日本ファジィ学会誌 Vol.8, No.3, pp.431-439(1996)

特集 ファジィクラスタリングとその応用

ファジィクラスタリングと回帰分析†

中森 義輝

1. はじめに

回帰分析によってデータからモデルを構築しようとするとき，分析者がまず熟慮することはモデルの構造である。全域的に1つの関数で表現できない場合，データを分割して局所的に有効なモデル群を構築するアプローチがとられる。このとき，もっとも単純な方法は，データ変換をしないで，局所的に線形なモデル群を構築することである。本稿では，説明変数が多く，視覚的にデータ分割を決定することが困難な場合に，ファジィクラスタリングを応用して，データの分類と回帰モデル群の構築を同時に行う方法を紹介する。ただし，データ分割は用いる説明変数によって異なるが，説明変数の選択問題までを同時に考えることは，さらに研究を要する課題である。

まず，Hathaway and Bezdek[1]による問題の定式化と解法アルゴリズムを簡単に紹介する。定式化のポイントは，クラスタ中心としていくつかの回帰超平面を仮定して，データとクラスタとの距離を回帰残差を用いて定義するところにある。定式化はスマートであるが，望ましい結果を得るためには少し工夫が必要である。つまり，回帰超平面は無限の広がりを持つので，遠く離れた複数のデータが1つのクラスタに属すことがあり，現実的でない結果を得ることがある。本稿では，クラスタの形状を考慮するアプローチをいくつか簡単に紹介した後に，特に，クラスタリング過程でクラスタの形状をダイナミックに変更するDave[2]の考え方に注目する。4章において，Hathaway and Bezdek[1]による問題の定式化にDave[2]の考え方を加味し，さらに筆者らが若干の改良を加えたアプローチを詳しく解説する。

2. 分類と回帰の同時分析

与えられたデータ $S=\{(\boldsymbol{x}_1,\boldsymbol{y}_1),\cdots,(\boldsymbol{x}_n,\boldsymbol{y}_n)\}$ を用いて，c 個のスイッチング回帰モデル[3]

$$\boldsymbol{y}=\boldsymbol{f}_i(\boldsymbol{x};\boldsymbol{\beta}_i)+\boldsymbol{e}_i,\quad i=1,2,\cdots,c \tag{1}$$

を同定する問題を考える。ここに，$\boldsymbol{y}\in\boldsymbol{R}^r$は目的変数，$\boldsymbol{x}\in\boldsymbol{R}^s$は説明変数，$\{\boldsymbol{\beta}_i\}\in\boldsymbol{R}^{s_i}$は回帰パラメータである。

$u_{ik}\in U\in\boldsymbol{R}^{c\times n}$をデータ$(\boldsymbol{x}_k,\boldsymbol{y}_k)$のクラスタ i へのメンバシップ値$(0\leq u_{ik}\leq 1)$とする。また，$\{u_{ik}\}$は以下の条件を満たすものとする。

$$u_{i1}+\cdots+u_{in}>0,\quad i=1,2,\cdots,c \tag{2}$$
$$u_{1k}+\cdots+u_{ck}=1,\quad k=1,2,\cdots,n \tag{3}$$

$\boldsymbol{f}_i(\boldsymbol{x}_k;\boldsymbol{\beta}_i)$による $\boldsymbol{y}_k$の近似誤差の尺度を

$$E_{ik}(\boldsymbol{\beta}_i)=\|\boldsymbol{y}_k-\boldsymbol{f}_i(\boldsymbol{x}_k;\boldsymbol{\beta}_i)\|^2 \tag{4}$$

とし，つぎのような評価規範を導入する。

$$J(U,\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_c)=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m E_{ik}(\boldsymbol{\beta}_i) \tag{5}$$

ただし，m はファジィ度を与えるスムージング・パラメータ$(m>1)$である。

詳しくは4章で記述するが，モデル(1)の同定は，U を固定し，J を最小化する$\{\boldsymbol{\beta}_i\}$の探索と，逆に$\{\boldsymbol{\beta}_i\}$を固定し，J を最小化する U の探索を繰り返すアルゴリズムによる。すなわち，以下の 1. と 2. を繰り返す。

† Fuzzy Clustering and Regression Analysis

* Yoshiteru NAKAMORI
甲南大学 理学部応用数学科
Department of Applied Mathematics, Konan University

1. まず，$U^{(l)}$を固定して，関数

$$\Phi(\beta_1,\cdots,\beta_c)=J(U^{(l)},\beta_1,\cdots,\beta_c) \tag{6}$$

の大域的最小値を与えるパラメータ$\beta_i^{(l)}$を求める。

2. つぎに，$\{\beta_i^{(l)}\}$を固定して，関数

$$\Psi(U)=J(U,\beta_i^{(l)},\cdots,\beta_c^{l}) \tag{7}$$

の大域的最小値を与える$U^{(l+1)}$を求める。

以上のアプローチはFuzzy c-Regression Models(FCRM)と呼ばれる。Hathaway and Bezdek[1]は予測，分類，説明変数間の従属性の発見，あるいは異なる構造のモデルの同定へ応用したいと述べている。しかし，適切なクラスタ数c，クラスタiの適切な次元(s_iに関係)を発見することは容易ではない。

