数理解析研究所講究録
1015 巻 1997 年 32-37

数理計画を用いたパターン分類

甲南大学　理学部　　中山弘隆 (Hirotaka Nakayama)

1. はじめに

パターン分類の問題に対し，与えられたデータからその規則性を抽出する技法は現在 computational intelligence の中で最も重要な課題の一つとなっている．統計的判別分析は古典的手法としてよく知られているが，データの分布に対する仮定が現実には厳しすぎることも多い ([2])．machine learning の立場からニューラルネットワークはデータ分布に依存せず，非線形性の強い問題にも適用でき，これまでにも数多くの報告がなされている ([2],[3])．一方，Mangasarian は非常に早い時期に数理計画を用いたパターン分類の手法 (MSM) を提案しており ([4])，近年には医療診断問題に適用しその有効性を主張する ([10]) とともに，カスケードニューラルネットワークとして解釈できることを報告している ([5],[8])．本報告では MSM の問題点を指摘し，それらを克服するいくつかの試みについて述べる．

2. Multisurface Method (MSM)

n 次元 Euclid 空間の部分集合 X 上に与えられたデータが 2 つのカテゴリー $\mathcal{A}$, $\mathcal{B}$ のいずれかに属するものとする． 行列 A の各行ベクトルはカテゴリー $\mathcal{A}$ に属するデータの座標ベクトルとし，行列 B の各行ベクトルはカテゴリー $\mathcal{B}$ に属するデータの座標ベクトルとする．簡単化のためにカテゴリー $\mathcal{A}$ に属する点の集合を A によって表し，カテゴリー $\mathcal{B}$ に属する点の集合を B で表すことにする．Mangasarian (1968) によって提案された MSM は繰り返し線形計画法を解くことにより，2 つの集合 A および B を分離する区分的に線形な曲面を求めようとするものである．その主な着想はできる限り多くのデータを正しく分類する互いに平行な 2 つの超平面 $g(\boldsymbol{u}) = \boldsymbol{x}^T\boldsymbol{u} = \alpha$, $g(\boldsymbol{u}) = \boldsymbol{x}^T\boldsymbol{u} = \beta$ を求めることである．

アルゴリズムは次のようになる．

Step 1． k 回目の繰り返しで次のＬＰを解く．(最初は $k = 1$ とする):

(I)　Maximize　$\phi_i(A,B) = \alpha - \beta$
　subject to

$$
\begin{aligned}
A\boldsymbol{u} &\geqq \alpha\mathbf{1} \\
B\boldsymbol{u} &\leqq \beta\mathbf{1} \\
-\mathbf{1} \leqq\ &\boldsymbol{u} \leqq \mathbf{1} \\
\boldsymbol{p}_i^T\boldsymbol{u} &\geqq \frac{1}{2}\left(\frac{1}{2} + \boldsymbol{p}_i^T\boldsymbol{p}_i\right) \qquad (1)
\end{aligned}
$$

ここで， $\boldsymbol{p}_i$ は $\boldsymbol{p}_1^T = (\frac{1}{\sqrt{2}}, 0, \ldots, 0), \boldsymbol{p}_2^T = (-\frac{1}{\sqrt{2}}, 0, \ldots, 0), \cdots, \boldsymbol{p}_{2n}^T = (0, \ldots, 0, -\frac{1}{\sqrt{2}})$ のうちの 1 つである．制約 (1) は無意味な解 $\boldsymbol{u} = 0, \alpha = 0, \beta = 0$ を避けるために，$\boldsymbol{u}^T\boldsymbol{u} \geqq \frac{1}{2}$ を線形近似したものである．

各 i $(1 \leqq i \leqq 2n)$ に対して LP 問題 (I) を解き，それらの解の中から最も多くのデータを正しく分類している超平面を採用する．その解を $\boldsymbol{u}^*, \alpha^*, \beta^*$　とし，それに対応する目的関数の値を $\phi^*(A,B)$ とする．

もし， $\phi^*(A,B) > 0$ ならば，完全な分離超平面 $g(\boldsymbol{u}^*) = (\alpha^* + \beta^*)/2$ を得る．$\tilde{A}^k = \{\boldsymbol{x} \in X |\ g(\boldsymbol{u}^*) \geqq (\alpha^* + \beta^*)/2\}$, $\tilde{B}^k = \{\boldsymbol{x} \in X |\ g(\boldsymbol{u}^*) < (\alpha^* + \beta^*)/2\}$ とおく．$\tilde{A}^k$ および $\tilde{B}^k$ はこの段階で正しく分類されたカテゴリー $\mathcal{A}$ および $\mathcal{B}$ の部分集合である．Step 3 に移る．

