

# 多目的最適化に基づくマルチクラスサポートベクトルマシン

巽 啓司

サポートベクトルマシンは、パターン認識のための有力な機械学習モデルの一つである。その基本モデルは2クラス識別器であり、クラス間のマージンの最大化により高い汎化性を保持している。本稿では、そのモデルのマルチクラス識別への拡張法として、各クラス間の幾何マージンを最大化する多目的最適化モデルを紹介する。まず、基本となる線形ハードマージンモデルについて詳解し、その非線形への拡張法や、一対多手法の情報をを用いたモデル、数値実験による識別性能評価の結果について報告する。

キーワード：サポートベクトルマシン、マルチクラス識別、マージン最大化、多目的最適化

## 1. はじめに

サポートベクトルマシン (SVM) は、統計学的学習理論の枠組みで提案された、パターン認識のための有力な機械学習モデルである [17]。機械学習とは、人間の「学習」能力を、人工知能上に実現する手法・技術のことであり、検索エンジン、医療診断、スパムメール検出、DNA 配列の分類、音声認識や文字認識といった幅広い分野で用いられている。なかでも、SVM は、数理計画法をたくみに援用して効率的に「学習」を行うモデルであり、その基本となる2クラス識別では、2個のクラスを識別する超平面と識別データとの距離を表す「マージン」を最大化する最適化問題を求解することで、高い汎化性能を保持している。

さらに、識別問題の多くが3個以上の識別先クラスをもつマルチクラス識別問題であることから、近年、このモデルをマルチクラス識別に拡張する手法が提案されている [1]。その多くは、もとの問題をいくつかの部分問題に分解し、各々の問題を学習した2クラスSVMを組み合わせ、一つの識別関数を構成する方法であり、一対多手法、一対一手法などが代表的な方法である [3, 7]。他方、マルチクラス識別問題をそのまま一つの最適化問題として定式化し、一度の求解により識別器を構成する、一括型手法と呼ばれる方法も提案されている [5]。この手法は、2クラスSVMの基本アイデアである「マージン最大化」を自然な形で拡張したモデルとみなすことができるが、一対多手法と比

べ、得られる識別器の能力に大きな差が見られず、必要とする計算量が相対的に大きいため、上記の2クラスSVMを組合せる手法が主に研究されてきた [1, 6]。

筆者らは、近年、一括型手法において求解に使用されている最適化問題が、SVMの基本アイデアである「マージン最大化」を正確に実現できておらず、その潜在的能力を十分に引き出せていないことを指摘し、マージンを正確に最大化する多目的最適化モデルとして、多目的マルチクラスSVM (MMSVM) を提案した [11]。さらに、ベンチマーク問題への適用によりそのモデルの高い汎化能力を確認している。本稿では、まず、多目的モデルの基本である線形ハードマージンモデルについて詳述し、その非線形への拡張モデル [12] や、一対多手法の情報をを用いて、汎化能力を保持しつつ、計算量を大幅に削減できるモデル [13, 15] の概要を紹介する。

## 2. サポートベクトルマシン

本節では、2クラス識別問題に対するSVMについて紹介する。一般に、2クラス識別問題とは、未知の確率分布  $P(x, y)$  に従う入力データ  $x^i \in \mathcal{R}^n$  とそのクラスラベル  $y_i \in \{1, -1\}$ ,  $i \in I$  が与えられたとき、同じ分布に従う未知の入力のクラスを予測する問題である。これらの与えられたデータは、教師データと呼ばれる。ここでは、簡単のため、教師データは線形分離可能であるとする。

SVMの目的は、未知の入力データ  $x$  を正しく識別する能力、つまり、汎化能力が最大となる識別関数

$$f(x) = w^\top x + b \quad (1)$$

たつみ けいじ  
大阪大学大学院工学研究科 電気電子情報工学専攻  
〒565-0871 大阪府吹田市山田丘 2-1

を構成することである。ただし、 $f(x)$  の値の正負により識別を行い、 $w \in \mathcal{R}^n, b \in \mathcal{R}$  は決定変数である。 $(w, b)$  の値を定める学習は、凸 2 次計画問題

