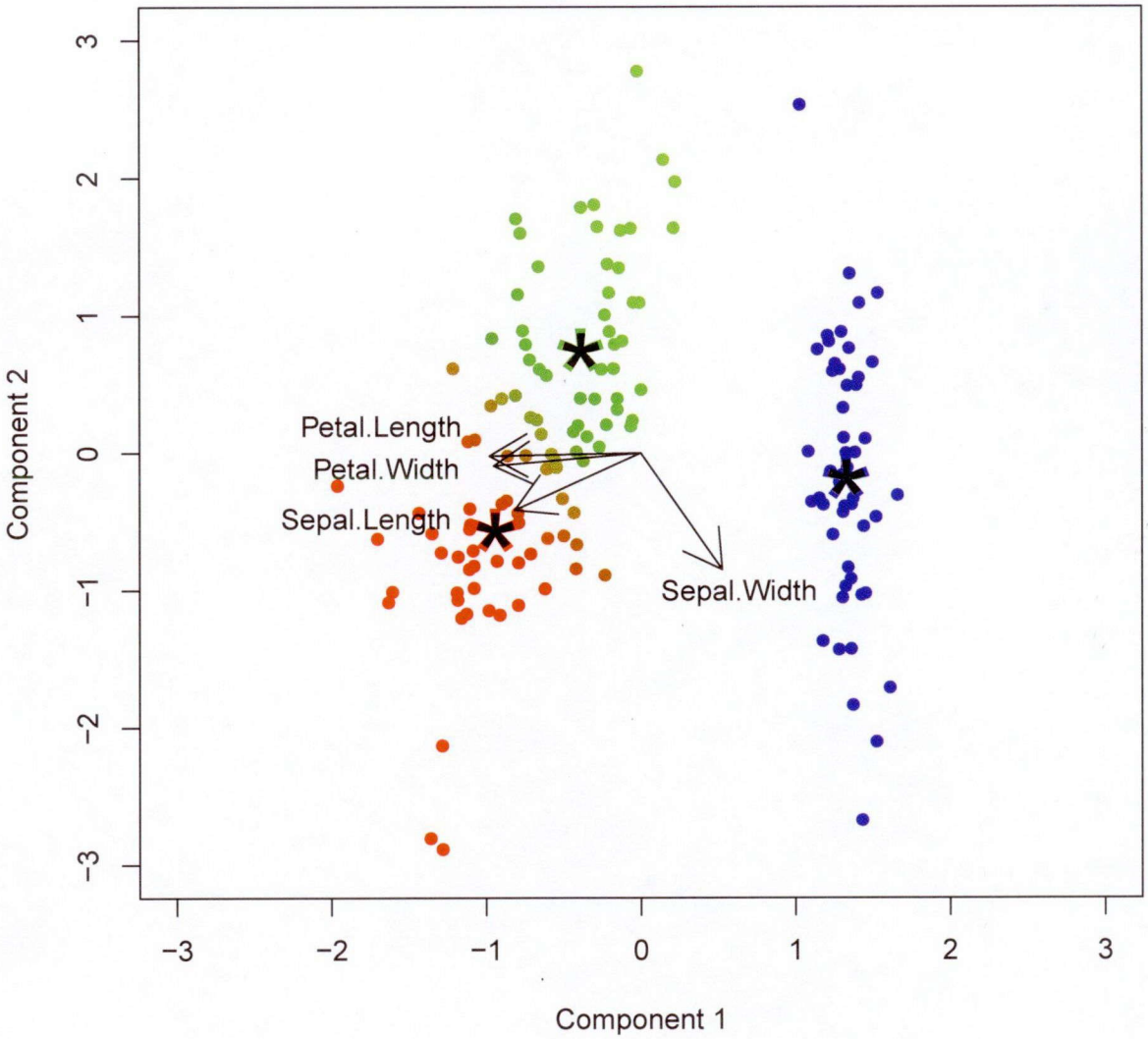

図 9　2 次元空間における散布図.「*」はクラスターセントロイドを示し，矢印は変数ベクトルを表す.

長さ（Petal Length）と花弁の幅（Petal Width）との関係が大きいが，全変数と大きな関係がある．それに対して縦軸は花弁（Petal）との関係はほとんど無く，がく片の長さ（Sepal Length）や，特に，がく片の幅（Sepal Width）関係が大きい．以上より，横軸は花の全体的な大きさを，縦軸はがく片の大きさを示していると解釈できる.

表 18　メンバーシップパラメータ行列．3 つのクラスターのうち最もメンバーシップ値が高いセルに網掛けを行った．

	クラスター 1	クラスター 2	クラスター 3
ネズミ (子)	0.827	0.166	0.007
ウシ (丑)	0.940	0.054	0.007
トラ (寅)	0.890	0.100	0.010
ウサギ (卯)	0.759	0.235	0.006
ウマ (午)	0.973	0.026	0.001
ヒツジ (未)	0.935	0.063	0.002
サル (申)	0.691	0.283	0.027
イヌ (戌)	0.835	0.153	0.012
イノシシ (亥)	0.936	0.061	0.003
ヒト (人)	0.415	0.329	0.257
カナリヤ (カ)	0.010	0.988	0.002
ワシ (鷲)	0.014	0.983	0.003
ペンギン (ペ)	0.105	0.781	0.114
ダチョウ (ダ)	0.371	0.623	0.006
ニワトリ (酉)	0.067	0.929	0.004
リュウ (辰)	0.027	0.017	0.956
ワニ (鰐)	0.017	0.127	0.856
ヘビ (巳)	0.016	0.058	0.926
クジラ (鯨)	0.018	0.031	0.950
マグロ (鮪)	0.000	0.001	0.999
イワシ (鰯)	0.000	0.002	0.998
タコ (蛸)	0.041	0.021	0.937

6.5.2　動物評定データ

本節では曖昧さが積極的に解釈に有用であることを例示する．22 種類の動物（ネズミ，ウシ，トラ，ウサギ，ウマ，ヒツジ，サル，イヌ，イノシシ，ヒト，カナリヤ，ワシ，ペ

表 19　2 次元空間における変数ベクトル

	次元 1	次元 2
大きさ	-0.190	-0.302
走る	0.613	-0.458
飛ぶ	0.113	0.648
泳ぐ	-0.679	-0.068
体毛	0.884	-0.017
脚	0.372	-0.383
蹄	0.395	-0.346
羽	0.327	0.878
鱗	-0.736	-0.082
肺呼吸	0.553	0.091

ンギン，ダチョウ，ニワトリ，ワニ，リュウ，ヘビ，クジラ，マグロ，イワシ，タコ）に関して，10 個の特徴（大きさ，走る，飛ぶ，泳ぐ，毛，足，蹄，羽，鱗，肺呼吸）について評定されたデータを用いる．10 個の特徴のうち，はじめの 5 個は，対象の動物がその特徴を「とても良く有する」から「まったく有しない」までの 5 段階で，残りの 5 個については，「足」は足の本数，「蹄」は蹄の数，「羽」，「鱗」，「肺呼吸」は特徴を有するか否かの 2 値によって評定されている．

方法

解の解釈可能性と，可視化できることを考慮して，クラスター数 $K = 3$，階数 $R = 2$，正則化パラメータ $\lambda = 0.3$ として提案手法を適用した．データ行列には変数ごとに標準化したものを使用した．初期値を 50 セット用意して最も目的関数が小さくなったものを解として採用した．

結果と考察

まず，どのように個体が分類されたかを検討する．メンバーシップ行列を表 18 に示す．3 つのクラスターのうち最もメンバーシップ値が高いセルに網掛けを行った．3 つのクラ