もし，$\phi^*(A,B) \leqq 0$ならば Step 2に移る．

Step 2．まず, 集合Aから $\boldsymbol{x}^T\boldsymbol{u}^* > \beta^*$ なる点を除く．この取り除かれた点の集合をA^kで表そう．分離超平面を $g(\boldsymbol{u}^*) = (\beta^* + \tilde{\beta})/2$ (ただし，$\tilde{\beta} = \mathrm{Min}\ \{\boldsymbol{x}^T\boldsymbol{u}^* |\ \boldsymbol{x} \in A^k\}$) とする．ここで，$\tilde{A}^k = \{\boldsymbol{x} \in X |\ g(\boldsymbol{u}^*) > (\beta^* + \tilde{\beta})/2\}$ とする．集合$\tilde{A}^k$ はこの段階で正しく分類されたカテゴリー $\mathcal{A}$ の部分集合である．集合 $X\backslash\tilde{A}^k$ をXとし，$A\backslash A^k$ をAとする．

次に，集合Bから $\boldsymbol{x}^T\boldsymbol{u}^* < \alpha^*$ なる点を除く．この取り除かれた点の集合を B^k で表す．分離超平面を $g(\boldsymbol{u}^*) = (\alpha^* + \tilde{\alpha})/2$ (ただし，$\tilde{\alpha} = \mathrm{Max}\ \{\boldsymbol{x}^T\boldsymbol{u}^* |\ \boldsymbol{x} \in B^k\}$) とする．ここで，$\tilde{B}^k = \{\boldsymbol{x} \in X |\ g(\boldsymbol{u}^*) < (\alpha^* + \tilde{\alpha})/2\}$ とする．集合 $\tilde{B}^k$ はこの段階で正しく分類されたカテゴリー $\mathcal{B}$の部分集合である．集合 $X\backslash\tilde{B}^k$ をX をとし，$B\backslash B^k$ を Bとする．

$k = k + 1$ として Step 1に戻る．

Step 3. 上で得られた各段階での分離超平面の適当な部分をとって集合A と B に対する区分的に線形な全体としての分離超平面を構成する．

注 最終段階(p回目とする)において, $\mathcal{A}$ の領域を $\tilde{A}^1 \cup \tilde{A}^2 \cup \ldots \cup \tilde{A}^p$，同様に $\mathcal{B}$の領域を $\tilde{B}^1 \cup \tilde{B}^2 \cup \ldots \cup \tilde{B}^p$ として得る．新たな点が与えられたとき，その点はどれかの$i = 1, \ldots p$に対し，$\tilde{A}^i$，もしくは$\tilde{B}^i$ のどちらかに属するので $i = 1, 2, \ldots, p$ の順にどちらに属するかを調べていけばよい．

3. 多目的計画法の適用

MSM は極めて非線形度の高い問題にも適用でき，しかもバックプロパゲーションによる多層ニューラルネットのように構造決定の困難さやあらかじめ定めておくべきパラメータが不必要であるというメリットがあり，医療診断の問題への適用によってその有効性も報告されている ([9])．しかし，我々のこれまでの観測ではMSMはあまりに複雑な分離曲面を作ることが往々にしてあり，容易に分かるようにこのことはデータのノイズ等にあまりにも敏感で汎化能力の劣化につながる．この原因として各段階で常に2つの平行な超平面を求めていることがあげられる ([7])．

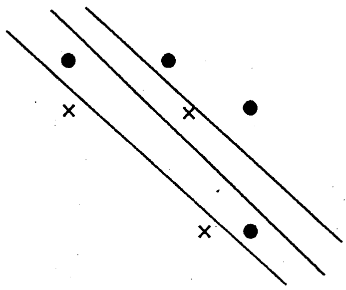

Fig. 1 Classification by MSM

そこで，片側に集合Bを置きながら集合Aの点をできる限り多く正しく分類する1つの超平面を求めることを試みる．これは次の多目的計画の問題を解くことによって達成される．