$$(P) \quad \min_{w, b} \frac{1}{2} \|w\|^2$$

$$\text{s.t. } y_i(w^\top x^i + b) \geq 1, i \in I,$$

として与えられる。問題 (P) の制約条件は、すべてのデータを正しく識別することを表している。また、(P) の最適解  $(w^*, b^*)$  に対し、 $\|w^*\|$  の逆数は、識別超平面  $(w^*)^\top x + b^* = 0$  と入力データ集合との距離である、マージン：

$$\min_{i \in I} \frac{|w^\top x^i + b|}{\|w\|}, \quad (2)$$

と一致することが知られており、(P) の目的関数の  $\|w\|$  の最小化は、マージンの最大化を意味する。マージン最大化は、識別器 (1) の汎化能力の向上につながる。ことが統計学的に示されており、SVM の汎化性能の理論的根拠となっている [17]。線形分離可能なデータに適用可能なモデル (P) は、ハードマージンモデルと呼ばれる (図 1)。

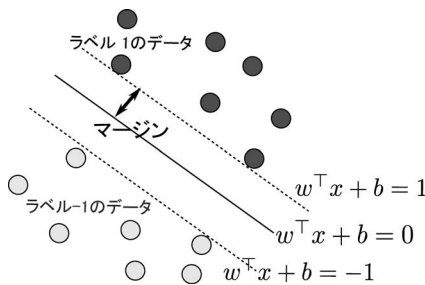


図 1 2 クラス SVM

問題 (P) の双対問題は、双対変数  $\alpha \in \mathcal{R}^{|I|}$  についての凸 2 次計画問題

$$(D) \quad \max_{\alpha} -\frac{1}{2} \sum_{i \in I} \sum_{j \in I} y_i y_j \alpha_i \alpha_j x^i{}^\top x^j + \sum_{i \in I} \alpha_i$$

$$\text{s.t. } \sum_{i \in I} y_i \alpha_i = 0, \alpha_i \geq 0, i \in I,$$

であり、その最適解  $\alpha^*$  により、識別関数を構成することも可能である。さらに、データを識別するのに適した特徴空間  $\mathcal{F}$  への非線形写像  $\phi: \mathcal{R}^n \rightarrow \mathcal{F}$  を用いて、 $\phi(x)$  に関する線形識別関数  $f(x) = w^\top \phi(x) + b$  を構成する非線形モデルもよく用いられる。このとき、 $\phi(x)$  の内積が、関数  $k(\cdot, \cdot) = \phi(\cdot)^\top \phi(\cdot)$  により与えられる場合、双対問題 (D) は、直接  $\phi(x)$  の値を求めなく求解可能であり、 $k$  として  $\phi(x)$  の存在が

保障された Mercer カーネルが用いられる。この方法は、カーネル法と呼ばれる [9]。この非線形 SVM についても、線形モデルと同様な汎化性の議論が成り立つ。

さらに、教師データの一部に例外データが含まれている場合は、(P) での各データを正しく識別する制約を緩和する変数を導入し、目的関数に緩和量に関するペナルティ項を加えたモデルが用いられ、ソフトマージンモデルと呼ばれる。本稿で扱う各モデルには、ソフトマージンモデルも存在しているが、紙面の都合上、ハードマージンモデルを中心に議論を行う。

### 3. マルチクラス識別問題

この 2 クラス SVM を、3 個以上のクラスをもつマルチクラス識別問題に拡張することを考える。マルチクラス識別問題とは、2 クラス識別問題と同様に、教師データとして、入力データ  $x^i \in \mathcal{R}^n$  とクラスラベル  $y_i \in M := \{1, \dots, m\}, i \in I := \{1, \dots, l\}$  がそれぞれ与えられたとき、教師および未知データに対して正しく識別を行う関数を構成することを意味する。2 クラス SVM の拡張モデルはさまざまに研究されているが、その多くは、もとのマルチクラス識別問題から、2 クラス識別部分問題を抽出し、そのおのおのを学習して得られた 2 クラス SVM を組み合わせて識別を行う方法である。その代表的なものに、 $m$  クラス内のすべてのクラスのペアについて、 $m(m-1)/2$  個の SVM を構成する一対一手法 [7] と、1 個のクラスと残りのすべてのクラス間を識別する  $m$  個の SVM を構成する一対多手法 [3] などがある。一対一手法を組み合わせる方法としては、2 クラス SVM をデータに適用する順序をグラフを用いて決定する方法 [10] がよく用いられ、一対多手法については、クラス  $p$  を識別する 2 クラス SVM の学習時に得られた  $w^p{}^\top x + b^p$  を、データ  $x$  がクラス  $p$  として識別された場合の確信度を表すと解釈し、識別関数

$$f(x) = \operatorname{argmax}_{p \in M} w^p{}^\top x + b^p \quad (3)$$

を構成することで、より確信度が大きいクラスに識別する方法が用いられる。他方、識別関数 (3) の変数  $(w, b)$  を決定するために、マージンを最大化する一つの最適化問題を解くことで、一度に変数を決定する一括型手法と呼ばれるモデルも知られている [5]。

