

標本部分空間法に基づく サポートベクトルマシンの開発

北村 拓也^{†1} 竹内 翔吾^{†1} 阿部 重夫^{†1}

本論文では、1) 部分空間法に基づくサポートベクトルマシン (SS-SVM) による特徴選択の評価と、2) 標本部分空間法に基づくサポートベクトルマシン (ES-SVM) の提案を述べる。1) の手法では、各クラスの共分散行列の固有値問題によって導いたすべての固有ベクトルに SS-SVM を用いて識別の観点から最適な重みを決定し、識別における重要度を意味する重みを基準として固有ベクトルの選択を繰り返しより良い識別を行うための部分空間の軸の選択を行う。2) の手法では、SS-SVM に各クラスの教師ベクトルの張る凸錐をそれぞれの部分空間とする錐制約部分空間法 (CRS) の識別関数を SS-SVM の識別関数に類似した形にした標本部分空間法 (ES) を適用することで固有値問題を避け、学習コストを削減する。ベンチマークデータを用いて、従来手法と比較することで提案手法の有効性を示す。

Empirical Subspace Based Support Vector Machines

TAKUYA KITAMURA,^{†1} SHOGO TAKEUCHI^{†1}
and SHIGEO ABE^{†1}

In this paper, we propose two methods: 1) the feature selection using subspace based support vector machines (SS-SVM), and 2) empirical subspace based support vector machines (ES-SVM). In 1), first using SS-SVM, we obtain the optimum weights associated with all the eigenvectors of the covariance matrix. Then, we iteratively select the eigenvectors according to the associated weights, which mean importance for classification, as the dictionaries of the class subspace for classification. In 2), similar to cone restricted subspace method (CRS), we define the subspace of a class using the mapping function to the empirical subspace associated with the class. Thus, we do not need to solve the eigenvalue problems, and the training cost is reduced. We demonstrate the effectiveness of the proposed methods over the conventional methods for benchmark datasets.

1. はじめに

パターン認識において、代表的な識別手法の 1 つとして部分空間法^{1),2)} が広く適用されてきた。部分空間法の多くは、未知の入力を各クラスごとに生成した低次元の部分空間との角度またはその部分空間における射影長を類似度としてクラスごとの類似度を基に識別を行う。代表的な部分空間法である CLAFIC 法¹⁾ では単純に射影長が最大となる部分空間に識別を行う。しかしながら、部分空間の各軸は識別の観点から重要性が異なる。たとえば、クラス部分空間のある軸がそのクラスの分布を表すことにあまり用いられなかった場合でも他の軸と同等に扱うため、その軸における類似度が大きい他のクラスに属する入力を誤識別するおそれがある。そこで、各軸に対するそのクラスの分散を各軸に重みとして付ける手法^{1),2)} が提案されている。しかしながら、重みの大きい軸において他のクラスに属する入力に対する類似度が同様に大きくなる場合にこの軸によって誤認識するおそれがあり、識別の観点から最適な重みとはいえない。このように、部分空間法において各クラス部分空間の各軸の重みは識別に大きく影響を与え、識別の観点から最適な重み付けを行うことは非常に重要である。

そこでサポートベクトルマシン (SVM)³⁾ を用いて各クラス部分空間のマージンを大きくし、誤認識を小さくするように識別の観点から最適な重み付けを行う部分空間法に基づくサポートベクトルマシン (SS-SVM)⁴⁾ が提案されている。SS-SVM では非線形問題に対応するためにカーネル法を用いたカーネル主成分分析 (KPCA)^{5),6)} を用いて各クラス部分空間を生成し、入力ベクトルにおいて重み付けされた全クラス類似度の中で対応するクラス類似度が最大となるように制約条件をおき、マージン最大化と誤認識最小化とを考慮した最適化を行う。このため、すべての重みを同時に決定する必要があり、最適化問題には同時定式化方式を用いている。現在まで提案されている SS-SVM には最小自乗 SVM (LS-SVM)³⁾ と線形計画 SVM (LP-SVM)³⁾ が用いられており、それぞれの手法を部分空間法に基づく LS-SVM (SSLS-SVM)、部分空間法に基づく LP-SVM (SSLP-SVM) とよぶ。また、SS-SVM の高速化手法が文献 7) で提案されている。SSLP-SVM では学習時に特徴選択を同時に行うが、SSLS-SVM では従来の部分空間法と同様に各軸におけるクラスの分布の分散を基に特徴選択を行うため、他のクラスの部分空間の重なりなどを考慮してお