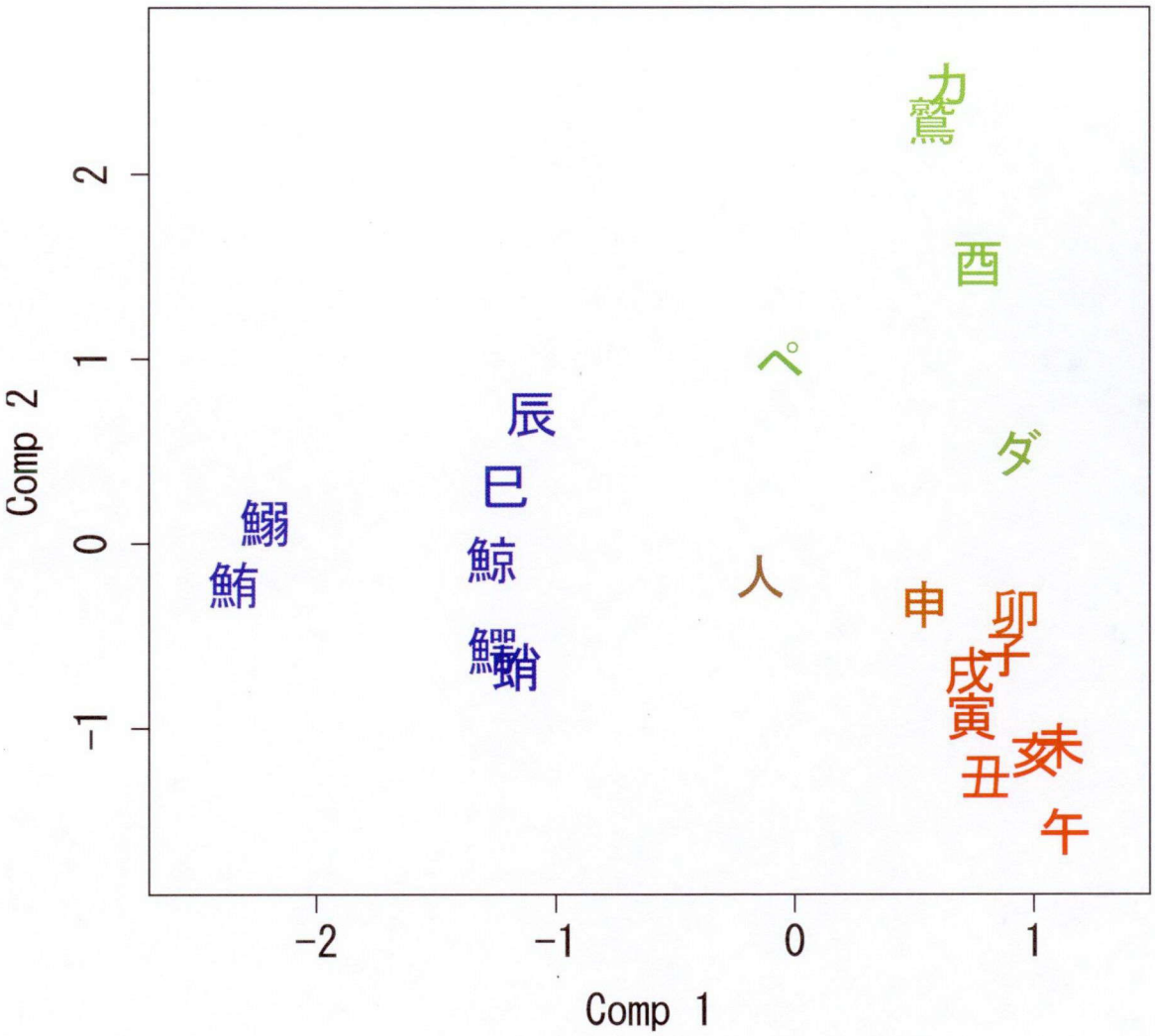

図 10　2 次元空間における 22 種の動物の布置

スターは，およそ哺乳類，鳥類，魚類・爬虫類であることがわかる．「ウマ」や「イノシシ」，「カナリヤ」や「ワシ」，「マグロ」や「イワシ」などは特定のクラスターにのみ高いメンバーシップ値が割り当てられており，これらの動物は各クラスターの特徴をよく反映している典型的個体といえる．それに対して「ペンギン」，「ダチョウ」，「ヒト」，「サル」などの動物には 2 つ以上のクラスターにまたがってメンバーシップ値が多く割り当てられている．これらの動物は各クラスターの特徴を多く共有するため，クラスター間の境界に位置し所属が曖昧な個体といえる．

次に，クラスターの特徴について検討する．2 次元空間における個体の布置を図 10 に示す（表 18 の動物名の後のカッコ内の一字によってプロットした）．各個体の色はメンバーシップの値を赤，緑，青の 3 色に変換したものである．したがって，ファジィな個体ほど中間色でプロットされる．図 10 の横軸，縦軸は 10 個の変数から新たに合成された 2 つの次元である．表 19 に 2 次元空間における変数ベクトルを示す．横軸と関係が大きい変数は「泳ぐ」，「鱗」，「肺呼吸」，「体毛」，縦軸と関係が大きい変数は「羽」，「飛ぶ」，「大きさ」である．「走る」，「脚」，「蹄」は両方の軸に同程度影響していることがわかる．以上から，横軸は水棲か否かを，縦軸は空を飛ぶか否かを示す軸と解釈できる．

以上より，提案手法は，個体を強制的にクラスターへ分類するのではなく，そのままファジィなメンバーシップ値を与えることにより，所属が曖昧な個体を発見できることが示された．また，次元縮約により個体の分布を可視化できるため，解釈を行いやすいことが示された．

6.6　クラスターサイズを考慮した次元縮約ファジィ K 平均法の改良

第 4 章 5 節での問題意識と同様に，クラスターに含まれるサンプルの数が大きく異なるように分類されるべき場合でも，サンプルの数が等しくなるように分類されてしまうことを避けるため，次元縮約ファジィ K 平均法に Miyamoto et al. (2008) のクラスターサイズ調整パラメータを導入して解決を試みる．すなわち，クラスターサイズ調整機能を有する次元縮約ファジィ K 平均法の目的関数とそのアルゴリズムを提案する．

6.6.1　目的関数

提案手法の目的関数は

$$F(\mathbf{C},\mathbf{U},\alpha) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|\mathbf{x}_i - \mathbf{c}_k\|^2 + \lambda^{-1}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\log\frac{u_{ik}}{\alpha_k} \tag{6.16}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 + \sum_{k=1}^{K} N_k\|\bar{\mathbf{x}}_k - \mathbf{c}_k\|^2 + \lambda^{-1}\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\log\frac{u_{ik}}{\alpha_k} \tag{6.17}$$

とあらわせる．この関数を最小にするパラメータ $\mathbf{C}$，$\mathbf{U}$，α を

$$\mathrm{rank}(\mathbf{C}) = R, \tag{6.18}$$

$$\sum_{k=1}^{K} u_{ik} = 1 \ \text{ and } \ 0 \le u_{ik} \le 1, \tag{6.19}$$

$$\sum_{k=1}^{K} \alpha_k = 1 \ \text{ and } \ 0 \le \alpha_k \le 1. \tag{6.20}$$

の制約のもとで求める．

6.6.2　アルゴリズム

所属度 u_{ik} は (6.16) を最適化することにより

$$u_{ik} = \frac{\alpha_k \exp(-\lambda \|\mathbf{x}_i - \mathbf{c}_k\|^2)}{\sum_{j=1}^{K} \alpha_j \exp(-\lambda \|\mathbf{x}_i - \mathbf{c}_j\|^2)} \tag{6.21}$$

と得られる．

階数 R のセントロイド行列 $\mathbf{C}$ は (6.17) を最適化することにより，次元縮約ファジィ K 平均法と同様に

$$\hat{\mathbf{C}} = \mathbf{W}^{-\frac{1}{2}} \mathbf{U}_R \mathbf{S}_R \mathbf{V}_R' \tag{6.22}$$

を得る．

クラスターサイズ調整パラメータ α は

$$\alpha_k = \frac{1}{N} \sum_{i=1}^{N} u_{ik} \tag{6.23}$$

により得られる．

アルゴリズム全体は以下のようになる．

0. 初期化: パラメータを初期化する．パラメータを初期化する．クラスター数 K，階数 R，および，正則化パラメータ λ を与える．
1. **U**-step: **C**，$\boldsymbol{\alpha}$ を所与とし，(6.21) を用いて **U** を更新する．
2. **C**-step: **U**，$\boldsymbol{\alpha}$ を所与とし，(6.22) を用いて **C** を更新する．
3. $\boldsymbol{\alpha}$-step: **U**，**C** を所与とし，(6.23) を用いて $\boldsymbol{\alpha}$ を更新する．
4. 収束判定: 収束基準を満たしていれば終了．そうでなければ 1. へ戻る．

6.6.3　データ解析

前節と同じ動物データに対して，クラスターサイズを考慮した次元縮約ファジィ K 平均法を適用する．パラメータの設定も前節と同様，カテゴリ数を K=3，階数を R = 2 正則化パラメータを λ = 0.3 とした．

図 11 に 2 次元空間における個体の布置を示す．各個体の色はメンバーシップパラメータ **U** の値を赤，緑，青の 3 色に変換したものである．クラスターサイズ調整機能を有する次元縮約ファジィ K 平均法では図 10 に示した次元縮約ファジィ K 平均法に比べ，全体的な個体布置は変わっていないが，緑のクラスターサイズが小さくなっていることがわかる．

緑のクラスターは鳥類に相当すると解釈されるが，鳥類の個体は他の 2 クラスターよりも個体数が少なく，ニワトリ，ペンギン，ダチョウのように，鳥類の典型ではない個体も多く含まれている．このようなデータの場合，ファジィな手法を用いても，適切なメンバーシップ値を得られるとは限らない．クラスターサイズ調整機能を有する次元縮約ファジィ K 平均法を用いると，クラスターサイズを柔軟に変化させ，ニワトリやペンギン，ヒトといった動物のファジイさをうまく表現できている．