これらの方法を比較すると、グラフを用いた一対一手法と識別関数 (3) を用いる一対多手法では適用する問題により、それぞれ識別性能に差がある一方で、同

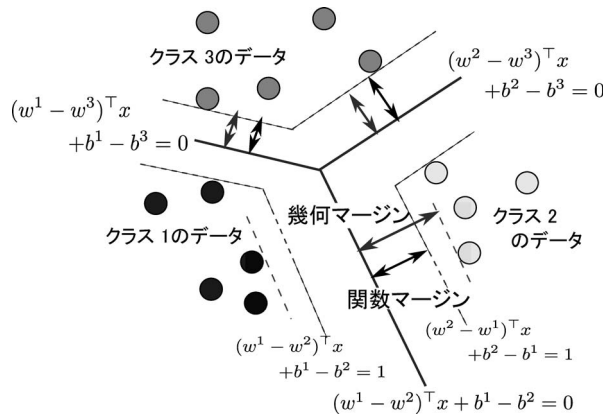


図2 一括型従来モデル (O)

じタイプの識別関数 (3) を構成する一対多手法と一括型手法では比較的よく似た汎化性能をもつ識別器が求まることが知られている。加えて、一括型手法は、全データを用いて全変数に関する最適化問題を解くため、三つの手法の中で最も多くの計算量を要することから、主に一対一手法、一対多手法に関する研究が行われてきた [1, 6].

筆者らは、一括型手法のモデルにおいて、汎化性を高めるためのマージン最大化が正確に図られていない点を指摘し、多目的最適化の視点から、マージンを正確に最大化することにより、汎化性能を向上させたモデルを提案した [11, 12]. また、一対多手法に基づいて、提案した多目的モデルで用いる最適化問題の解集合を限定することで、求解時間を大幅に短縮しつつも高い汎化性能をもつ識別器が得られることを示した [13, 15]. これらのモデルについて以下の節で紹介する。

## 4. 一括型多目的マルチクラスモデル

### 4.1 一括型手法

前節で述べたとおり、一括型手法では、データ  $x$  がクラス  $p$  に識別された場合の確信度を表す関数を用いて、識別関数を (3) 式のように構成する。このとき、

$$(w^p - w^q)^T x + (b^p - b^q) = 0, \quad q \neq p, \quad p, q \in M, \quad (4)$$

は、クラス  $p$  と  $q$  の境界を定める識別超平面である。 $(w, b)$  の決定には、2 クラス SVM と同様に、すべてのペア  $(p, q), q \neq p \in P$  に対して  $1/\|w^p - w^q\|$  を最大化するモデルが提案されている。

$$(O) \quad \begin{aligned} \min_{w, b} \quad & \frac{1}{2} \sum_{p \in M} \sum_{q \in M} \|w^p - w^q\|^2 \\ \text{s.t.} \quad & (w^p - w^q)^T x^i + (b^p - b^q) \geq 1, \\ & i \in I_p, \quad q \neq p, \quad p, q \in M. \end{aligned}$$

ここで  $I_p$  は、教師データのラベルがクラス  $p$  となるデータの添字集合  $I_p := \{i \in I \mid y^i = p\}$  を表す。同じ視点から、複数の一括型モデルが提案されているが、そのすべてが (O) と実質的に同一のモデルであることが知られている [5]. しかし、モデル (O) では、2 クラス SVM と異なり、マージン最大化が正確に行われていない。(O) の最適解  $(w^*, b^*)$  において  $1/\|w^{p^*} - w^{q^*}\|$  の値 (関数マージン) は、クラス  $p, q$  間の識別超平面 (4) と二つの支持超平面  $(w^p - w^q)^T x + (b^p - b^q) = \pm 1$  間の距離を表すが、その支持超平面内に必ずしも教師データがあるとは限らない (図 2)。つまり、関数マージンは、各クラス間の識別超平面とそれに最も近いデータへの距離である、幾何マージン：

$$d_{pq}(w, b) = \min_{i \in I_p \cup I_q} \frac{|(w^p - w^q)^T x^i + (b^p - b^q)|}{\|w^p - w^q\|}, \quad q > p, \quad p, q \in M, \quad (5)$$

と必ずしも一致しないことが示されている [11].