(II) Maximize $(A\boldsymbol{u} - \alpha\mathbf{1})$
subject to

$$\begin{aligned} B\boldsymbol{u} - \beta\mathbf{1} &\leqq \mathbf{0} \\ -\mathbf{1} \leqq \boldsymbol{u} &\leqq \mathbf{1} \end{aligned}$$

実際に問題(II)を解くには，適当なスカラー化によって通常の数理計画の問題に帰着させる:

(II') Maximize $\sum_{i=1}^{m}(y_i^{+}-y_i^{-})$
subject to

$$\begin{array}{rcl} A\boldsymbol{u}-\beta\mathbf{1} & = & \boldsymbol{y}^{+}-\boldsymbol{y}^{-} \\ B\boldsymbol{u}-\beta\mathbf{1} & \leqq & \mathbf{0} \\ -\mathbf{1} \ \leqq \ \boldsymbol{u} & \leqq & \mathbf{1} \\ \boldsymbol{y}^{+},\ \boldsymbol{y}^{-} & \geqq & \mathbf{0} \end{array}$$

この手法により Fig. 2のような我々の直感に近い分離曲面が得られる．一般には， A と B の役割の対称性を考え， (II')に加え，次の問題を解く．

(III') Minimize $\sum_{i=1}^{m}(z_i^{+}-z_i^{-})$
subject to

$$\begin{array}{rcl} A\boldsymbol{u}-\alpha\mathbf{1} & \geqq & \mathbf{0} \\ B\boldsymbol{u}-\alpha\mathbf{1} & = & \boldsymbol{z}^{+}-\boldsymbol{z}^{-} \\ -\mathbf{1} \ \leqq \ \boldsymbol{u} & \leqq & \mathbf{1} \\ \boldsymbol{z}^{+},\ \boldsymbol{z}^{-} & \geqq & \mathbf{0} \end{array}$$

(II'), (III')を解く際，無意味な解$\boldsymbol{u}=0,\ \boldsymbol{y}=0,\ \boldsymbol{z}=0,\ \alpha=0, \beta=0$を避けるために制約(1)を付加する．(II')と(III')の解のうち正しく分類できているデータの数が多い方を採用する．全体としての分離曲面の求め方はMSMと同様である．

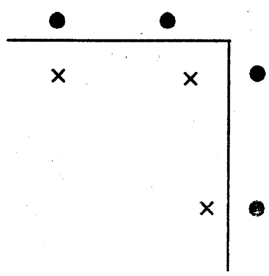

Fig. 2 Classification by (II')

【ゴールプログラミングの適用】

多目的計画法と同様，超平面の片側に一方の集合を置いてはいるが，別の集合の点を分類させるとき誤認識のデータのみを考慮し，できる限りその程度を小さくさせようとする所が多目的計画法とは異なる別の方法も考案されている．この手法はBenett-Mangasarian (1992)によって提案され，RLPD(Robust Linear Programming Discrimination)と命名されているが，容易に分かるように上記多目的計画法の適用の際，ゴールプログラミングとして定式化したものである(Nakayama-Kagaku, to appear).

4. ファジィ数理計画法の適用

上記の手法はいずれも与えられた教師データに対し完全な分類を行おうとするものである．その結果，程度の大小はあれ，分離曲面が複雑になることは避けられず，汎化能力の劣化を引き起こす．ところで，完全に分類するよりも，分類が困難なデータに対しては判定を保留する方が実際的であることも多い．このように判定保留の部分をグレーゾーンとして残す方法も提案されている．これはファジィ数理計画法を用いることによって容易に実現できる．

たとえば，問題(II')において，制約 $B\boldsymbol{u}-\beta\mathbf{1} \leqq \mathbf{0}$ を

$$B\boldsymbol{u}-\beta\mathbf{1} \lesssim \mathbf{0}$$

のようにファジィ化する．ここで，$\lesssim$は“almost $\leqq$”を意味する．不等式 $B\boldsymbol{u}-\beta\mathbf{1} \leqq \mathbf{0}$ の満足度はメンバーシップ関数 $M_B(\boldsymbol{x})$ によって与えられる．