6.7　まとめ

本章では，個体のクラスタリングと変数の次元縮約を同時に達成する次元縮約 K 平均法をエントロピー正則化によりファジィ化し，クラスターへの所属が明確でない個体を表現できる次元縮約ファジィ K 平均法の目的関数を提案し，パラメータの推定アルゴリズムを開発した．

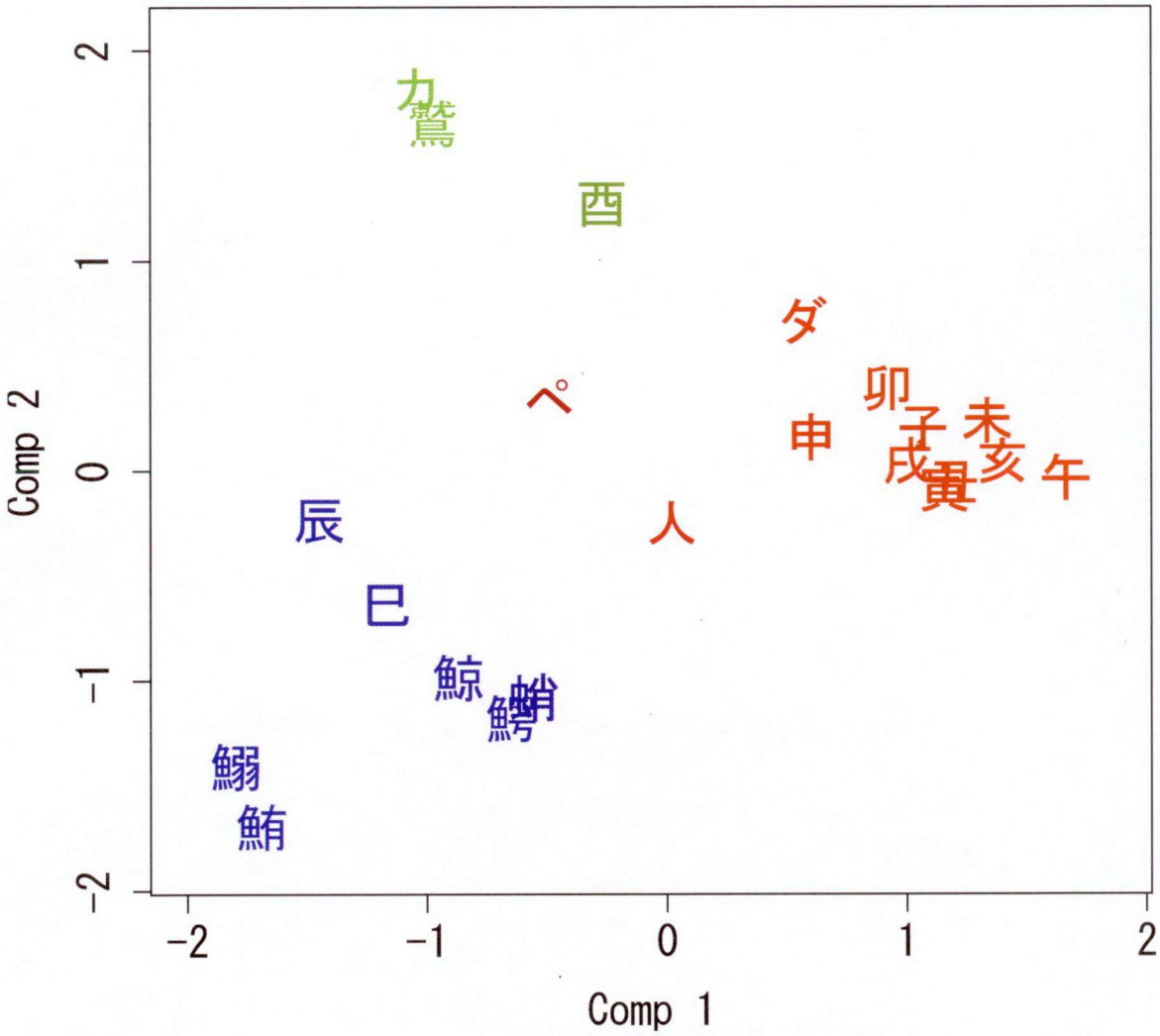

図 11　2 次元空間における個体の布置

数値実験により，クラスター構造に関係しないノイズ変数が加わっても，高い精度でクラスタリングすることができるが，クラスター間分離度が下がるとクラスタリング結果に大きく影響することが示された．データ解析結果から，クラスターが明確に分かれていないとき，ファジィなメンバーシップ値が解釈に有用であることが示された．

さらに，クラスターサイズを調整する機能を有した次元縮約ファジィ K 平均法の目的関数を提案しパラメータの推定アルゴリズムを示した．この手法により，K 平均法の問題点であり，その拡張手法である次元縮約ファジィ K 平均法においても問題となるクラス

ターサイズが均等に分類される傾向を解消することができる．データ解析によりその有用性が示された．

第 7 章

総合考察

本論文内で提案された手法については，それぞれの章で考察を行ったが，本章では章を超えて得られる知見について考察を行う．各提案手法は，全てパラメータのファジィ化をともなったクラスタリング法である．まず，提案手法を総括し，データ解析手法としての位置づけを考察する．次に，提案手法をデータ解析手法としてではなく，認知モデルとしての解釈可能性を探る．最後に計量心理学における分類について展望を述べる．

7.1　総括と提案手法の位置づけ

本節では第 4 章から第 6 章で提案されたクラスタリング手法の総括を行い，クラスタリング手法もしくは多変量データ解析手法としての位置づけについて考察する．

多変量データ解析手法は様々な観点から分類できる．例えば，外的基準が有るか無いか，変数の尺度の水準は何か，変数の個数はいくつか，潜在変数を用いるか否か，などである (柳井, 1994)．ほかにも，志向性は探索的か確認的か，主要目的は空間表現か分類かそれとも因果分析か，といった分類も考えられる (足立, 2006)．足立 (2006) によれば，探索とは「何か結論を出すというよりも，データが有する傾向の概略を発見して，今後の研究に役立てる」ことを意味する．

クラスタリング手法は，外的変数を用いない教師なし学習による，対象のグループ分けであり，探索的に用いられる．パラメータがファジィ化されていることは，探索的に用いられるクラスタリングにおいて非常に有用であると考えられる．これまでに提案されてきた方法と，本論文で提案された方法を，ファジィ化の方法と対象の観点からの整理した

表 20　ファジィ化の方法と対象による K 平均法の分類

方法＼対象	個体パラメータ	変数パラメータ
べき乗型	べき乗型ファジィ c 平均法	W-K 平均法
エントロピー正則化	第 6 章，正則化ファジィ c 平均法	第 4 章
リパラメトライズ	第 5 章	

(表 20).

従来，K 平均法においてファジィ化の対象として考えられてきたパラメータは個体に対して与えられるメンバーシップパラメータであった．メンバーシップパラメータをファジィ化する方法として Bezdek (1980) によるべき乗型アプローチ，Miyamoto & Mukaidono (1997) によるエントロピー正則化アプローチが提案されている．第 6 章ではエントロピー正則化法を用いて，次元縮約 K 平均法におけるメンバーシップパラメータのファジィ化を行った．さらに第 3 のアプローチとしてリパラメトライズによるファジィ化を第 5 章において提案した．

近年では，解釈可能性の向上を目指し，変数パラメータに対してファジィ化の方法が採られている (Huang et al., 2005; Nishida, submitted). Huang et al. (2005) はべき乗型アプローチを用いて，変数パラメータにファジィネスを導入し，W-K 平均法を提案した．本論文の第 4 章ではエントロピー正則化による変数パラメータのファジィ化を行った．

7.1.1　変数に対するパラメータのファジィ化

研究の初期段階においてデータの変数の数が多い場合，解釈が困難になり，変数を減らしたい状況が生まれてくる．このようなニーズに対し，第 4 章で提案した変数選択をともなったクラスタリングが応じられる．

第 4 章では，K 平均法においては潜在的に固定化されている変数に対する係数パラメータを顕在化した定式化を行い，ファジィ理論を援用することにより，変数に対する係数パラメータを連続値として推定可能なアルゴリズムを開発した．この方法では係数パラメータを解釈することにより，変数選択を可能にし，クラスターの特徴を容易に理解できる．