### 4.2 多目的モデル

一括型従来モデル (O) では、幾何マージン最大化の意味での汎化性向上が必ずしも望めないため、筆者らは、複数の幾何マージンを多目的最適化の意味で最大化する多目的マルチクラス SVM (multiobjective multiclass SVM: MMSVM) を提案した [11].

$$(M1) \quad \begin{aligned} & \max_{w,b} \quad d_{12}(w,b), d_{13}(w,b), \dots, d_{(m-1)m}(w,b) \\ & \text{s.t.} \quad (w^p - w^q)^\top x^i + (b^p - b^q) \geq 1, \\ & \quad \quad \quad i \in I_p, q \neq p, p, q \in M. \end{aligned}$$

このモデルでは、すべてのデータを正しく識別する制約のもとで、幾何マージン (5) が最大化されており、さらに、マルチクラス識別においては複数のクラスペアが存在するため、(M1) は、おのおのの幾何マージンを目的関数とする多目的最適化問題として定式化されている。また、(M1) を解くために新たな変数ベクトル  $\sigma \in \mathcal{R}^{m(m-1)/2}$  を導入した多目的最適化問題

$$(M2) \quad \begin{aligned} & \max_{w,b,\sigma} \quad \frac{\sigma_{12}}{\|w^1 - w^2\|}, \dots, \frac{\sigma_{(m-1)m}}{\|w^{(m-1)} - w^m\|} \\ & \text{s.t.} \quad (w^p - w^q)^\top x^i + (b^p - b^q) \geq \sigma_{pq}, \\ & \quad \quad \quad i \in I_p, q > p, p, q \in M, \\ & \quad \quad \quad (w^q - w^p)^\top x^i + (b^q - b^p) \geq \sigma_{pq}, \\ & \quad \quad \quad i \in I_q, q > p, p, q \in M, \\ & \quad \quad \quad \sigma_{pq} \geq 1, q > p, p, q \in M, \end{aligned}$$

が提案されている。このモデルに関しては、変数  $\sigma$  の導入により、二つの支持超平面  $(w^p - w^q)^\top x - (b^p - b^q) = \pm \sigma_{pq}$  を各ペア  $(p, q)$  ごとに自由に平行移動でき、パレート最適解において、 $\sigma_{pq}/\|w^p - w^q\|$  が幾何マージン  $d_{pq}(w, b)$  に一致することや、(M2) の求解により (M1) のパレート最適解が求解可能であることが示されている [11]。次に (M2) のパレート最適解の求解方法について考察する。

### 4.3 多目的モデルの求解法

前述の多目的モデル (M2) の求解法として、多目的最適化問題に対するスカラー化手法を用いて導出した方法が提案されている [11]。ここでは、 $\varepsilon$  制約法に基づいた方法を紹介する。まず、(M2) に目的関数の中で、1 個のクラスの組  $(r, s)$  についてはそのまま目的関数とし、残りの目的関数は、適当な定数  $\varepsilon_{pq}, q > p, (p, q) \neq (r, s), p, q \in M$  を用いて、制約に変更した問題を考える。この問題の最適解は、もとの多目的最適化問題の弱パレート解となることが知られている [4]。さらに、求解のため、 $\sigma_{rs}$  を正の定数  $c_{rs}$  で固定すると以下のような 1 目的最適化問題が得られる<sup>1</sup>。

$$(E-P) \quad \begin{aligned} & \max_{w,b,\sigma} \quad \frac{c_{rs}}{\|w^r - w^s\|} \\ & \text{s.t.} \quad \frac{\sigma_{pq}}{\|w^p - w^q\|} \geq \varepsilon_{pq}, \\ & \quad \quad \quad q > p, (p, q) \neq (r, s), p, q \in M, \\ & \quad \quad \quad (w^p - w^q)^\top x^i + (b^p - b^q) \geq \sigma_{pq}, \\ & \quad \quad \quad i \in I_p, q > p, p, q \in M, \\ & \quad \quad \quad (w^q - w^p)^\top x^i + (b^q - b^p) \geq \sigma_{pq}, \\ & \quad \quad \quad i \in I_q, q > p, p, q \in M, \\ & \quad \quad \quad \sigma_{pq} \geq 1, q > p, p, q \in M, \\ & \quad \quad \quad \sigma_{rs} = c_{rs}. \end{aligned}$$