$$M_B(\boldsymbol{x}) = \begin{cases} 1, & -\frac{\boldsymbol{x}^T\boldsymbol{u}-\beta}{e} \geqq 1.0 \\ -\frac{\boldsymbol{x}^T\boldsymbol{u}-\beta}{e}, & 0 < -\frac{\boldsymbol{x}^T\boldsymbol{u}-\beta}{e} < 1 \\ 0, & -\frac{\boldsymbol{x}^T\boldsymbol{u}-\beta}{e} \leqq 0 \end{cases}$$

$$M_A(\boldsymbol{x}) = \begin{cases} 0, & \frac{\boldsymbol{x}^T\boldsymbol{u}-\alpha}{e} - 1 \geqq 0 \\ \frac{\boldsymbol{x}^T\boldsymbol{u}-\alpha}{e} - 1, & -1.0 < \frac{\boldsymbol{x}^T\boldsymbol{u}-\alpha}{e} - 1 < 0 \\ -1, & \frac{\boldsymbol{x}^T\boldsymbol{u}-\alpha}{e} - 1 \leqq -1 \end{cases}$$

よく知られたファジィ数理計画の手法を用いれば上記のファジィ化は次の LP 問題を解くことに帰着される．

(IV) Minimize $\sum_{j=1}^{k} y_j - w\lambda$
subject to

$$\begin{aligned} -A\boldsymbol{u}+\beta\mathbf{1} &\leqq \boldsymbol{y} \\ B\boldsymbol{u}-\beta\mathbf{1}+e\lambda\mathbf{1} &\leqq e\mathbf{1} \\ -\mathbf{1} \leqq \boldsymbol{u} &\leqq \mathbf{1} \\ \boldsymbol{y} &\geqq \mathbf{0} \end{aligned}$$

(V) Minimize $\sum_{i=1}^{m} z_i - w\lambda$
subject to

$$\begin{aligned} A\boldsymbol{u}-\alpha\mathbf{1}-e\lambda\mathbf{1} &\geqq -e\mathbf{1} \\ B\boldsymbol{u}-\alpha\mathbf{1} &\leqq \boldsymbol{z} \\ -\mathbf{1} \leqq \boldsymbol{u} &\leqq \mathbf{1} \\ \boldsymbol{z} &\geqq \mathbf{0} \end{aligned}$$

【注】 パラメータ e は分離超平面の“グレーゾーン”の幅を規定し，w はメンバーシップ関数を考慮する程度を規定するウェイトである．明らかに，分離曲面のあいまい度はこれらのパラメータの取り方に依存する．これらは問題に応じて経験によって定める必要があるが，その値によっては分離曲面が大きく異なることがあり，注意を要する．．