従来のクラスタリング法においてファジィ化されるパラメータはメンバーシップであったが，本提案手法においては，変数に対する係数パラメータをファジィ化している．この

係数パラメータは従来の K 平均法では目的関数に表れず，本提案手法が導入したものである．係数パラメータは 0 以上 1 以下の連続値として推定されるため，変数に対する重要度として解釈することができる．係数の値が 0 に近ければ，それに対応する変数はクラスタリングにおいて重要な役割を果たしていないことになり，係数の値を見ることにより変数選択が可能となる．逆に係数の値が大きい変数は，クラスターを特徴付ける重要な変数であると解釈できる．

このような解釈は従来の K 平均法ではかなわず，係数パラメータを導入した本提案手法によって可能となる．本提案手法の観点から見れば，係数パラメータが目的関数に表れない K 平均法は，潜在的に係数パラメータが変数数の逆数に固定されていると解釈できる．すなわち，係数パラメータを推定する本提案手法は，K 平均法を一般化した解析手法と位置づけることができる．

7.1.2　個体に対するパラメータのファジィ化

探索的にクラスタリングを用いる場合，個体の所属を明確に決めてしまうのは，適切とはいえない場合がある．クラスターが明確に分離している場合は，クリスプクラスタリングを用いてメンバーシップパラメータを推定しても問題無いが，クラスター境界が明確でない場合は，クラスターへの所属をはっきりと決めてしまうのではなく，所属には曖昧さがあることを理解しながら解釈を行う方が妥当である．第 5 章と第 6 章において，メンバーシップパラメータをファジィ化した，クラスタリング法を提案した．

第 5 章では，行列表現による K 平均法の定式化を用いることによって，従来行われてきたファジィ K 平均法とは異なるファジィクラスタリング法の目的関数を提案し，パラメータを推定する交互最適化アルゴリズムを開発した．

K 平均法ではメンバーシップパラメータをファジィ化するに当たって，ファジィパラメータの導入や正則化項を用いるといった目的関数を変形する方法がとられてきた．もちろんこれらの方法でも，メンバーシップパラメータをファジィ化するという目的を達成できるが，ファジィパラメータや正則化パラメータを解析の事前に決定しなくてはならない．しかしながら，これらのチューニングパラメータの決定法は，整備されていない．

本提案手法は行列表現による K 平均法の定式化を用いることによって，上記のようなチューニングパラメータなしにメンバーシップパラメータのファジィ化を達成した．したがって，分析者が決定しなくてはならない事前のパラメータは通常の K 平均法と同様ク

ラスター数のみである．

また，本提案手法の行列表記による目的関数は，次元縮約法として用いられる主成分分析と同一である．ただし，主成分分析には階数の制約が課されるが，本提案手法では階数に関する制約はなく，ファジィクラスタリングとしての制約が課される．このような特徴から本提案手法は，メンバーシップパラメータを解釈することにより，クラスタリング法として用いることはもちろんであるが，係数行列を解釈することにより，主成分分析のように解釈を行うことも可能である．すなわち，ファジィクラスタリングの特徴と，主成分分析の特徴を併せ持った中間的手法といえる．

第 6 章では，次元縮約とクラスタリングを単一の目的関数の最小化によって達成する次元縮約 K 平均法を拡張し，ファジィなメンバーシップパラメータを得られるような目的関数を提案し，そのパラメータを推定する交互最適化アルゴリズムを開発した．クラスターが明確に分離していない場合，ファジィなメンバーシップを用いることで，個体がクラスター所属する程度が連続値になり，より適切に解釈可能となった．

次元縮約 K 平均法は変数が多いデータセットに対してクラスタリングを行いたい場合に有用な方法である．K 平均法を基礎としているため，そのメンバーシップ値は 2 値として得られる．本提案手法は，次元縮約 K 平均法を正則化法により，ファジィなメンバーシップパラメータを求められるよう拡張したものである．対象のデータに内在するクラスター構造が明確に分離している場合は，ファジィクラスタリングを用いても結果の解釈においてはクリスプクラスタリングとさほど変わりない．しかしながら，我々が手にするデータにはクラスター構造が明確に分離されていないものも多く存在する．このようなデータに対しては，ファジィクラスタリングを用いて初めて，その曖昧なクラスター構造を把握できる．

特に，次元縮約 K 平均法のように，低次元空間における表現を目的とした場合，情報が圧縮されるため，クラスター境界が曖昧になる可能性が生じる．本提案手法は，個体がクラスター所属する程度が連続値になり，より適切に解釈可能となった．

7.2　認知モデルとしての教師なし分類学習手法

本論文の目的は，データ解析手法としての教師なし分類学習の手法，すなわちクラスタリング法を開発することであったが，本節では教師なし分類学習の手法をデータ解析手法

ではなく，計算論的認知モデルとして扱うことの可能性を考察する．

7.2.1　認知機能へのモデルベースアプローチ

認知機能を解明しようとする研究のアプローチには，実験・調査を用いた実証的アプローチと計算機上で構築されたモデルを用いたモデルベースアプローチに大別される．後者のモデルベースアプローチはさらに 2 つに分けられる．一つは記号論的なアプローチであり，もう一つは統計学的なアプローチである．

記号論的アプローチは，古典的な認知科学や人工知能研究において用いられていたもので，認知プロセスを記号操作として記述する (Newell & Simon, 1976)．代表的な記号論的アプローチは ACT-R(Anderson, 2007) である．ただし，ACT-R は特定の認知機能，プロセスに対するモデルではなく，IF-THEN ルールの組み合わせによって汎用的に認知機能を記述しうる統合認知アーキテクチャ (Newell, 1994) であり，細分化された認知機能の研究の統合を目指すものである．

それに対して，統計学的なアプローチは，数理モデルによって記述される．人工知能研究において統計学的アプローチを取り，人間が行っているような学習を計算機によって実現しようとする分野は機械学習や統計的学習 (Bishop, 2006; Hastie, Tibshirani, & Friedman, 2001) と呼ばれる．

人間の認知機能を数理モデル化するという点で，認知科学，認知心理学でなされてきた数理モデル研究は機械学習に近いものがある．ただし，機械学習で良いとされるモデル・アルゴリズムには誤差が小さいことや正答率が高いことが要求されるが，認知科学，認知心理学で良いとされるモデルは行動・心理実験で得られた結果を記述・再現できるかということが要求される点においては違いがある．

7.2.2　データ解析と認知機能の目的

数理モデルを扱う心理学関連分野に計量心理学と数理心理学がある．計量心理学は統計学をバックグラウンドとし，心理データの解析手法の開発を目的とする．数理心理学は心理学で得られた知見から数理モデルを構築し，心理現象の解明を目的とする（図 12）．目的は異なると言いながらも，両者を厳密に線引きすることは難しい．たとえば，因子分析は，現在では多変量データ解析手法の 1 つとして広く用いられているが，もともとは知

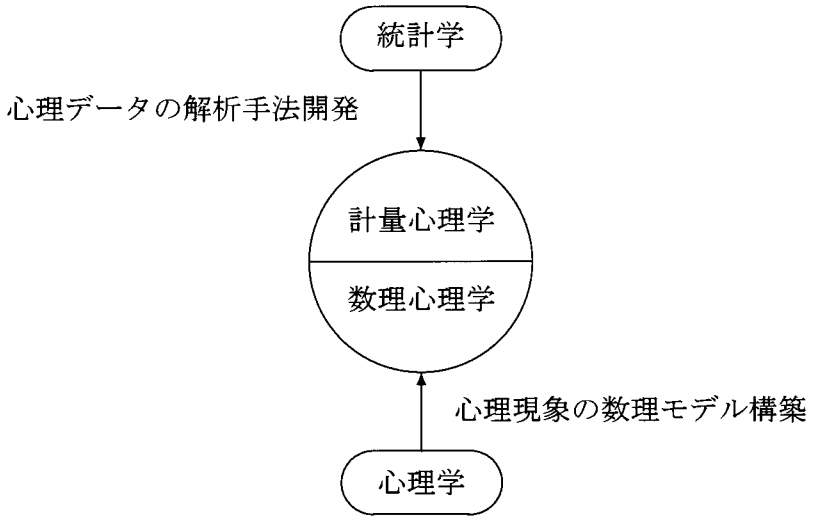

図 12　数理モデルを扱う心理学関連分野の関係

能のモデルとして出発した (Spearman, 1904). このような例を考えると，データ解析手法におけるモデルは，特定の認知機能に対するモデルではないが，それを用いて認知機能を解釈することは可能であると考えられる．その理由として，データ解析と認知機能には目的が同じであるもの多くあることが挙げられる．認知機能の役割には，入力されるデータ（刺激）に何らかの処理を行い，結果の予測，パターンの発見，情報の圧縮などの目的を達成することが挙げられる．このような目的はデータ解析においても同様に挙げられるもので，目的を達成するためのデータ解析手法が存在する．すなわち，判別分析，クラスタリング，主成分分析などである．

判別分析はクラスラベルとともなった入力データからクラス分類法を学習し，未知の入力があったときに学習済みの分類法を適用し，どのクラスに分類されるかを予測する．これは心理学における教師あり概念学習といえ，概念識別の数理モデルを構成する際，判別分析の枠組みを利用できる (Ashby, 1992). 概念学習モデルの ALCOVE (Kruschke, 1992) や SUSTAIN(Love, Medin, & Gureckis, 2004) も変数への注意量をパラメータとして導入した放射基底関数ネットワークによる判別分析と考えることができる．また，概念学習におけるプロトタイプ，イグザンプラーといった仮説は，混合分布モデルの両極限として表現できる (Rosseel, 2002). より記述的なモデルとして，メタヒューリスティックアルゴリズムを導入したり，誤差の最小化ではなく知識の複雑性を考慮した効用関数を最適化する概念学習モデルもある (Matsuka, Sakamoto, & Chouchourelou, 2008).