定数  $c_{rs}$  を十分に大きな正数を選んだ場合、(E-P) については以下のような定理が得られる。

**定理 1.** (M2) の任意のパレート最適解  $(w^*, b^*, \sigma^*)$  に対して、 $\frac{c_{rs}}{\sigma_{rs}^*}(w^*, b^*, \sigma_{-rs}^*)$  が、(E-P) の最適解となるような  $\varepsilon_{-rs}$  が存在する。

**定理 2.** (E-P) の最適解  $(w^*, b^*, \sigma_{-rs}^*)$  に対し、 $(w^*, b^*, (\sigma_{-rs}^*, c_{rs}))$  は (M2) の弱パレート最適解である。

ここで、 $\sigma_{-rs}, \varepsilon_{-rs}$  は、おのおの、 $\sigma, \varepsilon \in \mathcal{R}^{m(m-1)/2}$  から要素  $\sigma_{rs}, \varepsilon_{rs}$  を取り除いたベクトルを、 $(w, b, (\sigma_{-rs}, c_{rs}))$  は、 $(w, b, \sigma_{-rs})$  に、要素  $\sigma_{rs}$  として  $c_{rs}$  を追加したベクトルを表す。

定理 1, 2 より、(E-P) の最適解は (M2) のパレート最適解であり、適切なペア  $(r, s)$  と定数  $c_{rs}, \varepsilon_{-rs}$  を選択することで、(E-P) の求解により、(M2) の任意のパレート最適解が得られることがわかる。そのため、パラメータの選択によりさまざまなパレート解を求めることも可能である。また、(E-P) の求解方法としては、この問題が 2 次錐計画問題に変形できることから、主双対内点法などの有効な求解法を適用可能である [2]。

さらに、(M2) に制約を緩和する変数を導入し、そのペナルティ関数を、目的関数として加えたソフトマージンモデルも提案されている [14]。線形ハード・ソフトマージン型の一括型従来モデルと MMSVM モデルをベンチマーク識別問題に適用し、得られた解を幾何マージンに関して比較すると、前者が後者の劣解となっていることが確認されている。また、MMSVM モデルのパレート最適解集合を求め、その汎化性を検証したところ、従来モデルと比べ高い識別率をもつさまざまな解が多数含まれていることや、幾何マージンの和が最大となる解の識別率が必ずしも高くないこと、すべてのパレート最適解が高い汎化性能をもつわけではないことなどがわかっている [11, 14]。

<sup>1</sup> 紙面の都合上、問題 (E-P) の決定変数を、 $(w, b, \sigma)$  とし記載したが、実質的な決定変数は  $(w, b, \sigma_{-rs})$  である。

#### 4.4 非線形モデル

多目的モデル (M2) を、2 クラス SVM と同様に、カーネル法を用いて非線形モデルに拡張することも可能である [12]. この場合、問題に適した適当な写像  $\phi: \mathcal{R}^n \rightarrow \mathcal{F}$  を用いて、非線形識別器

$$f(x) = \arg \max_{p \in M} \{w^p \top \phi(x) + b^p\}. \quad (6)$$

を構成することが目的となる. そのため、(M2) の  $x$  を  $\phi(x)$  で置き換えたモデルを考え、 $\mathcal{F}$  における幾何マージンを最大化することにより  $(w, b)$  を求めることを考える. ただし、カーネル法を適用可能にするため、新しい決定変数  $\beta_i^p, i \in I, p \in M$  を用いて、

$$w^p = \sum_{i \in I} \beta_i^p \phi(x^i), \quad p \in M, \quad (7)$$

として、 $w$  を置き換える. このとき、 $\phi(x)$  はすべて内積の形でのみ現れるため、モデルは以下のように与えられる.