^{†1} 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University

らず、識別の観点から必ずしも良い軸の選択とはいえない。また SSLP-SVM では重みに非負制約を設けているが、重みが負となる軸が識別において必要な可能性がある。また、これらの手法では各クラスごとの共分散行列の固有値問題を解く必要があるため、大規模データにおいて部分空間の生成における学習コストが膨大となる。

一方、固有値分解を解く必要のない部分空間法として、錐制約部分空間法 (CRS)⁸⁾ が提案されている。CRS では、特徴ベクトルの要素が非負値をとることを前提として、特徴ベクトルはクラスごとに一部の領域に密集して存在することが多いため、教師ベクトルそのものを凸錐のクラス部分空間とおき、入力ベクトルと各クラス凸錐部分空間との角度を類似度として識別を行う。このため、固有値問題を解く必要がなくなり、部分空間生成における学習コストが大幅に削減できる。しかしながら、入力ごとに凸錐との角度を最小自乗法を用いて計算するため識別コストが膨大となりうる。さらに通常のパターン認識問題において線形分離可能であることは稀である。そのため、属するクラスから逸脱した教師ベクトルが存在したとき、この部分空間の張る凸錐は広くなり、クラス部分空間どうして重なりができる。その重複した領域は未分類領域となる。

そこで本論文では、従来の SS-SVM と CRS のこれらの問題を解消するための、1) 最適な重みを基にした特徴選択⁷⁾を行う SSLS-SVM の評価、2) 標本部分空間法に基づくサポートベクトルマシン (ES-SVM) の提案を述べる。1) の手法では、KPCA により各クラスごとの共分散行列の固有ベクトルを求め、すべての固有ベクトルに対し SSLS-SVM により重みを決定し、それらの重みを基に以下に述べる条件を満たすまで学習を繰り返す。条件として、2 つの基準を設ける。1 つ目は SSLP-SVM と同様に非負制約を設ける。負の重みとなる軸において属するクラスのパターンより他のクラスパターンにおける類似度が大きくなりやすいことを意味しており、これらの軸を削減することで属するクラスのパターンのみの類似度が大きくなる軸のみを選択することができる。しかしながら、前述のように負の重みを持つ軸が識別において重要である可能性があり、汎化能力の低下につながりうる。そこで、2 つ目の基準として各軸の重みの絶対値の大小を用いる。重みの絶対値が小さい軸は識別に影響を与えにくく、識別において冗長な軸として削除する。実験によりこれら 2 つの条件を設けた特徴選択の比較を行い、さらに SSLP-SVM においても非負制約を設けずに重みを最適化する手法を用いて従来の SSLP-SVM と比較することにより SS-SVM において負の重みを持つ軸が識別において重要となりうるかを確認する。2) の手法では、CRS と同様に各クラスの教師ベクトルを各クラス部分空間の軸とし、CRS の識別関数を入力ベクトルの重み付けをした各軸への射影長の総和とすることにより SS-SVM の決定関数と同じ形にする

ことにより、他のクラス部分空間への重み付けされた射影長と等しくなる領域以外では未分類領域を存在しないようにする。ここで、識別関数が錐部分空間との角度を表していないことからこの部分空間法を以下では教師ベクトルを部分空間とする標本部分空間法に基づく SVM (ES-SVM) とよぶ。ES-SVM では、SS-SVM と同様に部分空間どうしの分離の最大化と誤識別の最大化を行うことにより識別の観点から最適な重みを付けを行う。ES-SVM においても重みが負となる軸が識別において重要となるかを確認するために非負制約を設ける場合と設けない場合で比較する。この手法により、SS-SVM と比べ部分空間の生成における学習コストが削減でき、また CRS と比べ識別におけるコストを削減するとともに、識別の観点から最適な分離超平面を決定するため汎化能力を向上させることができる。

以下、2 章では従来の SS-SVM を説明し、3 章では重みに基づく特徴選択手法について述べ、4 章では ES-SVM について述べる。5 章においては、ベンチマークデータを用いた計算機実験で従来の SS-SVM と提案手法を比較することにより提案手法の有効性を検証し、6 章で結論を述べる。