上記の方法を図1のようなテストデータ[2]に適用し，図2のようなデータの分類と回帰直線

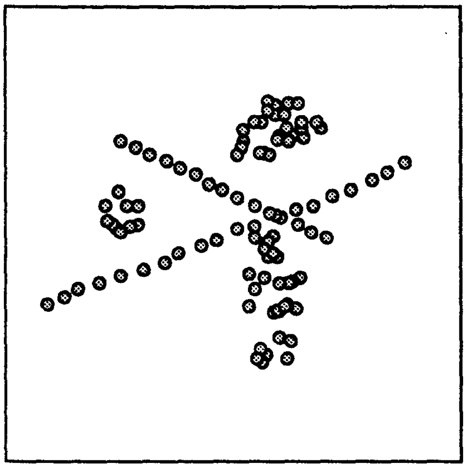


図1 異なる部分構造を持つテストデータ

を得た。ただし，スムージング・パラメータmを1.2，クラスタ数cを5とし，メンバシップ値の最も大きいクラスタにクリスプ分割して表示している。図3はクラスタ数cを6とした場合である。図2，図3より，形状の異なるクラスタが存在するとき，この方法ではうまくいかないことが理解できる。

3. クラスタ形状の考慮

前節の方法を改良する前に，クラスタの形状を考慮するクラスタリング手法を紹介する。前節の方法もその流れの中の1つである。ここでは，変数に区別をつけないので，データベクトルを$\boldsymbol{w}_k=(\boldsymbol{x}_k,\boldsymbol{y}_k)$とおく。なお，$\boldsymbol{w}_k$は行ベクトルであることに注意する。

3.1 ファジィ共分散の利用

クラスタの形状を考慮するアプローチとして最初に言及しなければならないのはGustafson and Kessel[4]の方法である。クラスタiの未知な中心を$\boldsymbol{v}_i$とし，M_iを未知な正値対称行列とする。パラメータを$\theta_i=\{\boldsymbol{v}_i,M_i\}$とおき，つぎのような距離を用いる。

$$D_{ik}(\theta_i)=(\boldsymbol{w}_k-\boldsymbol{v}_i)M_i(\boldsymbol{w}_k-\boldsymbol{v}_i)^{\top} \tag{8}$$

ただし，$\top$は転置を表す。つぎのようなクラスタリングの規範を採用する。

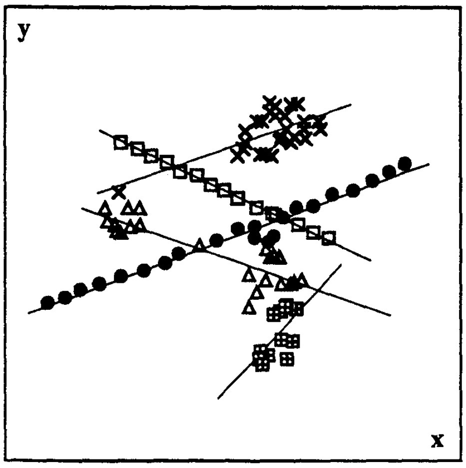


図2 FCRMによる分類と回帰(5分割)

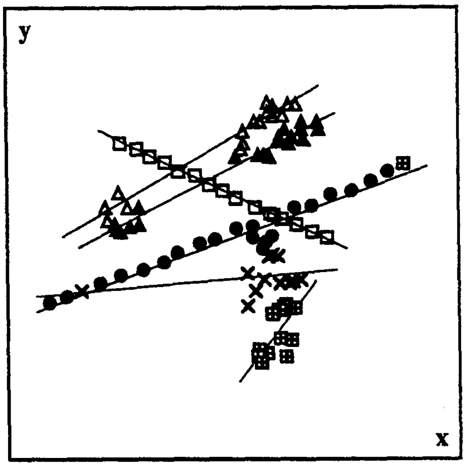


図3 FCRMによる分類と回帰(6分割)

$$J(U,\theta_1,\cdots,\theta_c)=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m\ D_{ik}(\theta_i) \qquad (9)$$

しかし，M_iが規範Jに線形に入るため，つぎのような拘束を入れる。

$$|M_i|=\rho_i>0\,(\text{fixed}),\quad i=1,2,\cdots,c \qquad (10)$$

ただし，$|M_i|$は行列式を表す。

ラグランジェ乗数法により，最適性の必要条件が以下のように求められる。まず，すべての$\theta_i=\{v_i,M_i\}$が与えられたとき，

$$u_{ik}=\left[\sum_{j=1}^{c}\left(\frac{D_{ik}(\theta_i)}{D_{jk}(\theta_i)}\right)^{\frac{1}{m-1}}\right]^{-1} \qquad (11)$$