クラスタリングの目的はクラスラベルがない入力データを少数の群に分類し，データのパターンを発見することである．これは，認知機能で言えば教師なし概念生成といえる．教師なし分類学習の方法は機械学習などの分野で盛んに研究されている．有名な概念クラスタリングの手法として COBWEB(Fisher, 1987) がある．COBWEB は category utility(Corter & Gluck, 1992; Gluck & Corter, 1985) と呼ばれる基準を最大化するクラス分類木を生成するクラスタリングアルゴリズムである．心理学における教師なし概念学習の確率モデルには，混合正規分布モデルを用いたモデル (Fried & Holyoak, 1984) や，ディリクレ過程混合モデルを用いたモデル (Anderson, 1991) がある．また，Pothos & Chater (2002) は単純原理による教師なし概念形成モデルを構築している．

主成分分析の目的は高次元のデータをなるべく情報を保ったまま低次元の成分で近似することと言える．データの次元を縮約，復元するという枠組みよって人間のパターン認識機能も体系づけられると考えられる (乾, 1995)．例えば，人間が持つランダムドット，幾何学図形，文字，顔などのパターンの類似性表現が主成分分析の結果と非常に似ていることが示されている (乾, 1995; Makioka, Inui, & Yamashita, 1996; 森崎・乾, 1995)．すなわち，視覚情報処理過程が主成分分析と似た性質を持っていることを示唆する．

7.2.3　教師なし概念生成の認知モデルとしてのクラスタリング

我々の周りにあふれている物や現象を認識し，把握するためには少数の群に分けて整理することが必要である．人間は情報を整理し，必要なときに素早く利用可能にするため，分類によって作られた集合にラベルをつけている．この集合は「概念」と呼ばれる．そして，概念に含まれる事例や，概念が持つ特性というものは「知識」と言い換えることができる．松香・本田・吉川 (2010) は「概念の性質を明らかにすることは，我々が持つ知識構造という直接的な論題のみならず，人間の認知システムの性質を明らかにしていく上でも非常に重要である」と述べている．例えば，我々は概念を用いて帰納的な推論を行っているし (Osherson, Smith, Wilkie, López, & Shafir, 1990)，確率判断においても概念や知識が重要な役割を果たす (Hattori & Nishida, 2009; 西田・服部, 2011)．このように，概念は思考の基盤であると考えられる．

概念は古くから心理学の研究対象であり，その話題は多岐にわたるが，最も重要なものは概念はどのように形成されるかという概念学習に関する問題である．概念学習も教師ラベルの有無によって教師あり概念識別と教師なし概念形成の 2 つに分けられる．人間が

概念を獲得する状況は，教師ラベルがない事例からの学習であることも多いであろう．概念を形成することは事例間の類似性（もしくは距離）から最適な分割を発見することすることに他ならない．つまりクラスタリングと言えよう．心理学的な術語を用いればクラスターは概念，個体は事例，変数は特徴と言い換えることができる．

人間が行っている教師なし分類学習をクラスタリングとするならば，その特徴はどのようなものだろうか．多変量解析法としてのクラスタリングとは何が同じで何が異なるのだろうか．おそらく人間が行うクラスタリングには，少なくとも，概念の曖昧さと特徴次元の選択・縮約という 2 つの特徴があると考えられる (西田, 2010b).

概念の曖昧さ

ある事例を見たときに概念 A と B の 2 つのどちらかに分類するかという問題の場合，A のほうにも属するような気もするし，B のほうに属するような気もするという場面は誰しも経験したことがあるだろう．このように概念構造は曖昧さを有すると考えられる．

古典的な概念に関する研究では特徴とその値（e.g. 足が 4 本ある，鱗を持っているなど）で概念は定義されるとしていた (Bruner, Goodnow, & Austin, 1956). しかし，ある概念を特徴づけられる，必要十分な定義的特徴は多くの場合存在しない．概念に含まれる平均的な事例から離れた事例は，同一人物が評定しても，文脈によって概念に含まれるか否かの判断が異なってくる (Mccloskey & Glucksberg, 1978). したがって，多くの事例に共有される複数の特徴が概念を形成させていると考えられる (Rosh & Mervis, 1976). この考えは「家族的類似性」とよばれる (Wittgenstein, 1958). すなわち家族的類似性を持つ事例を，概念としてまとめるためにはファジィな分類が必要となる．

特徴次元の選択・縮約

例えば，ある生物を見たときにそれをどの概念に分類するかという課題に対して，人間はどのように振舞っているだろうか．K 平均法などのクラスタリング手法においては，分類対象となる事例が持つ特徴の次元数はあらかじめ与えられており，与えられた全ての特徴を用いて分類を行う．

しかし，人間が概念形成を行う場合には事例のどの特徴を使えばよいかということは事前にはわからない．また，7 ± 2 チャンク程度 (Miller, 1956) といわれている短期記憶の容量に代表される，認知的な限界を考慮すると多くの特徴を用いているとは考えにくい．

人間は高次元の情報を縮約し利用できるとされており (Markman, 1999)，概念形成に効果的な特徴を自ら抽出もしくは次元縮約していると考えられる．

以上のような観点からすると，変数選択や次元縮約による情報の圧縮とファジィクラスタリングを同時に達成するモデルが，人間の概念形成の認知モデルとして妥当であると考えられる．

7.2.4　提案手法の概念形成モデルとしての適用可能性

第 4 章で提案した変数選択をともなう K 平均法は，変数がクラスタリングに重要かそうでないかを識別できるという点で，上記の特徴次元の選択を達成できる．新たに導入された変数に対する係数パラメータは，概念識別モデルである ALCOVE (Kruschke, 1992) などで採用されている，選択的注意パラメータと同等であると解釈できる．また，メンバーシップパラメータのファジィ化も達成できる (西田, 2011).

第 6 章で提案した次元縮約をともなうファジィ K 平均法は，変数の次元縮約とファジィ分類を達成している．したがって，これらの方法は人間の教師なし概念形成モデルとしても扱うことができると考えられる．

人間がどのように情報圧縮しているのか，すなわち次元縮約を行っているのか，変数選択を行っているのかという方略に関しては実験研究による結果が待たれる．実験的に得られたデータと，これらのモデルによる予測値を照合することで，モデルを修正し，さらに精緻化することができると考えられる．

しかしながら，認知モデルとして不十分であるところもある．一般に精度とコストはトレードオフの関係にある．分類の性能を上げるためには学習時間を増やさなければならない．しかしながら，学習時間を増やしても，ある程度以上の精度を得ることは難しい．人間は学習に無限の時間を費やせるわけではない．したがって，なるべく短い時間で，ほどほどの精度が得られるような学習方略をとっていると考えられる．機械学習や，統計的学習において，学習は誤差の最小化と考えられていることが多いが，認知モデルにおいては学習時間などの学習にかけるコストも，同時に最小化する対象であると考えられる．

7.3　計量心理学における教師なし分類学習の展望

本論文で提案した手法は，全て K 平均法を基礎としている．そのため，K 平均法において問題とされる点をそのまま引き継いでいる．しかし，その問題も近年では解決法が提案されつつある．例えば，K 平均法で問題とされてきた，クラスタサイズが均等に分類される傾向は，クラスタサイズパラメータの導入によって解決が試みられている．本論文においても，第 4 章と第 6 章において提案手法にクラスタサイズパラメータを導入した改良手法を提案している．以下では，本論文では達成できなかった問題や課題を述べる．

7.3.1　分類境界の非線形化

K 平均法は有用なクラスタリング法であるが，線形分離可能なクラスターしか得ることができない．そのため分類境界が非線形の場合でもクラスタリング可能な方法が開発されている．非線形分類境界を得るための有力なアプローチとしてカーネル法がある (赤穂, 2008; 福水, 2010; Vapnik, 2000).