$$\begin{aligned} \max_{\beta, b, \sigma} \quad & \frac{\sigma_{12}}{\|\beta^1 - \beta^2\|_K}, \dots, \frac{\sigma_{(m-1)m}}{\|\beta^{m-1} - \beta^m\|_K} \\ \text{s.t.} \quad & (\beta^p - \beta^q)^\top \kappa(x^i) + (b^p - b^q) \geq \sigma_{pq}, \\ & i \in I_p, q > p, p, q \in M, \\ \text{(NM)} \quad & (\beta^q - \beta^p)^\top \kappa(x^i) + (b^q - b^p) \geq \sigma_{pq}, \\ & i \in I_q, q > p, p, q \in M, \\ & \sigma_{pq} \geq 1, q > p, p, q \in M. \end{aligned}$$

ここで、

$$\begin{aligned} \kappa(x^i) &:= (\kappa(x^1, x^i), \dots, \kappa(x^l, x^i))^\top, i = 1, \dots, l, \\ K &:= (\kappa(x^1), \dots, \kappa(x^l)), \|u\|_K := u^\top K u, \end{aligned}$$

であり、Mercer カーネルの性質より  $K$  は半正定値である [9]. また、 $w$  を (7) 式により限定しても、もとの問題の最適解が得られることが知られている [8]. さらに、(NM) の求解のために、行列  $K$  を対角化し、固有値が正である固有空間の和に実行可能解を限定することで、線形ハードマージンモデルと同様に、(NM) のパレート最適解を最適解にもつ 1 目的 2 次錐計画問題が導出でき、その 2 次錐計画問題の求解により、線形モデルと同様に、(NM) のパレート最適解を求めることが可能である [12].

#### 4.5 数値実験結果

本節では、一括型の従来モデル ((O) の双対問題にカーネル法を適用したモデル: DNO), と非線形多目的マルチクラス SVM (NM) をベンチマーク問題 [16] に適用した結果を示す [12]. 10 分割交差検証法 (10-fold cross validation) を用いて、二つのモデルを比較

した. この検証法は、与えられた入出力データを 10 等分し、そのうちの 9 個を教師信号として学習を行い、残りの 1 個で識別率を求める実験を 10 通り行い、正しく識別する割合の平均を求める方法である. 最適化ツールとして MathWorks Matlab 7.0.1<sup>2</sup> および Mosek version 5.0<sup>3</sup> を用い、カーネル関数としては、多項式カーネル  $k(u, v) = (u^\top v + 1)^d$ , RBF カーネル  $k(u, v) = \exp(-\gamma \|u - v\|^2)$  を用いた<sup>4</sup>. また、(NM) のパレート最適解を得るため 1 目的 2 次錐計画問題を求解し、 $c_{rs} = 10$ ,  $\varepsilon_{-rs}$  の値として、(DNO) により得られた識別超平面に対する幾何マージンを用いた. 固定するクラスのペア  $(r, s)$  としては、各ベンチマーク問題に対して、すべてのクラスペアを選択しておのおの 2 次錐計画問題を求解し、その中で最も汎化性能が高いものを表 1 に示した. 表中の太字は、おのおの問題に対する最良の識別率を、“-” は対応するモデルで求解できなかったことを表す. 表 1 より、各問題ごとに最良の識別値が得られたモデルはすべて (NM) であることがわかる.

表 1 各ベンチマークデータに対する汎化性能の比較 (%)

ベンチマーク 問題	多項式		RBF	
	DNO	NM	DNO	NM
Wine	99.94	97.75	98.88	<b>100.00</b>
Dermatology	95.90	96.72	95.90	<b>96.99</b>
Balance Scale	<b>100.00</b>	<b>100.00</b>	97.44	98.24
Iris	93.33	94.00	<b>95.33</b>	<b>95.33</b>
Vehicle	-	-	74.23	<b>76.24</b>
Car	97.86	<b>98.15</b>	97.57	97.45
Glass	-	-	<b>65.89</b>	<b>65.89</b>

また、表 2 は、二つの手法を Vehicle に適用した際に得られた幾何マージンの 1 例であり、(DNO) で得られた解は、幾何マージンに関して、(NM) により得られた解の劣解になっていることがわかる. 他の多くのケースにおいても、(DNO) では幾何マージンが正確に最大化できていないことを確認している.

### 5. 一対多手法に基づく多目的モデル

最後に、一括型と同じ識別関数を構成する一対多手法に基づいた多目的モデル [13, 15] を紹介する. 一対多手法において、 $m$  個の 2 クラス識別問題を解いて得られた解を  $(\bar{w}^p, \bar{b}^p), p \in M$  とする. この  $\bar{w}^p$  と新た

<sup>2</sup><http://www.mathworks.co.jp/>

<sup>3</sup><http://www.mosek.com/>

<sup>4</sup> 関数中のパラメータ  $d, \gamma$  は実験により適当なものを選択した.