逆に，すべてのu_{ik}が与えられたとき，

$$\boldsymbol{v}_i=\frac{\sum_{k=1}^{n}(u_{ik})^m\boldsymbol{w}_k}{\sum_{k=1}^{n}(u_{ik})^m} \qquad (12)$$

$$M_i=(\rho_i|P_i|)^{\frac{1}{n}}\ P_i^{-1} \qquad (13)$$

ただし，P_iはつぎのファジィ共分散行列である。

$$P_i=\frac{\sum_{k=1}^{n}(u_{ik})^m(\boldsymbol{w}_k-\boldsymbol{v}_i)^{\top}\ (\boldsymbol{w}_k-\boldsymbol{v}_i)}{\sum_{k=1}^{n}(u_{ik})^m} \qquad (14)$$

非正則の問題など詳しいことは4章で記述する。

アルゴリズムは$\theta_i^{(0)}=\{\boldsymbol{v}_i^{(0)},\ M_i^{(0)}\}$という初期設定から出発し，$\{D_{ik}^{0}(\theta_i^{(l)})\}$と$\{u_{ik}^{(l)}\}$を計算し，$\theta_i^{(l+1)}$を更新して，収束条件

$$\max_{i,k}\{|u_{ik}^{(l+1)}-u_{ik}^{(l)}|\}\leq\varepsilon \qquad (15)$$

が満たされるまで繰り返すというものである。

この方法はかなりうまくいくが，筆者らの経験によれば，以下のような問題もある。すなわち，メンバシップ値が単調に変化しないことがあり，どこでストップするかの判定が難しい。また，クラスタの形状は結果であって，あらかじめ仮定できないし，クラスタリング過程でそれをコントロールすることもできない。つまり，ρ_iの設定は難しい。

3.2 2つの規範のバランス

データ分布の線形構造発見のためのファジィクラスタリングは，J. C. Bezdekらの仕事が最初であろう。彼らはFuzzy c-Means (FCM) [5]を拡張し，Fuzzy c-Lines (FCL) [6]，さらに一般化したFuzzy c-Varieties (FCV) [7]について考察している。これらは，データとクラスタとの距離を直線あるいは超平面との距離で定義する方法である。

直線や超平面は無限の広がりを持つから，FCMとFCVのバランスをとる規範

$$J_{fce}=(1-\alpha)J_{fcm}+\alpha J_{fcv},\quad 0\leq\alpha\leq1 \qquad (16)$$

が提案されている[7]。ただし，J_{fcm}，J_{fcv}はそれぞれFCMとFCVの規範であるが，詳細は省略する。本特集の宮本[8]を参照されたい。この方法は，Fuzzy c-Elliptotypes (FCE)と呼ばれる。

あらかじめαの値を設定しなければならないので，いくつかのαについて試行して，結果を比較することになる。形状の異なるクラスタが混在しているときは，なかなか満足できる分類は得られない。そこで，つぎのような方法が開発された。

3.3 適応型クラスタリング

ここでは，Dave[2]により提案された，クラスタリング中にαをダイナミックに変更する方法を紹介する。クラスタリングの評価規範は，

$$J_{afc}=\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m Z_{ik} \qquad (17)$$

$$Z_{ik}=(1-\alpha_i)\,G_{ik}+\alpha_i V_{ik} \qquad (18)$$

ただし，G_{ik}がデータとクラスタのファジィ重心$\boldsymbol{v}_i$との2乗距離であるのに対し，V_{ik}はデータと線形多様体との2乗距離である。

各クラスタiについて，そのファジィ散布行列

$$S_i=\sum_{k=1}^{n}(u_{ik})^m(\boldsymbol{w}_k-\boldsymbol{v}_i)^{\top}(\boldsymbol{w}_k-\boldsymbol{v}_i) \qquad (19)$$

の固有値$\{\lambda_{ij}\}$を用いて，α_iをつぎのように定める。

$$\alpha_i=1-\frac{\min_j\{\lambda_{ij}\}}{\max_j\{\lambda_{ij}\}},\quad i=1,2,\cdots,c \qquad (20)$$

クラスタリングの各ラウンドにおいて，クラスタの形状に応じて適応的に荷重パラメータα_iを変

更する。この方法はAdaptive Fuzzy c-Elliptotypes(AFC)と呼ばれる。

この方法はなかなかのアイデアである。あらかじめクラスタの形状に関する仮定をおかなくてもよく，異なる形状のクラスタが存在するときでも，かなりうまくいく。だだし，メンバシップ値の最大変化は単調減少ではなく，初期設定のみならず，終了判定条件により結果が異なる。また，クラスタが密集している場合は失敗することが多い。