2. 部分空間法に基づくサポートベクトルマシン

2.1 クラス部分空間の生成

SS-SVM ではクラス部分空間の生成に標本特徴空間^{9),10)}上で定義した KPCA^{5),6)}を用いている。ここで、 n クラス m 次元の入力 \mathbf{x} における識別問題について考える。入力ベクトル \mathbf{x} を N 個の一次独立な教師ベクトル \mathbf{x}_{k_i} ($i = 1, \dots, N$) の張る N 次元標本特徴空間に写像関数 $\mathbf{h}(\mathbf{x}) = (H(\mathbf{x}_{k_1}, \mathbf{x}), \dots, H(\mathbf{x}_{k_N}, \mathbf{x}))^T$ により写像する。ここで、一次独立なデータはコレスキー分解を用いて選択する。このとき、高次元特徴空間への写像関数を $\mathbf{g}(\mathbf{x})$ としたとき、 $H(\mathbf{x}, \mathbf{x}') = \mathbf{g}^T(\mathbf{x})\mathbf{g}(\mathbf{x}')$ である。このとき、以下の共分散行列の固有値問題を解くことにより、クラス i 部分空間の r_i 個の軸 φ_{ik} を決定する。

$$\frac{1}{|X_i|} \sum_{j \in X_i} \mathbf{h}(\mathbf{x}_j) \mathbf{h}^T(\mathbf{x}_j) \varphi_{ik} = \lambda_{ik} \varphi_{ik} \quad (1)$$

ここで、 X_i はクラス i の教師ベクトル集合、 $|X_i|$ は X_i の教師ベクトル数、 λ_{ik} は φ_{ik} に対応する固有値を示す。また、固有値を大きい順に並べ替え、累積寄与率 a を用いて各クラス部分空間の軸の選択を行う。このとき $a(r_i) = \sum_{j=1}^{r_i} \lambda_{ij} / \sum_{j=1}^N \lambda_{ij} \times 100$ (%) と定義し、 a が閾値 κ 以上になるまで対応する固有ベクトルを軸として選択し、部分空間の次元数 r_i を決定する。

2.2 重みの最適化

SS-SVM において、各クラス部分空間の各軸における類似度を以下のような列ベクトルで定義する。

$$\mathbf{f}_i(\mathbf{x}) = \left(\frac{(\boldsymbol{\varphi}_{i1}^T \mathbf{h}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{i1}\|^2 \|\mathbf{h}(\mathbf{x})\|^2}, \dots, \frac{(\boldsymbol{\varphi}_{ir_i}^T \mathbf{h}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{ir_i}\|^2 \|\mathbf{h}(\mathbf{x})\|^2} \right)^T \quad (2)$$

このとき、決定関数を類似度の列ベクトルを用いて

$$D_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}) + b_i \quad (3)$$

と表す。ここで、 \mathbf{w}_i をクラス i 部分空間の k 番目の軸の重み w_{ik} の列ベクトルとする。このとき、入力ベクトル \mathbf{x} を $D_i(\mathbf{x})$ が最大となるクラス i に分類する。5章の実験では、従来の SSLP-SVM においてバイアス項が識別に悪影響を及ぼす場合はバイアス項を省き、実験を行っている。また、SSLS-SVM においてバイアス項 b_i を省く。

通常の SS-SVM では、各クラス間の部分空間を分離するためにマージンを最大化し、誤認識を最小化するように重みを最適化するため、最適化問題は同時定式化方式となる。そのとき、SSLS-SVM の最適化問題は以下のように表せる。

$$\text{minimize } Q(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2} \sum_{i=1}^n \|\mathbf{w}_i\|^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ y_j \neq i}}^M \frac{CM}{2n |X_{y_j}|} \xi_{ji}^2 \quad (4)$$

$$\text{subject to } \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) = 1 - \xi_{ji} \\ \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (5)$$

ここで M は全教師ベクトル数、 C はマージン最大化と誤認識最小化のトレードオフを決定するパラメータであり、 ξ_{ji} はスラック変数である。

SSLP-SVM では SSLS-SVM と異なり、最適化問題に重みの非負制約を設けている。またバイアス項 b_i は正または負の値をとるため異なる正の変数によって正負の値をとるようにする。すなわち、 $b_i = b_i^+ - b_i^-$ ($b_i^+ \geq 0, b_i^- \geq 0$) とする。さらに、不等号制約を等号制約に変換するため、非負のスラック変数 u_{ji} ($i = 1, \dots, n, i \neq y_j, j = 1, \dots, M$) を用いる。このときの SSLP-SVM の最適化問題は以下のように表せる。