表2 2つのモデルを Vehicle に適用した際に得られた解の幾何マージンの比較

モデル	$d_{12}$	$d_{13}$	$d_{14}$	$d_{23}$	$d_{24}$	$d_{34}$
DNO	0.04718	0.01500	0.01765	0.01508	0.01875	0.01206
$NM_{(r,s)=(1,2)}$	0.07045	0.02510	0.02409	0.01814	0.02225	0.01241
$NM_{(r,s)=(1,3)}$	0.04811	0.05593	0.01792	0.01755	0.02036	0.01223
$NM_{(r,s)=(1,4)}$	0.04864	0.01578	0.04891	0.01722	0.02432	0.01230
$NM_{(r,s)=(2,3)}$	0.04860	0.02248	0.01858	0.05139	0.01931	0.01228
$NM_{(r,s)=(2,4)}$	0.04904	0.01741	0.02139	0.01617	0.05437	0.01234
$NM_{(r,s)=(3,4)}$	0.05144	0.02405	0.02282	0.02046	0.02565	0.01373

な決定変数  $\alpha_p, p \in M$  を用いて, (M2) での変数  $w$  を,

$$w^p = \alpha_p \bar{w}^p, p \in M, \quad (8)$$

と限定することを考える. つまり, 識別関数として,

$$f(x) = \operatorname{argmax}_{p \in M} \{\alpha_p \bar{w}^{p\top} x + b^p\}, \quad (9)$$

を考える. このとき, (M2) は,

$$\begin{aligned} & \text{(M-OA)} \\ & \max_{\alpha, b, \sigma} \frac{\sigma_{12}}{\|\alpha_1 \bar{w}^1 - \alpha_2 \bar{w}^2\|}, \dots, \frac{\sigma_{(m-1)m}}{\|\alpha_{m-1} \bar{w}^{m-1} - \alpha_m \bar{w}^m\|} \\ & \text{s.t. } (\alpha_p \bar{w}^p - \alpha_q \bar{w}^q)^\top x^i + (b^p - b^q) \geq \sigma_{pq}, \\ & \quad i \in I_p, q > p, p, q \in M, \\ & (\alpha_q \bar{w}^q - \alpha_p \bar{w}^p)^\top x^i + (b^q - b^p) \geq \sigma_{pq}, \\ & \quad i \in I_q, q > p, p, q \in M, \\ & \sigma_{pq} \geq 1, q > p, p, q \in M. \end{aligned}$$

と書き直すことができる. このモデルは, 実行可能解集合を限定した MMSVM であるが,  $\bar{w}^p$  の組合せ方に  $\alpha_p$  により自由度を与え, その候補の中で幾何マージンが最大となるものを求める, 一対多手法の改良手法とみなすこともできる. このモデルに対しても, MMSVM と同様に,  $\varepsilon$  制約法に基づき, (M-OA) のパレート最適解を最適解にもつ 1 目的 2 次錐最適化問題の導出が可能であり, 容易に求解可能である [13].

さらに, (M-OA) にカーネル法を適用して非線形に拡張することも可能であり [15], (NM) では  $l \times l$  行列  $K$  の対角化が必要となるのに対し, 非線形 (M-OA) では,  $m(m-1)/2$  個の  $2 \times 2$  行列の対角化のみが必要となる. 加えて, (NM) と非線形 (M-OA) の変数の数は, おおの,  $m(m+2l+1)$ ,  $m(m-1)/2$  であり, 一般に, クラス数  $m \ll$  データ数  $l$  であるため, 後者のモデルでは, 変数の数が大幅に削減されていることが分かる. これらの事実より, (M-OA) の求解のための 2 次錐計画問題の求解に要する計算量は大幅に削減でき, 実際に, ベンチマーク問題の Vehicle (4 クラス, 18 次元, データ 846 個) に非線形モデルを適用し

た場合, 平均して, 非線形型 ( $\varepsilon$ -P) の 0.5%, 一対多手法の 26.1% 程度の計算時間で, (M-OA) は求解可能である. また, その汎化性能は, 多くの識別問題において一対多手法よりも高く, MMSVM に対し,  $\varepsilon$  制約法で得られる  $m(m-1)/2$  個のパレート最適解集合間で比較した場合も, (M-OA) が高い汎化性能をもつ解が得られていることが確認されている [15]. これらの結果より, (8) による実行可能解の限定は, 有効な方法であるといえる.