4. 適応的データ分類と回帰分析

ここでは，Hathaway and Bezdek[1]による問題の定式化にDave[2]の方法を付け加えたアプローチを詳しく解説する。ただし，問題設定をより一般的にするために，回帰モデルの目的変数群と説明変数群の他に，データ分割のみに影響を与える条件変数群を導入する。条件変数を導入するのは，データ分割には重要な役割を果たすが，回帰モデルの説明変数としては不適当な変数の存在，あるいは，その逆のケースがありうることを考慮するものである。たとえば，少数の離散的な値のみを取る質的変数に近い変数を条件変数とすることが考えられる。これは，実はファジィ予測モデルの構築を強く意識したものである。ファジィモデリングとの関連，および，具体的な例については領家・中森[9]を参照されたい。

なお，ここでは詳細に記述するため，前節までと重複する部分があることをお断りする。

4.1 記号の定義

目的変数を $y_1, y_2, \cdots, y_r$, 説明変数を $x_1, x_2, \cdots, x_s$, 条件変数を $z_1, z_2, \cdots, z_t$とする。ここで，説明変数集合と条件変数集合には共通部分があってよく，共通な変数の個数を u とする。すなわち，

$$u = |\{x_1, x_2, \cdots, x_s\} \cap \{z_1, z_2, \cdots, z_t\}| \tag{21}$$

ただし，ここでは$|\cdot|$は集合の要素数を表す。したがって，クラスタリング空間は $R^{r+s+t-u}$である。これらの変数に対して n 組のデータが与えられているとする。ただし，データの単位のばらつきによる悪影響を避けるために，平均0，分散1に標準化しておく。そのため，回帰式を元の単位で記述するためには逆変換が必要である。

データベクトルとデータ行列を以下のような記号で表す。

- 目的変数データベクトル

$$\boldsymbol{y}_k = (y_{k1}, y_{k2}, \cdots, y_{kr}), \quad k = 1, 2, \cdots, n$$

- 説明変数データベクトル

$$\boldsymbol{x}_k = (x_{k1}, x_{k2}, \cdots, x_{ks}), \quad k = 1, 2, \cdots, n$$

- 条件変数データベクトル

$$\boldsymbol{z}_k = (z_{k1}, z_{k2}, \cdots, z_{kt}), \quad k = 1, 2, \cdots, n$$

- 目的変数データ行列

$$Y = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1r} \\ y_{21} & y_{22} & \cdots & y_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nr} \end{bmatrix}$$

- 説明変数データ行列

$$X = \begin{bmatrix} \boldsymbol{x}_1^o \\ \boldsymbol{x}_2^o \\ \vdots \\ \boldsymbol{x}_n^o \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{x}_1 \\ 1 & \boldsymbol{x}_2 \\ & \vdots \\ 1 & \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1s} \\ 1 & x_{21} & \cdots & x_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{ns} \end{bmatrix}$$

ファジィクラスタリングは，各データをいずれかのクラスタに完全に帰属させる従来のクラスタリングと異なり，各データの各クラスタへの帰属度，すなわちメンバシップ値を求めるものである。u_{ik}をデータベクトル$(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{z}_k)$のクラスタ i へのメンバシップ値とし，つぎの行列を導入する。

- メンバシップ行列

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{c1} & u_{c2} & \cdots & u_{cn} \end{bmatrix}$$

ただし，c はクラスタ数であり，初期設定として与えなければならない。メンバシップ値$\{u_{ik}\}$は以下の条件を満たすものとする。

(i) $0 \leq u_{ik} \leq 1, \quad \forall i, \forall k$

(ii) $0<\sum_{k=1}^{n} u_{ik}<n, \quad \forall i$

(iii) $\sum_{i=1}^{c} u_{ik}=1, \quad \forall k$

条件(ii)はすべてのクラスタは空でないという要請，条件(iii)は各データについて，すべてのクラスタへの帰属度の合計を1とする要請である。

ところで，クラスタリングはすべての変数の空間 $\boldsymbol{R}^{r+s+t-u}$ において実行する。この空間におけるデータベクトルをつぎのように定義しておく。

- クラスタリングデータベクトル
  $\boldsymbol{w}_k=(\boldsymbol{x}_k, \boldsymbol{y}_k, \boldsymbol{z}_k)=(w_{k1}, w_{k2}, \cdots, w_{kv})$

ただし，$v=r+s+t-u$ である。クラスタのファジィ重心を次式で定義する。

$$\boldsymbol{v}_i=\frac{\sum_{i=1}^{c}(u_{ik})^m \boldsymbol{w}_k}{\sum_{k=1}^{n}(u_{ik})^m}, \quad i=1,2,\cdots,c \tag{22}$$