元のデータに対して非線形変換を行うと，非線形特徴や高次モーメントが表現されデータの性質をとらえやすくなる場合がある (福水, 2010). したがって，データをもとの空間から高次元の特徴空間へ写像することを考える．多くの線形のデータ解析は内積計算に拠っているため，特徴空間における内積を計算することが求められるが，写像をうまく選択することにより特徴空間における内積がカーネル関数によって与えられる．よって，高次元特徴空間での計算が効率的に行える．これを利用することにより様々な線形のデータ解析手法を変換後のデータに適用できる．たとえば，(Girolami, 2002) はカーネル法を用いたデータを高次元特徴空間へ写像して分類を実行し，非線形な分類境界を得るカーネル K 平均法を提案している．また，メンバーシップをファジィ化したカーネルファジィ c 平均法も提案されている (Miyamoto et al., 2008).

本論文で提案した手法も K 平均法を基礎としており，線形分離可能なクラスターしか得ることができない．カーネル法を用いることによって，非線形な分類境界を得ることができる拡張を行うことができると考えられる．

7.3.2　制約付きクラスタリング

近年，盛んに研究が行われているクラスタリング手法に，制約付きクラスタリング (Basu, Davidson, & Wagstaff, 2009) や半教師ありクラスタリング (Chapelle, Schölkopf, & Zien, 2006) といったものがある．これらの方法は，データに含まれる個体のうちいくつかについて所属クラスターに関する事前知識が与えられているときに適用可能な手法である．

その先駆けは，K 平均法において，ある個体とある個体は必ず同じクラスターに所属するという must-link 制約や，必ず異なるクラスターに所属するという cannot-link 制約を導入した制約付き K 平均法である (Wagstaff, Cardie, Rogers, & Schroedl, 2001)．事前情報として，一部の個体がどのクラスターに所属するかが分かっているときには，クラスラベルが既知である個体のラベル情報が，クラスラベルが未知である個体へ伝播するような方法が取られることもある Zhou, Bousquet, Lal, Weston, & Schölkopf (2004)．個体についての事前情報を用いるのではなく，クラスター構造に関する事前知識を導入することも考えられる．足立 (2011) は，クラスターに所属する個体の数を指定してクラスタリングを行うサイズ固定クラスタリングを提案している．

我々が手にするデータは，全く外部情報が無いものや，完全に外部情報が付随しているデータばかりではない．部分的に外部情報を得られている場合には，その情報を積極的に用いることでよりよい解析が可能になると考えられる．

7.3.3　クラスタリングをともなった同時解析

解析法 A で得られた結果に対して，さらに解析法 B を用いることはタンデムな解析と呼ばれ，推奨されない．そのため，クラスタリングと他の解析手法を組み合わせた単一の目的関数を定義し，最適化する同時解析手法が多く開発されている．

本論文で扱った次元縮約 K 平均法 (De Soete & Carroll, 1994) は，単一の目的関数を最小化することにより，次元縮約とクラスタリングを同時に達成するものであった．同じ目的を持つ手法として Vichi & Kiers (2001) や，多次元尺度法，多次元展開法，および，最適尺度法とクラスタリングを組み合わせた手法などがある (Adachi, 2000; Bock, 1987; DeSarbo, Jedidi, Cool, & Schendel, 1990; Heiser, 1993)．他にも，クラスタリングによっ

て得られたクラスターごとに回帰分析を行う手法がある (西田, 2010a; Späth, 1979).

第 6 章で提案したファジィ次元縮約 K 平均法は，次元縮約 K 平均法を直接ファジィパラメータを得られるように拡張したものであるが，第 4 章や第 5 章で提案したクラスタリング手法と，他の多変量解析法を組み合わせた解析法を考えることもできる.

7.3.4　クラスター数の決定法

K 平均法におけるクラスター数の決定法には，ギャップ統計量 (Tibshirani et al., 2001) や CCC(Sarle, 1983) といった方法が提案されているが，未だ定番と呼ばれるものがない. ファジィクラスタリングの分野においても，様々なクラスター数決定法が提案されている (佐藤, 2009). 例えば，ファジィ共分散行列の行列式やそのトレースを用いる方法がある (Gath & Geva, 1989). しかしながら，一般に受け入れられている方法はない (宮本, 2009).

一般に，クラスター数を増やして行けば，二乗誤差の観点による適合度は向上する. 極論すれば，個体の数だけクラスターを考えればデータを完全に説明できることになる. しかし，分類の考えからして意味の無いことであるし，不必要にモデルを複雑にすることは避けなくてはならない. したがって，適合度とクラスター数の両方を考慮した，クラスター決定法の開発が待たれる.

要約

第 1 章 序章

分類とは物や現象を種類や性質などの基準により，似たもの同士をまとめ，似ていないものを区別することである．分類の目的は多数の対象を把握するために，少数の群に分けて整理することといえる．分類のための基準を構成することを分類学習という．分類学習は外的基準の有無によって，教師あり学習と教師なし学習の 2 種類に分けられる．分類に関する研究は，生物学，心理学，統計学など多くの分野で研究がなされてきた．特に，統計学や心理学の分野でどのような研究が行われてきたかを概観した．本論文では，分類学習のうちデータ構造の探索を目的とする教師なし分類学習，すなわちクラスタリングを計量心理学の立場から扱う．

クラスタリングの対象となるデータの形式には，個体 × 変数の行列で表現される多変量データと，個体 × 個体の行列で表現される関連性データがある．本論文で対象とするデータは，前者の多変量データである．クラスタリングは外的基準がないため，類似性を用いて分類を行う．類似性は距離によって表現されることが多く，一般的にはユークリッド距離が用いられる．代表的な 3 つのクラスタリング法，K 平均法，階層的クラスタリング，混合分布によるクラスタリングを概観し，それらの特徴をまとめた．

ファジィクラスタリングと呼ばれるファジィ理論を取り入れたクラスタリング法がある．ファジィ理論を用いたクラスタリング技法を発展させることは，理論的にも応用的にも重要である．特に，探索的に用いられるクラスタリング法は，パラメータがファジィな連続値で与えられる方が，柔軟な解釈を行うことができると考えられる．本論文の目的は，ファジィ理論を用いて結果の解釈を容易にし，より情報を引き出すことができるクラスタリング法を開発することである．

第 2 章 数学的準備

本論文で必要となる線形代数，数値最適化，および解析結果の評価指標に関する数学的基礎をまとめ，後に続く章の準備を行った．はじめに，行列・ベクトルの表記法，基本演算，基本的定義を整理し，逆行列，固有値分解，特異値分解などについてまとめた．次に，パラメータ推定に必要となるラグランジュ乗数法と正則化法を概説した．最後に，クラスタリングの分類結果の評価指標となる Adjusted Rand Index を紹介した．

第 3 章 ファジィクラスタリング

通常のクラスタリングでは，個体の集合を互いに排他的部分集合に分割することを目標とする．ある個体がクラスターに属するか否かが明確な状態である．しかしながら，実際のデータに関しては，クラスターへの所属が明確でない状態もある．そのように所属が曖昧な個体を強制的に分類したとしても，データから有益な情報を引き出しているとは言えない．そこで，クラスターへの所属が明確でない状態を，そのままを数値で表現することを考える．ファジィ集合の概念を用いて，個体がクラスターにどの程度属するかを表現する方法が，ファジィクラスタリングである．ファジィクラスタリングに対し，個体の集合を明確に分割するクラスタリングは，クリスプクラスタリングと呼ばれる．

K 平均法はメンバーシップパラメータに制約があったため，取り得る値は 0 か 1 に限られていた．ファジィ c 平均法は，メンバーシップ値が 0 以上 1 以下の連続値をとることを許し，個体がどの程度クラスターに所属するかということを表現できるよう，K 平均法の制約を緩和したものである．ファジィ c 平均法として様々な手法が提案されているが，本章ではべき乗型ファジィ c 平均法とエントロピー正則化ファジィ c 平均法，さらに，エントロピー正則化ファジィ c 平均法に，クラスターサイズパラメータを導入して拡張した，クラスターサイズ調整エントロピー正則化ファジィ c 平均法を概観した．

第 4 章 K 平均法における変数選択法の開発

我々が手にするデータにはクラスター構造に関与する本質的な変数と，クラスター構造には関係のない変数が混在していることがある．このようなデータにおいては，クラスター構造に関与する変数のみを用いて分析を行ったほうが，良い結果を得られる場合がある．本章では正則化法を用いた K 平均法における変数選択法を開発した．

まず，K 平均法においては潜在的に固定化されている，変数に対する重み係数を顕在化した定式化を行った．そして，ファジィ理論を用いることにより，変数に対する重み係数を推定可能な K 平均アルゴリズムを開発した．数値実験の結果より，提案手法は通常の

K 平均よりも高い精度でクラスタリング可能であり，有用な手法であることが示された．データ解析の結果より，提案手法は変数に対する重み係数を解釈することにより，変数選択を可能にし，クラスターの特徴を容易に理解できることが示された．