## 6. さいごに

本稿では, マルチクラス識別問題に対する一括型手法として, 多目的マルチクラス SVM を紹介し, 一括型手法の基本アイデアである「マージン最大化」を正確に行えば, 識別器の汎化性能の向上が可能であることを示した. また, 従来法により得られる情報を用いて 2 次錐計画問題を解くことで, 多目的モデルの汎化性の高いパレート最適解が得られることを紹介した.

しかし, この多目的モデルでは, 目的関数の幾何マージンは間接的に汎化性を表す指標でしかなく, 得られた解をその目的関数値のみで評価することは困難である. 加えて, 大規模なマルチクラス識別問題の求解時には, 多数のパレート最適解を求めることも困難となる. そのため, 現在, 筆者らは, 効率的に汎化性の高い解を求めるモデル・求解方法の研究に取り組んでいる.

## 参考文献

- [1] S. Abe, *Support Vector Machines for Pattern Classification*, Springer, 2005.
- [2] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Mathematical Programming*, Ser. B, 95, pp. 3–51, 2003.
- [3] L. Bottou, C. Cortes, J. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. Muller, E. Sackinger, P. Simard and V. Vapnik, "Comparison of classifier methods: A case study in handwriting digit recognition," *Proc. Int. Conf. Pattern Recognition*, pp. 77–87, 1994.
- [4] M. Ehrgott, *Multicriteria Optimization*, 2nd edi-

tion, Springer, Berlin, 2005.

- [5] Y. Guermeur, “Combining discriminant models with new multiclass SVMs,” Neuro COLT2 Technical Report Series, NC2-TR-2000-086, 2000.
- [6] C. W. Hsh and C. J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Transactions on Neural Networks*, Vol. 13, No. 2, pp. 181–201, 2002.
- [7] U. Kressel, “Pairwise classification and support vector machines,” in *Advances in Kernel Methods – Support Vector Learning*, (eds.) B. Schölkopf, C. Burges, and A. J. Smola, MIT Press, Cambridge, pp. 255–268, 1999.
- [8] B. Schölkopf, R. Herbrich and A.J. Smola, “A generalized representer theorem,” *Lecture Notes in Computer Science, 2001*, Vol. 2111/2001, pp. 416-426, 2001.
- [9] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda and B. Shölkopf, “An introduction to kernel-based learning algorithms,” *IEEE Transactions on Neural Networks*, Vol. 12, No. 2, pp. 181–201, 2001.
- [10] J. C. Platt, N. Cristianini and J. Shawe-Taylor, “Large margin DAG’s for multiclass classification,” *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT Press, Vol. 12, pp. 547–553, 2000.
- [11] K. Tatsumi, R. Kawachi, K. Hayashida and T. Tanino, “Multiobjective multiclass support vector machines maximizing geometric margins,” *Pacific Journal of Optimization*, Vol. 6, No. 1, pp.115–141, 2010.
- [12] K. Tatsumi, R. Kawachi and T. Tanino, “Nonlinear extension of multiobjective multiclass support vector machine,” *Proc. IEEE-SMC 2010*, pp. 1338–1343, 2010.
- [13] K. Tatsumi, M. Tai and T. Tanino, “Multiobjective multiclass support vector machine based on the one-against-all method,” *Proc. WCCI 2010*, pp. 2844–2850, 2010.
- [14] K. Tatsumi, M. Akao, R. Kawachi and T. Tanino “Performance evaluation of multiobjective multiclass support vector machines maximizing geometric margins,” *Numerical Algebra, Control and Optimization*, Vol. 1, No. 1, pp. 151–169, 2011.
- [15] K. Tatsumi, M. Tai and T. Tanino, “Nonlinear extension of multi-objective multiclass support vector machine maximizing margins based on one-against-all method,” *Proc. IJCNN 2011*, pp. 1571–1576, 2011.
- [16] UCI benchmark repository of artificial and real data sets, University of California Irvine, <http://www.ics.uci.edu/~mlern/databases/>
- [17] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd edition, NY, Springer, 2000.