ここで，m はファジィ度を表す1より大きい実数である。また，クラスタ i の散布行列を次式で定義する。

$$S_i=\sum_{k=1}^{n}(u_{ik})^m(\boldsymbol{w}_k-\boldsymbol{v}_i)^{\top}(\boldsymbol{w}_k-\boldsymbol{v}_i) \tag{23}$$

クラスタ i に対応する線形モデルを

$$y_j=a_{j0}^i+a_{j1}^i x_1+\cdots+a_{js}^i x_s \tag{24}$$

と表し，つぎの行列を導入する。

- 回帰係数行列

$$A_i=\begin{bmatrix} a_{10}^i & a_{20}^i & \cdots & a_{r0}^i \\ a_{11}^i & a_{21}^i & \cdots & a_{r1}^i \\ \vdots & \vdots & \ddots & \vdots \\ a_{1s}^i & a_{2s}^i & \cdots & a_{rs}^i \end{bmatrix}$$

クラスタ i に対応する線形モデルによるデータ y_k の予測残差ベクトル，および全データの予測残差行列は以下のように表現される。

- 予測残差ベクトル
  $\boldsymbol{y}_k-\boldsymbol{x}_k^{\circ} A_i, \quad i=1,2,\cdots,c, \quad k=1,2,\cdots,n$
- 予測残差行列
  $Y-XA_i, \quad i=1,2,\cdots,c$

4.2 データの分類と回帰の同時分析

データ分類と部分線形モデルを同時に決定するために最小化すべき目的関数として次式を用いる。

$$J(U; A_1,\cdots,A_c; \bar{z}_1,\cdots,\bar{z}_c) =\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m E_{ik}(A_i, \bar{z}_i) \tag{25}$$

ここで，$E_{ik}(A_i, \bar{z}_i)$ は次式で定義する。

$$E_{ik}(A_i, \bar{z}_i)=\alpha_i\times(\boldsymbol{y}_k-\boldsymbol{x}_k^{\circ} A_i)(\boldsymbol{y}_k-\boldsymbol{x}_k^{\circ} A_i)^{\top} +(1-\alpha_i)\times\eta\times(\boldsymbol{z}_k-\bar{z}_i)(\boldsymbol{z}_k-\bar{z}_i)^{\top} \tag{26}$$

ただし，η は2つの項のスケールを調整するパラメータ，$\bar{z}_i$ はクラスタ i の条件変数データの中心ベクトルである。$\boldsymbol{v}_i$ ではなく $\bar{z}_i$ を用いるのは，クラスタリング結果を用いてファジィ予測モデルなどを構築しやすいようにとの配慮であるが，画像認識などの認識問題に対しては $\boldsymbol{v}_i$ を用いた方がよい。

ここでは3.3節の方法を採用して，クラスタのデータ分布構造に応じて $E_{ik}(A_i, \bar{z}_i)$ の2つの項の重視度をダイナミックに変更するアプローチをとる。これを担当するパラメータが $\alpha_1, \alpha_2, \cdots, \alpha_c$ である。すなわち，λ_{ij} をクラスタ i のファジィ散布行列 S_i の第 j 固有値とするとき，

$$\alpha_i=1-\frac{\min_j\{\lambda_{ij}\}}{\max_j\{\lambda_{ij}\}}, \quad i=1,2,\cdots,c \tag{27}$$

により適応的に規範の第1項と第2項のバランスをとりながらクラスタリングを実行する。つまり，線形関係が抽出できそうな場合は最小固有値 $\min\{\lambda_{ij}\}$ が0に近くなることから，α_i は1に近くなり，$E_{ik}(A_i, \bar{z}_i)$ の第1項が重視される。そうでない場合は第2項を重視してクラスタリングを行う。

クラスタリングはパラメータ m, c, η を与え，規範 J を最小化するように

$\{u_{ik}, A_i, \bar{z}_i; i=1,2,\cdots,c, k=1,2,\cdots,n\}$

を決定していくプロセスである。3種類のパラメータを同時に決定することは困難であるから，2種類のパラメータが与えられたとき，残りのパラメータを最適性の必要条件から決定するアプローチをとる。最適性の必要条件はそれぞれ以下のように求められる。

(a) 条件変数データのクラスタ中心ベクトル

メンバシップ行列 U と回帰係数行列 A_1, $A_2, \cdots, A_c$が与えられたとき，最小化すべき規範は

$$J_1(\bar{z}_1, \cdots, \bar{z}_c) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m(z_k-\bar{z}_i)(z_k-\bar{z}_i)^\top$$

であるから，必要条件$\partial J_1/\partial \bar{z}_i = O_{1\times t}$(零ベクトル)より

$$\bar{z}_i = \frac{\sum_{k=1}^{n}(u_{ik})^m z_k}{\sum_{k=1}^{n}(u_{ik})^m}, \quad i=1,2,\cdots,c \tag{28}$$