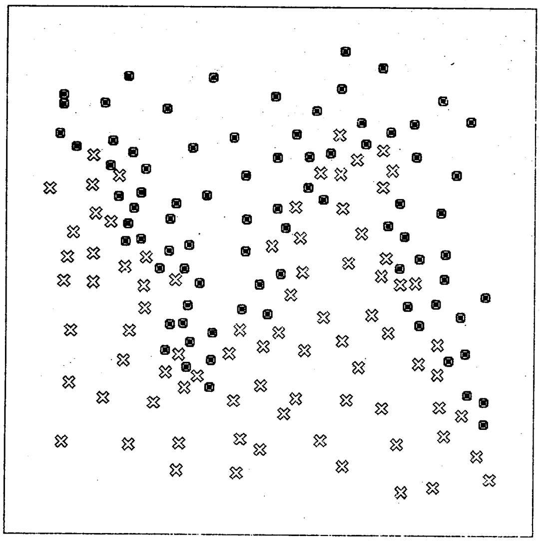

Fig. 3. Data

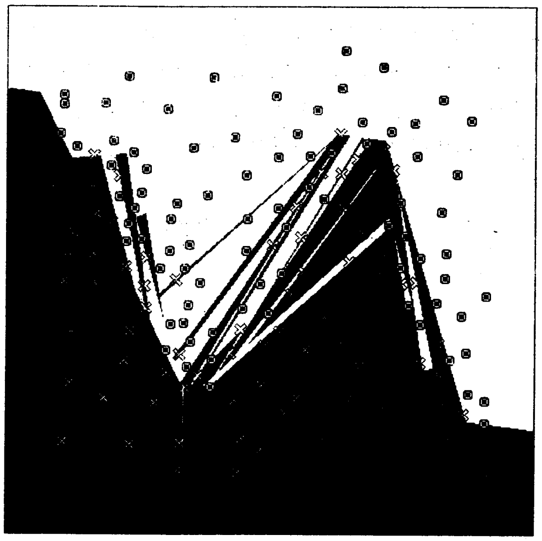

Fig. 4. MSM

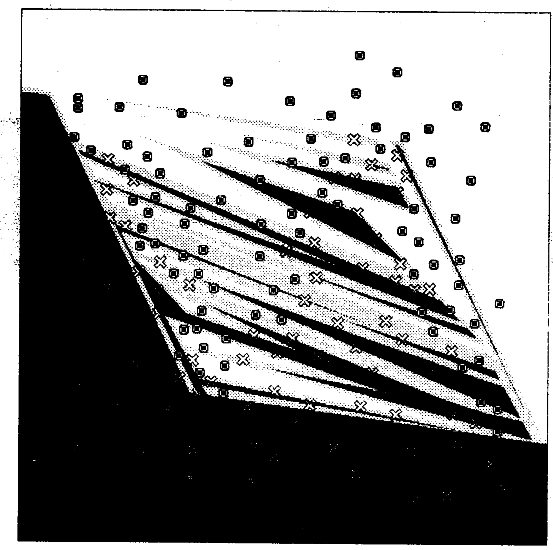

Fig. 5(a). Fuzzy MSM (e=0.01)

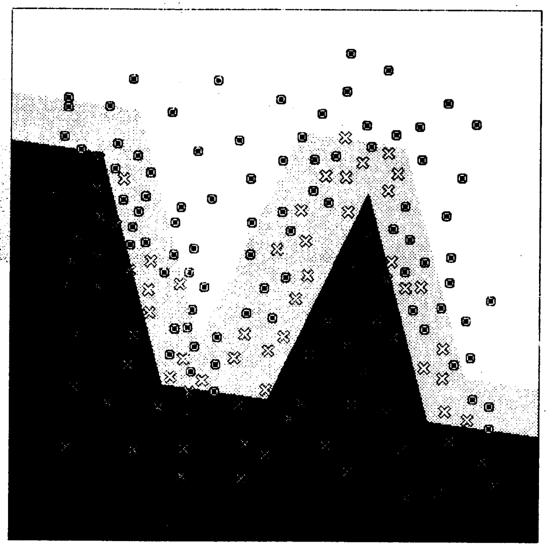

Fig. 5(b). Fuzzy MSM (e=0.05)

5. 結び

上で述べた手法はいずれも計算時間が比較的少なくてすむという利点がある．しかもデータが次々と追加されていくとき，最初から計算をやり直すのではなく，現在までの情報を利用して必要なだけ規則を修正するという追加学習が実際問題では重要であるが，この追加学習も容易に行えるというのが特徴である．株式の売買決定問題への応用例についてはNakayama-Kagaku (to appear)を参考にされたい．

参考文献

[1] K. P. Benett and O. L. Mangasarian, Robust Linear Programming Discrimination of Two Linearly Inseparable Sets, *Optimization Methods and Software*, Vol. 1, pp. 23-34, 1992.

[2] Bishop, C.M. (1995): *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford

[3] Hassoun, M.H. (1995): *Fundamentals of Artificial Neural Networks*, The MIT Press, Cambridg, Massachusetts

[4] Mangasarian, O.L. (1968): Multisurface Method of Pattern Separation, *IEEE Transact. on Information Theory*, **IT-14**, 801-807

[5] Mangasarian, O.L. (1993): Mathematical Programming in Neural Networks, *ORSA J. on Computing*, **5**, 349-360

[6] H. Nakayama and N. Kagaku, (to appear): Pattern Classification by Linear Goal Programming and its Extensions, *J. of Global Optimization*

[7] Nakayama, H. and Nasu, K. (1994): Pattern Classification by Linear Programming, *presented at XI-th Intenational Conf. on Multiple Criteria Decision Making*, Coimbra/Portugal

[8] Takiyama, R. and Fukudome, K. (1992): A Comparison of Serial and Parallel Processings by Neurons, *Proc. International Joint Conf. on Neural Networks, Beijin/China*, **II**, 811-814

[9] Wolbeg, W.H. and Mangasarian, O.L. (1990): Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology, *Proc. Natl. Acad. Sci. USA*, **87**, 9193-9196