$$\text{minimize } Q(\mathbf{w}, \mathbf{b}, \boldsymbol{\xi}, \mathbf{u}) = \sum_{i=1}^n \sum_{k=1}^{r_k} w_{ik} + \sum_{j=1}^M \sum_{i \neq y_j, i=1}^n \frac{CM}{n |X_{y_j}|} \xi_{ji} \quad (6)$$

$$\text{subject to } \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) + (b_{y_j}^+ - b_{y_j}^-) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) - (b_i^+ - b_i^-) = 1 - \xi_{ji} + u_{ji}$$

$$\text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (7)$$

$$\xi_{ji}, u_{ji} \geq 0 \text{ for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (8)$$

$$w_{ik} \geq 0 \text{ for } i = 1, \dots, n, k = 1, \dots, r_k \quad (9)$$

ここで、 $\mathbf{b} = (b_1, \dots, b_n)^T$, $\mathbf{u} = (\dots, u_{ji}, \dots)^T$ である。また SSLS-SVM とは異なり、重み w_{ik} が 0 となりうる。 $\mathbf{f}_i(\mathbf{x})$ の k 番目の要素を $f_{ik}(\mathbf{x})$ とすると、 $w_{ik} = 0$ となる $f_{ik}(\mathbf{x})$ は識別に影響を与えないため、対応する軸を削除することができる。すなわち SSLP-SVM では累積寄与率を用いることなく識別と同時に特徴選択を行うことができる。

3. 最適な重みに基づく特徴選択

SSLS-SVM では、式 (1) により求めた固有値の大きい固有ベクトルを条件を満たすまでクラス部分空間の軸として選択する。しかしながら、これらの固有値は識別においての重要性を意味してはいないため、固有値が大きい軸が識別の観点から冗長な軸であることがあり、逆に小さい軸が識別において重要である場合がある。すなわち、固有値に基づく特徴選択はより良い識別を行うための特徴選択とはいえない。この点、重みの最適化と同時に特徴選択を行える SSLP-SVM ではこのような問題は生じない。しかしながら、重みに非負制約を設けているため、識別において重要な軸であるが重みが負になりうるため削除される軸が存在しうる。

本章では、この問題点を解決するために SSLS-SVM における最適な重みに基づいた特徴選択と負の重みを許す SSLP-SVM の学習について述べる。

3.1 非負制約による SSLS-SVM の特徴選択

通常の部分空間法では各軸に対する重みは非負である。これに対して SSLS-SVM により最適化された重みの中でクラス i の負の重みを持つ軸の類似度 $f_{ik}(\mathbf{x})$ はある入力 \mathbf{x} がクラス i のデータである場合より他のクラスのデータである場合の方が大きな値を示しやすいことを意味している。そこで、重みが負である軸が有用であるか調べるために重みが負となる軸を削除する特徴選択を行う。まず、 $\kappa = 99.9$ (%) とおき、式 (1) により求めた固有ベクトルの中から軸を選択し、これらを用いて式 (4), (5) の最適化問題を解く。これにより各固有ベクトルにおける重みが決定でき、これらの重みの中で正の重みをとる固有ベクトルのみを選択し、それらを用いて再び式 (4), (5) の最適化問題を解く。この処理を全重みが正となるまで繰り返す。この手法は線形カーネルを用いた SVM における各軸の重みの大きさに基づく特徴選択法¹¹⁾ と本質的には同じである。

5章の実験において、カーネル関数として $\exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$ で表される RBF カーネル

を用いる．ここで γ はカーネルパラメータを表す．このとき，SSLS-SVM ではカーネルパラメータ γ とマージンパラメータ C を決定する必要がある．五分割交差検定を用いて決定する．また，この特徴選択を行う SSLS-SVM の学習アルゴリズムを以下に示す．

Algorithm 1

- Step 1** 五分割交差検定を用いて γ と C の値を決定する．このとき $\kappa = 99.9$ (%) とおく．
Step 2 Step 1 で決定した γ を用いて，コレスキー分解により教師データから一次独立なデータを選択し，これらを用いて標本特徴空間への写像関数 $\mathbf{h}(\mathbf{x})$ を求める． $i = 1$ とおく．
Step 3 式 (1) よりクラス i の固有ベクトル φ_{ik} と固有値 λ_{ik} を求める．
Step 4 κ を用いて固有ベクトルを選択し，クラス i 部分空間の次元数 r_i を決定する．
Step 5 式 (2) を用いて $\mathbf{f}_i(\mathbf{x}_j)$ ($j = 1, \dots, M$) を求める． $i = n$ ならば Step 6 に進み， $i \neq n$ ならば Step 3 に戻る．
Step 6 Step 1 で決定した C を用いて最適化問題 (4)，(5) を解くことにより重み \mathbf{w}_i ($i = 1, \dots, n$) を求める．
Step 7 全軸における重みが $w_{ik} > 0$ ($i = 1, \dots, n, k = 1, \dots, r_i$) ならば学習を終了する．そうでなければ，重みが負となる軸を削除し，クラス i における重みが正となる軸の数を r_i として Step 6 に戻る．