(b) 回帰係数行列

メンバシップ行列 U と条件変数データのクラスタ中心ベクトル$\bar{z}_1, \bar{z}_2, \cdots, \bar{z}_c$が与えられたとき，最小化すべき規範は

$$J_2(A_1, \cdots, A_c) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m(\boldsymbol{y}_k - \boldsymbol{x}_k^o A_i)(\boldsymbol{y}_k - \boldsymbol{x}_k^o A_i)^\top$$

であるから，最適性の必要条件は

$$\frac{\partial J_2}{\partial A_i} = \frac{\partial}{\partial A_i} tr[(Y-XA_i)^\top D_i(Y-XA_i)] = O_{s\times r}$$

ただし，$O_{s\times r}$は零行列，trA は行列 A の対角成分の和を表す。これより次式を得る。

$$A_i = [X^\top D_i X]^{-1}X^\top D_i Y \tag{29}$$

ただし，

$$D_i = \begin{bmatrix} (u_{i1})^m & 0 & \cdots & 0 \\ 0 & (u_{i2})^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (u_{in})^m \end{bmatrix}$$

(c) メンバシップ行列

条件変数データのクラスタ中心ベクトル$\bar{z}_1$, $\bar{z}_2, \cdots, \bar{z}_c$と回帰係数行列 $A_i, A_2, \cdots, A_c$，およびパラメータ $\alpha_1, \alpha_2, \cdots, \alpha_c, \eta$ が与えられたとき，最小化すべき規範は次式となる。

$$J_3(U) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m E_{ik} + \sum_{k=1}^{n}\mu_k\left(\sum_{i=1}^{c}u_{ik}-1\right)$$

ただし，$\mu_1, \mu_2, \cdots, \mu_n$はラグランジュ乗数である。$u_{ik}=s_{ik}^2$と置き，最適性の必要条件$\partial J_3/\partial s_{ik}=0$から

$$s_{ik}\left(ms_{ik}^{2(m-1)}E_{ik}+\mu_k\right)=0$$

すなわち，

$$s_{ik}^2 = \left(\frac{-\mu_k}{mE_{ik}}\right)^{\frac{1}{m-1}}$$

を得る。$\sum_{i=1}^{c}s_{ik}^2=1$ を利用して$(-\mu_k)^{\frac{1}{m-1}}$を求め，最終的に次式を得る。

$$u_{ik} = \left[\sum_{j=1}^{c}\left(\frac{E_{ik}}{E_{jk}}\right)^{\frac{1}{m-1}}\right]^{-1} \tag{30}$$

パラメータ $\alpha_1, \cdots, \alpha_c$および η の値はメンバシップ行列 U を求めるときにのみ必要となることに注意する。ここで，第 k データに対して，$E_{ik}=0$ となるクラスタの数が$|I|$であれば，

$$\begin{cases} u_{ik} = \dfrac{1}{|I|}, & E_{ik}=0 \\ u_{ik}=0, & E_{ik}>0 \end{cases} \tag{31}$$

によりメンバシップ値を計算する。

以下にクラスタリング・アルゴリズムを記述する。

ステップ 1：$l=0$ とする。パラメータ m, c, η を与える。収束判定値 ε 与える。メンバシップ値の初期値

$$\{u_{ik}^{(l)};\ i=1,2,\cdots,c,\ k=1,2,\cdots,n\}$$

を与える。

ステップ 2：次式により回帰係数を求める。

$$A_i^{(l)} = [X^\top D_i^{(l)} X]^{-1}X^\top D_i^{(l)} Y$$

ステップ 3：次式により条件変数データのクラスタ中心ベクトルを求める。

$$\bar{z}_i^{(l)} = \frac{\sum_{k=1}^{n}(u_{ik}^{(l)})^m z_k}{\sum_{k=1}^{n}(u_{ik}^{(l)})^m}, \quad i=1,2,\cdots,c$$

ステップ 4：次式により全変数データのクラスタ中心ベクトルを求める。

$$v_i^{(l)}=\frac{\sum_{k=1}^{n}(u_{ik}^{(l)})^m w_k}{\sum_{k=1}^{n}(u_{ik}^{(l)})^m},\quad i=1,2,\cdots,c$$

クラスタの散布行列

$$S_i^{(l)}=\sum_{k=1}^{n}(u_{ik}^{(l)})^m(w_k-v_i^{(l)})^{\top}(w_k-v_i^{(l)})$$

の固有値$\{\lambda_{ij}^{(l)}\}$を求め,

$$\alpha_i^{(l)}=1-\frac{\min_j\{\lambda_{ij}^{(l)}\}}{\max_j\{\lambda_{ij}^{(l)}\}},\quad i=1,2,\cdots,c$$

を計算する。

ステップ 5：次式によりメンバシップ値を更新する。

$$u_{ik}^{(l+1)}=\left[\sum_{j=1}^{c}\left(\frac{E_{ik}^{(l)}}{E_{jk}^{(l)}}\right)^{\frac{1}{m-1}}\right]^{-1}$$