さらに，本章で提案された変数選択をともなう K 平均法に，クラスターサイズ調整パラメータを導入する事により，K 平均法の問題の 1 つである，クラスターに含まれるサンプルの数が等しくなりやすい傾向を解消する手法を開発した．数値実験の結果，クラスターサイズに偏りがあるデータセットに対しては，通常の K 平均法や変数選択をともなう K 平均法よりも良い成績を示すことが明らかとなった．

第 5 章 データ行列の最小二乗近似によるファジィクラスタリング法の開発

従来のファジィ c 平均法では，通常の K 平均法の目的関数にファジィパラメータの導入や，正則化項を加えることによって，もとの目的関数を変形させることにより，メンバーシップパラメータのファジィ化を達成していた．本章では K 平均法の目的関数を変形させることなく直接ファジィメッンバシップ値を推定する手法を開発した．したがって，提案手法は，従来の c 平均法と異なり，ファジィパラメータや正則化パラメータなどのファジィ化のために必要なチューニングパラメータを事前に設定しなくてよい．

まず，行列表現による K 平均法の定式化を用いることによって，従来行われてきたファジィ c 平均法とは異なるファジィクラスタリング法の目的関数を提案した．そして，メンバーシップパラメータをリパラメトライズすることにより，ファジィメンバーシップを推定する交互最適化アルゴリズムを開発した．数値実験の結果より，提案手法は，べき乗型 c 平均法とは異なるメンバーシップ値を推定するが，クラスタリング手法としては妥当であることが示された．データ解析の結果より，提案手法は主成分分析と類似した結果を出力することが示された．提案手法は主成分分析と目的関数が一致し，制約条件のみが異なるため，ファジィクラスタリングの特徴と，主成分分析の特徴を併せ持った中間的手法といえ，両手法と同様な解釈が可能である．

第 6 章 次元縮約をともなうファジィ K 平均法の開発

次元縮約とクリスプクラスタリングを単一の目的関数の最小化によって達成する次元縮約 K 平均法がある．クリスプクラスタリングをファジィクラスタリングへ拡張するメリットに，局所解の低減，適合度の向上，計算の安定化，妥当な解釈を可能にする，などがある．本章では，個体がクラスターから発生する事後確率として解釈できるファジィメ

ンバーシップを推定可能な，エントロピー正則化ファジィ c 平均アルゴリズムを用いて，次元縮約 K 平均法のファジィ拡張を行う．

まず，個体のクラスタリングと変数の次元縮約を同時に達成する次元縮約 K 平均法をエントロピー正則化によりファジィ化し，クラスターへの所属が明確でない個体を表現できる次元縮約ファジィ K 平均法の目的関数を提案した．そして，パラメータを推定する反復アルゴリズムを開発した．数値実験の結果により，提案手法はノイズ変数が多い条件でも高い精度で分類可能であり，クラスタリング手法として有用であることが示された．データ解析の結果より，クラスターが明確に分かれていないとき，ファジィなメンバーシップ値が解釈に有用であることが示された．

さらに，クラスターサイズを調整する機能を有した次元縮約ファジィ K 平均法の目的関数を提案し，パラメータの推定アルゴリズムを示した．この手法により，クラスターサイズが均等に分類されやすいという K 平均法の問題点を解消することができる．データ解析によりその有用性が示された

第 7 章 総合考察

第 4 章，第 5 章，第 6 章を踏まえて提案された手法についてクラスタリング手法もしくは多変量データ解析手法としての位置づけを行い，総合的考察を行った．また，教師なし分類手法としてのクラスタリングを，心理的なモデルとしてとらえることの可能性を考察した．最後に，計量心理学におけるクラスタリング法の展望を述べた．

多変量データ解析手法は様々な観点から分類できる．その 1 つに解析法の志向性は探索的か確認的かというものがある．クラスタリング手法は，外的変数を用いない教師なし学習による，対象のグループ分けであり，探索的に用いられる．本論文で行ったように，ファジィ化されたパラメータを解釈することは，探索的に用いられるクラスタリングにおいて非常に有用であると考えられる．ファジィ化の対象となるパラメータと，ファジィ化の方法という 2 つの観点から，ファジィ理論を用いたクラスタリング手法を分類し，解析手法の位置づけを行った．本論文で提案した手法は，連続的なパラメータ値を得ることによって柔軟な解釈が可能になっており，探索的データ解析において有用である．

数理モデルを扱う心理学関連分野として，心理データの解析手法の開発を目的とする計量心理学と，数理モデルによる心理現象の解明を目的とする数理心理学があるが，両者の厳密な線引きは難しい．そこで，教師なし分類学習の手法をデータ解析手法ではなく，計

算論的認知モデルとして扱うことの可能性を考察した．人間が教師なし学習状況において概念を形成することは，事例間の類似性から最適な分割を発見することすることに他ならず，クラスタリングといえる．第4章や第6章で提案した手法は，人間が行なっているクラスタリングの特徴として仮定される，概念の曖昧さや特徴次元の選択・縮約を備えている．したがって，これらの手法は，人間の教師なし概念形成モデルとしても扱うことができると考えられる．

クラスタリング手法には，解決されていない問題もあるが，そのような問題に対処する新たな方法や，解釈を容易にし，より精度の高い分類を可能にする方法が多く提案されている．たとえば，非線形な分類境界を得られるカーネル法を用いた手法，部分的に得られている外部情報を積極的に用いる制約付きクラスタリングや半教師つきクラスタリング手法，クラスタリングと他の解析手法を組み合わせて同時解析を行う手法，などがある．本論文で提案した手法と上記のような方法を組み合わせて，新たなクラスタリング法を開発することで，より良い解釈を行う事ができると考えられる．最後に，K 平均法において，いまだ未解決の大きな問題は，定番と呼ばれるクラスター数の決定法がないことである．理論的に整備されたクラスター数決定法の開発が望まれる．

引用文献

Adachi, K. (2000). Growth curve representation and clustering under optimal scaling of repeated choice data. *Behaviormetrika*, **27**(1), 15–32.

足立浩平 (2006). 多変量データ解析法 –心理・教育・社会系のための入門– ナカニシヤ出版

足立浩平 (2011). 最小二乗置換によるサイズ固定クラスタリング データ分析の理論と応用, **1**(1), 11–22.

赤穂昭太郎 (2008). 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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, **98**(3), 409–429.

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?*. New York: Oxford University Press.

Arabie, P., & Hubert, L. (1994). Cluster analysis in marketing research. In R. P. Bagozzi (Ed.), *Advanced methods in marketing research.*, 160–189. Oxford: Blackwell.

Ashby, F. G. (1992). Multidimensional models of categorization. In F. Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.. pp. 335–362.

Basu, S., Davidson, I., & Wagstaff, K. L. (Eds.) (2009). *Constrained Clustering: Advances in Algorithms, Theory, and Applications*. Boca Raton, FL: Chapman and Hall/CRC.

Bezdek, J. C. (1980). A convergence theorem for fuzzy ISODATA clustering algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-2**(1), 1–8.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Bock, H. H. (1987). *On the interface between cluster analysis, principal component analysis, and multidimensional scaling*. pp. 17–34. New York: Reidel.

Bruner, J., Goodnow, J., & Austin, A. (1956). *A Study of Thinking*. New York: Wiley.

Carroll, J. D., & Arabie, P. (1983). INDCLUS: An individual differences generalization of the adclus model and the mapclus algorithm. *Psychometrika*, **48**(2), 157–169.

Chapelle, O., Schölkopf, B., & Zien, A. (Eds.) (2006). *Semi-supervised learning*. Cambridge, Massachusetts: MIT Press.

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, **111**(2), 291–303.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, **39**(1), 1–38.

DeSarbo, W. S. (1982). GENNCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, **47**(4), 449–475.

DeSarbo, W. S., Jedidi, K., Cool, K., & Schendel, D. (1990). Simultaneous multidimensional unfolding and cluster analysis: An investigation of strategic groups. *Marketing Letters*, **2**, 129–146.

De Soete, G., & Carroll, J. D. (1994). *K*-means clustering in a low-dimensional euclidean space. In E. Diday, Y. Lechevallier, M. Schader, P. Bertrand, & B. Burtschy (Eds.), *New approaches in classification and data analysis*., 212–219. Berlin: Springer-Verlag.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, **3**, 32–57.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**(3), 211–218.

Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, **2**, 139–172.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of*

Eugenics, **7**, 179–188.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**, 234–257.

福水健次 (2010). カーネル法入門 –正定値カーネルによるデータ解析– 朝倉書店

Gan, G., Ma, C., & Wu, J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. Philadelphia: SIAM.

Gath, I., & Geva, A. B. (1989). Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(7), 773–781.