3.2 重みの絶対値に基づく SSLS-SVM の特徴選択

3.1 節では，重みが負となる軸を削除することにより特徴選択を行った．しかしながら，重みが負の軸においてもその絶対値が大きければ大きいほど，識別により大きな影響を及ぼすが悪影響を及ぼすとは限らない．識別の観点から最適化した重みであるため識別に良い影響を及ぼす可能性が高い．また，重みの絶対値が小さければ小さいほど識別に影響をあまり与えない軸であり，識別の観点から冗長な軸であるといえる．そこで，重みの絶対値が小さい軸を削除することで特徴選択を行う．このとき，選択となる絶対値の閾値 θ を以下に定義する．

$$\theta_i = \sqrt{T} \sum_{j=1}^{r_i} |w_{ij}| \quad (10)$$

ここで T は閾値の値を調整するパラメータを表す．また， $|w_{ij}| > \theta_i$ に対応する軸のみを選択し，すべての重みが閾値以上になるまで学習を繰り返す．

5 章の実験では，パラメータ T を五分割交差検定により決定する．以下に学習アルゴリ

ズムを示す．

Algorithm 2

- Step 1** $\kappa = 99.9$ (%) とおき，五分割交差検定を用いて γ と C と T の値を決定する．
Step 2 Step 1 で決定した γ を用いて，コレスキー分解により教師データから一次独立なデータを選択し，これらを用いて標本特徴空間への写像関数 $\mathbf{h}(\mathbf{x})$ を求める． $i = 1$ とおく．
Step 3 式 (1) よりクラス i の固有ベクトル φ_{ik} と固有値 λ_{ik} を求める．
Step 4 κ を用いて固有ベクトルを選択し，クラス i 部分空間の次元数 r_i を決定する．
Step 5 式 (2) を用いて $\mathbf{f}_i(\mathbf{x}_j)$ ($j = 1, \dots, M$) を求める． $i = n$ ならば Step 6 に進み， $i \neq n$ ならば Step 3 に戻る．
Step 6 Step 1 で決定した C を用いて最適化問題 (4)，(5) を解くことにより重み \mathbf{w}_i ($i = 1, \dots, n$) を求める．また，Step 1 で決定した T を用いて式 (10) より閾値 θ_i ($i = 1, \dots, n$) を決定する．
Step 7 全軸における重みが $|w_{ik}| > \theta_i$ ($i = 1, \dots, n, k = 1, \dots, r_i$) ならば学習を終了する．そうでなければ， $|w_{ik}| < \theta_i$ ($i = 1, \dots, n, k = 1, \dots, r_i$) となる軸を削除し，残った軸の数を r_i として Step 6 に戻る．

5 章の計算機実験により SSLS-SVM において 3.1 節の特徴選択法と本手法を比較することにより，非負制約を設けるべきかどうかを確認する．

3.3 負の重みを許す SSLP-SVM

前述に示したように，SSLP-SVM では学習過程で同時に特徴選択を行うため，前節までに SSLS-SVM に適用した特徴選択を行う必要はない．しかしながら，従来の SSLP-SVM は最適化問題に非負制約が設けてあり，SSLS-SVM と同様に識別において重要な軸を削除している恐れがある．そこで，SSLP-SVM の非負制約を省くことにより負の重みを許す手法を提案する．負の重みを許すために最適化問題 (6)–(9) から制約条件 (9) を省く．このとき，SSLP-SVM の最適化問題は以下のように表せる．

$$\text{minimize } Q(\mathbf{w}, \mathbf{b}, \boldsymbol{\xi}, \mathbf{u}) = \sum_{i=1}^n \sum_{k=1}^{r_k} w_{ik} + \sum_{j=1}^M \sum_{i \neq y_j, i=1}^n \frac{C M}{n |X_{y_j}|} \xi_{ji} \quad (11)$$

$$\text{subject to } \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) + (b_{y_j}^+ - b_{y_j}^-) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) - (b_i^+ - b_i^-) = 1 - \xi_{ji} + u_{ij} \\ \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (12)$$

$$\xi_{ji}, u_{ji} \geq 0 \text{ for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (13)$$