ただし，$E_{ik}^{(l)}=0$となるiが存在する場合は(31)による。

ステップ 6：収束判定条件

$$\max_{i,k}\left\{\left|u_{ik}^{(l+1)}-u_{ik}^{(l)}\right|\right\}<\varepsilon$$

を満たせば終了。満たさなければ，$l=l+1$として，**ステップ 2** へ戻る。

以上の方法をここではAdaptive Fuzzy c-Regression Models(AFCR)[9]と呼ぶ。図1のような2次元のデータ分布が与えられたとき，上記の手法によれば図4のようなデータの分類と回帰直線が同定される。ここで，図の縦軸，横軸をそれぞれ目的変数y,説明変数xとみなし,条件変数zはxをそのまま用いている。図4においては,図2と同様，スムージング・パラメータmを1.2,クラスタ数cを5とし，メンバシップ値の最も大きいクラスタにクリスプ分割して表示している。なお，$\eta=1.0$としている。データの密集部が完全には分類されていない。

図5はクラスタ数cを6とした場合である。このとき，パラメータ$\alpha_1,\alpha_2,\cdots,\alpha_6$が収束していく様子を図6に示す。クラスタの形状に適応して変化していることが読みとれる。たとえば，線形構造を捉えているクラスタ3と5においてはα_3,α_5は1.0に収束し，逆に，クラスタ6ではα_6は0.0に近づいている。

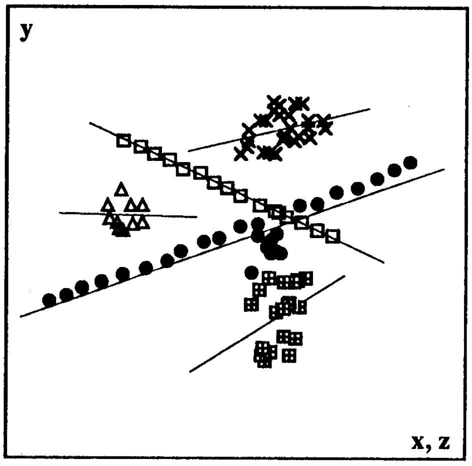


図4　AFCRによる分類と回帰(5分割)

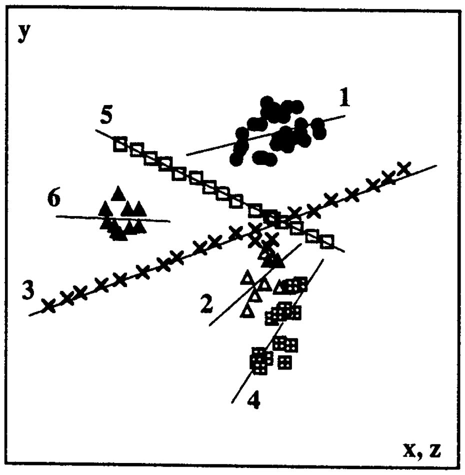


図5　AFCRによる分類と回帰(6分割)

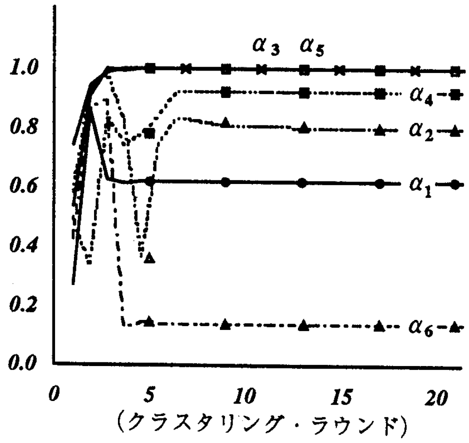


図6　パラメータの収束過程(6分割)

5. おわりに

ファジィクラスタリングを利用したデータの分類と回帰の同時分析法について解説した。しかし，実際の応用ではなかなかすっきりとはいかない。それは，説明変数の選択問題が絡むからである。また，クラスタ数をいくらにするかについても，決断が困難なケースがしばしばある。本稿で紹介しなかった方法や例題などは中森[10]を参照されたい。

本稿では，一般的に記述するため，複数の目的変数について同時に回帰モデルを同定するように定式化した。しかし，ある1つの目的変数に注目してデータの分類と回帰を行った方が，その目的変数について精度のよい予測モデルが構築できることは言うまでもない。

なお，本稿の例題では，スムージング・パラメータ m の値を1.2としているが，筆者らの経験では m の値は1に近い，すなわちクリスプなクラスタリングに近い方がうまくいく場合が多い。ただし，ファジィクラスタリングにおいては，一般に，m の値が小さいほど結果の初期分割への依存度が高くなるので，初期分割を変更して試行を繰り返して比較しなければならない。