Girolami, M. (2002). Mercer kernel-based clustering in feature space. *IEEE Transactions on Neural Networks*, **13**(3), 780–784.

Gluck, M. A., & Corter, J. E. (1985). Information, uncertainty, and the utility of categories. In *Program of the Seventh Annual Conference of the Cognitive Science Society.*, 283–287.

Hadamard, J. (1923). *Lectures on Cauchy's problem in linear partial differential equations*. New Haven: Yale University Press.

Hastie, T., Tibshirani, T., & Friedman, J. H. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag.

Hattori, M., & Nishida, Y. (2009). Why does the base rate appear to be ignored? The equiprobability hypothesis. *Psychonomic Bulletin and Review*, **16**(6), 1065–1070.

Heiser, W. J., & Groenen, P. J. F. (1997). Cluster differences scaling with a within-cluster loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, **62**, 63–83.

Heiser, W. J. (1993). *Clustering in low-dimensional space*. pp. 162–173. Berlin Heidelberg NewYork: Springer.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–68.

Hruschka, H. (1986). Market definition and segmentation using fuzzy clustering methods. *International Journal of Research in Marketing*, **3**(2), 117–134.

Huang, J. Z., Ng, M. K., Rong, H., & Li, Z. (2005). Automated variable weighting in *k*-means type clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(5),

657–668.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, **2**, 193–218.

Hwang, H., Dillon, W. R., & Takane, Y. (2010). Fuzzy cluster multiple correspondence analysis. *Behaviormetrika*, **37**, 111–133.

乾敏郎 (1995). 視覚情報の競合と統合　認知科学, **2**(2), 5–20.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22–44.

Lance, G. N., & Williams, W. T. (1967). A general theory of classificatory sorting strategies. I. Hierarchical systems. *Computer Journal*, **9**, 373–80.

Laviolette, M., Seaman, J. W., Barrett, J. D., & Woodall, W. H. (1995). A probabilistic and statistical view of fuzzy methods. *Technometrics*, **37**, 249–261.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, **111**, 309–332.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability.*, 281–297.

Makioka, S., Inui, T., & Yamashita, H. (1996). Internal representation of two-dimensional shape. *Perception*, **25**, 949–966.

Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.

松香敏彦・本田秀仁・吉川詩乃 (2010). プロトタイプ理論再考　認知科学, **17**, 95–109.

Matsuka, T., Sakamoto, Y., & Chouchourelou, A. (2008). Modeling a flexible representation machinery of human concept learning. *Neural Networks*, **21**, 289–302.

McBratney, A. B., & Moore, A. W. (1985). Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, **35**, 165–185.

Mccloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets ? *Memory & Cognition*, **6**, 462–472.

McLachlan, G. J., & Basford, K. E. (1987). *Mixturemodels: Inference and applications to*

clustering. New York: Mercel Dekker.

McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81–97.

宮岸聖高・市橋秀友・本多克宏 (2001). K-L 情報量正則化 FCM クラスタリング法　日本ファジィ学会誌, **13**(4), 406–417.

宮本定明 (1999). クラスター分析入門　–ファジィクラスタリングの理論と応用–　北森出版

宮本定明 (2009). ファジィクラスタリングの有用性について　知能と情報（日本知能情報ファジィ学会誌, **21**(6), 1008–1017.

Miyamoto, S., Ichihashi, H., & Honda, K. (2008). *Algorithms for Fuzzy Clustering: Methods in c-Means Clustering with Applications*. Berlin: Springer.

Miyamoto, S., & Mukaidono, M. (1997). Fuzzy *c*-means as a regularization and maximum entropy approach. In *Proceedings of the 7th International Fuzzy Systems Association World Congress*. Vol. 2., 86–92.

宮本定明・馬屋原一孝・向殿政男 (1998). ファジィ *c*-平均法とエントロピー正則化法におけるファジィ分類関数　日本ファジィ学会誌, **10**(3), 548–557.

森崎礼子・乾敏郎 (1995). 顔の類似性判断における脳内表現の検討　電子情報通信学会技術研究報告. HIP, ヒューマン情報処理, **95**, 37–42.

Nakatani, L. H. (1972). Confusion-choice model for multidimensional psychophysics. *Journal of Mathematical Psychology*, **9**(1), 104–127.

Newell, A., & Simon, H. A. (1976). Computer science as empirical enquiry. *Communications of the ACM*, **19**(3), 113–126.

Newell, A. (1994). *Unified Theories of Cognition*. Cambridge, Massachusetts: Harvard University Press.

西田豊 (2010a). クラスタワイズ非線形回帰分析　日本行動計量学会第 38 回大会発表論文抄録集, 128–129.

西田豊 (2010b). 曖昧さを表現する概念形成モデル　知能と情報 (日本知能情報ファジィ学会誌), **22**(4), 434–442.

西田豊 (2011). 概念形成における特徴次元への重み付けのモデル化　日本認知科学会第28回大会発表論文集, P1–30.

Nishida, Y. (submitted). Variable selection in *K*-means clustering via regularization method.

西田豊・服部雅史 (2011). 基準率無視と自然頻度の幻想: 等確率性仮説に基づく実験的検討　認知科学, **18**(1), 173–189.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39–57.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, **97**(2), 185–200.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, **317**, 314–319.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, **26**, 303–343.

Rosh, E., & Mervis, C. G. (1976). Family resemblance: Studies in the internal structure of category. *Cognitive Psychology*, **7**, 382–439.

Rosseel, Y. (2002). Mixture model of categorization. *Journal of Mathematical Psychology*, **46**, 178–210.

齋藤堯幸・宿久洋 (2006). 関連性データの解析法　–多次元尺度構成法とクラスター分析–　共立出版

Sarle, W. S. (1983). Cubic clustering criterion. *SAS Technical Report A-108, Cary, NC: SAS Institute Inc.*

佐藤義治 (2009). 多変量データの分類　–判別分析・クラスター分析–　朝倉書店

de Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot.
(小林英夫 (訳) (1972). 一般言語学講義　岩波書店).

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**(2), 461–464.

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, **86**(2), 87–123.

Shiina, K. (1986). A maximum likelihood nonmetric multidimensional scaling procedure for word sequences obtained in free-recall experiments. *Japanese Psychological Research*,

28(2), 53–63.

Späth, H. (1979). Algorithm 39: clusterwise linear regression. *Computing*, **22**, 367–373.

Spearman, C. (1904). "General intelligence," Objectively determined and measured. *The American Journal of Psychology*, **15**(2), 201–292.

Suk, H. W., & Hwang, H. (2010). Regularized fuzzy clusterwise ridge regression. *Advances in Data Analysis and Classification*, **4**(1), 35–51.

Sung, K. K., & Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 39–51.

高根芳雄 (1995). 制約付き主成分分析法 –新しい多変量データ解析法– 朝倉書店

田中豊・垂水共之 (編) (1995). 統計解析ハンドブック –多変量解析– 共立出版

ten Berge, J. M. F. (1993). *Least Squares Optimization in Multivariate Analysis*. Liden: DSWO Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, **58**(1), 267–288.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in the data set via the gap statistic. *Journal of the Royal Statistics Society: Series B*, **63**(2), 411–423.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solution of Ill-posed Problems*. Washington: Winston & Sons.

豊田秀樹 (2008). データマイニング入門 東京図書

Tran, D., & Wagner, M. (2000). Fuzzy entropy clustering. In *The 9th IEEE international conference on fuzzy systems*., 152–157.

Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**, 327–352.

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. 2nd ed. New York: Springer.

Vichi, M., & Kiers, H. A. L. (2001). Factorial *k*-means analysis for two way data. *Computational Statistics and Data Analysis*, **37**, 49–64.

Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (2001). Constrained *K*-means clustering with background knowledge. In *Proceedings of the 9th International Conference on Machine Learning*., 577–584.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, **58**(301), 236–244.

Wittgenstein, L. (1958). *Philosophical investigations*. 2nd ed. Oxford: Blackwell.

柳井晴夫 (1994). 多変量データ解析法 –理論と応用– 朝倉書店

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, **8**, 338–353.

Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*., 321–328.: MIT Press.

謝辞

本論文は，筆者が大阪大学 大学院人間科学研究科 人間科学専攻 博士後期課程在学中に行った研究をまとめたものである．本論文の執筆にあたり，大阪大学 大学院人間科学研究科の足立浩平教授にご指導いただいた．大阪大学 大学院基礎工学研究科の狩野裕教授，大阪大学 大学院人間科学研究科の篠原一光准教授には，本論文をご精読いただき，貴重なご意見をいただいた．大阪大学 大学院人間科学研究科 宮本友介助教には，研究生活や計算機に関して，あたたかいご支援をいただいた．研究室の皆さまには，議論を通じて多くの示唆をいただいた．以上，ここに記して謝意を表する．