ところで，線形モデルの代わりに積極的に非線形モデルを仮定して，ファジィクラスタリングを適用することも可能である。たとえば，A_iを対称行列として

$$E_{ik}(A_i, \boldsymbol{v}_i) = ([(\boldsymbol{w}_k - \boldsymbol{v}_i) A_i (\boldsymbol{w}_k - \boldsymbol{v}_i)^\top]^{\frac{1}{2}} - 1)^2$$

とおけば，楕円体

$$(\boldsymbol{w}_k - \boldsymbol{v}_i) A_i (\boldsymbol{w}_k - \boldsymbol{v}_i)^\top = 1$$

状のデータ分布を発見できる[12]。あるいは，

$$E_{ik}(A_i, \boldsymbol{b}_i, c_i) = (\boldsymbol{w}_k A_i \boldsymbol{w}_k^\top + \boldsymbol{b}_i \boldsymbol{w}_k^\top + c_i)^2$$

とおけば，2次曲線

$$\boldsymbol{w}_k A_i \boldsymbol{w}_k^\top + \boldsymbol{b}_i \boldsymbol{w}_k^\top + c_i = 0$$

を発見することができる[12]。さらに，超直線，超平面などへのデータフィッティングを考慮するファジィクラスタリングの研究が進められている[13]。

謝 辞

本稿を執筆するにあたり，大阪大学大学院生の領家美奈さんの協力を得た。感謝の意を表したい。

参 考 文 献

[1]R. J. Hathaway and J. C. Bezdek : Switching regression models and fuzzy clustering. *IEEE Trans. on Fuzzy Systems*, Vol.1, No.3, pp.195-204, 1993.

[2]R. N. Dave : An adaptive fuzzy c-elliptotype clustering algorithm. *In* : Proc. of NAFIPS 90 : Quater Century of Fuzziness, Vol.I, pp.9-12, 1990.

[3]D. W. Hosmer, Jr. : Maximum likelihood estimates of the parameters of a mixture of two regression lines. *Communications in Statistics*, Vol.3, No.10, pp.995-1005, 1974.

[4]D. E. Gustafson and W. C. Kessel : Fuzzy clustering with a fuzzy covariance matrix. *In* : Proc. of IEEE CDC, pp.761-766, San Diago, CA, 1979.

[5]J. C. Bezdek : *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[6]J. C. Bezdek et al. : Detection and characterization of cluster substructure I. linear structure : fuzzy c-lines. *SIAM J. Appl. Math.*, Vol.40, No.2, pp.339-357, 1981.

[7]J. C. Bezdek et al. : Detection and characterization of cluster substructure II. fuzzy c-varieties and convex combinations thereof. *SIAM J. Appl. Math.*, Vol.40, No.2, pp.358-372, 1981.

[8]宮本定明：ファジィクラスタリングあれこれ. 日本ファジィ学会誌，第8巻，第3号，pp.423-430，1996.

[9]領家美奈・中森義輝：適応型ファジィクラスタリングによる分類と回帰の同時分析. 日本ファジ

ィ学会誌, 第8巻, 第1号, pp.136-146, 1996.

[10]中森義輝：ファジィモデリング. オーム社, 1994.

[11]R. N. Dave and K. Bhaswan：Adaptive fuzzy c-shells clustering and detection of ellipses. *IEEE Trans. on Neural Networks*, Vol.3, No. 5, pp.643-662, 1992.

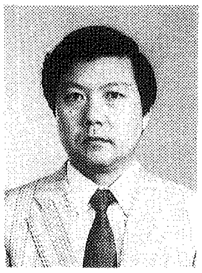
[12]R. Krishnapuram, H. Frigui and O. Nasraoui：Quadratic shell clustering algorithms and their applications. *Pattern Recognition Letters*, Vol.14, No.7, pp.545-552, 1993.

[13]R. Krishnapuram, H. Frigui and O. Nasraoui：Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation：Part I and II. *IEEE Trans. on Fuzzy Systems*, Vol.3, No.1, pp.44-60, 1995.

(1996年3月15日　受付)

[問い合わせ先]
〒658 神戸市東灘区岡本8-9-1
甲南大学 理学部 応用数学科
中森 義輝
TEL ：078-435-2533
FAX ：078-452-5507
Email：nakamori@edu3.math.konan-u.ac.jp

著者紹介



中森 義輝 (なかもり よしてる)

甲南大学 理学部応用数学科

1979年 京都大学大学院工学研究科数理工学専攻博士課程修了。1981年より甲南大学理学部応用数学科勤務。現在教授。1984年10月より1985年11月まで, 国際応用システム解析研究所(オーストリア)研究員。1986年4月より環境庁国立環境研究所客員研究員。1992年9月より大連理工大学客員教授兼務。日本ファジィ学会, 計測自動制御学会, 環境科学会, IEEEなどの会員。