5章の計算機実験により SSLP-SVM において従来の非負制約を設けた手法と本手法を比較することにより、非負制約を設けるべきかどうかを確認する。

4. 標本部分空間法に基づくサポートベクトルマシン

4.1 CRS における決定関数

SS-SVM を含む多くの部分空間法では固有値問題を解くことにより各クラス部分空間を生成する。そのため、一次独立なデータ数 N の次元の固有値分解をクラス数回、すなわち n 回解く必要があるため、 N の値が大きければ、学習コストが膨大となる。ここで、CRS では特徴ベクトルの要素がすべて正であることを前提とし、各クラスのデータクラスごとに一部の凸錐の領域にのみ密集して存在していると考え、その領域をクラス部分空間とおき、入力 \mathbf{x} とその凸錐クラス部分空間との角度を類似度として識別を行う手法である。この凸錐部分空間はクラスごとの教師ベクトルそのものを部分空間の軸とすることで求めることができるため非常に容易に求めることができ、固有値問題による学習コストをすべて削減することができる。通常の CRS では以下のように入力 \mathbf{x} とクラス i 部分空間への正射影ベクトルとの角度 $\beta_i(\mathbf{x})$ を決定関数 $D_i(\mathbf{x})$ として定義する。

$$D_i(\mathbf{x}) = \beta_i(\mathbf{x}) = \arcsin \left(\sqrt{\min_{\alpha_{ij} \geq 0} \left\| \mathbf{x} - \sum_{j \in X_i} \alpha_{ij} \mathbf{x}_j \right\|^2 / \|\mathbf{x}\|^2} \right) \quad (14)$$

ここで、 α_{ij} は教師ベクトル \mathbf{x}_j における係数を表す。この決定関数 $D_i(\mathbf{x})$ は未知の入力があるときに非負最小自乗法を用いて係数ベクトルを計算し決定する。このため、大規模データにおいて識別コストが膨大となる。また、通常のパターン認識の問題では線形分離可能な問題はほとんど現実にはなく、非線形問題であり、各クラス凸錐部分空間どうしで重なりがあるとき、その領域内では対応するクラスの決定関数 $D_i(\mathbf{x})$ が 0 となるため未分類領域となる。

4.2 SVM による重みの最適化

CRS の問題点を解消するために重複する領域を小さくするためカーネル法を用い、さらに CRS の決定関数を SS-SVM の決定関数と類似した形に変更し、各クラスの標本をクラス部分空間の軸とし錐の制約を設けない標本特徴空間法 (ES) に SS-SVM と同様にマージン最大化と誤認識最小化の SVM の概念を導入する ES-SVM を提案する。

はじめに錐部分空間の重複を小さくするため高次元特徴空間への写像関数 $g(\mathbf{x})$ を用いて入力空間から高次元特徴空間に写像する。次に各クラスごとの教師ベクトルそのものをクラ

ス部分空間の軸とするため、SS-SVM の決定関数を以下の各クラスの教師ベクトルである軸における重み付けされた射影長の総和の形に変更する。

$$D_i(\mathbf{x}) = \sum_{j \in X_i} \frac{w_{ij} \mathbf{g}^T(\mathbf{x}) \mathbf{g}(\mathbf{x}_{i_j})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{i_j})\|} = \sum_{j=1}^{N_i} \frac{w_{ij} H(\mathbf{x}, \mathbf{x}_{i_j})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{i_j})\|} \quad (15)$$

ただし、 \mathbf{x}_{i_j} ($i = 1, \dots, n, j = 1, \dots, |X_i|$) はクラス i の j 番目の教師ベクトルを表し、 w_{ij} はクラス i 部分空間の j 番目の軸における重みである。ここで、高次元特徴空間上のベクトル \mathbf{x} のノルムは $\|\mathbf{g}(\mathbf{x})\| = \sqrt{H(\mathbf{x}, \mathbf{x})}$ と容易に求めることができる。また類似度 $f_i(\mathbf{x})$ は

$$\mathbf{f}_i(\mathbf{x}) = \left(\frac{H(\mathbf{x}, \mathbf{x}_{i_1})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{i_1})\|}, \dots, \frac{H(\mathbf{x}, \mathbf{x}_{i_{|X_i|}})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{i_{|X_i|}})\|} \right)^T \quad (16)$$

と定義でき、決定関数を式 (3) と同様に表せる。ここで重み w_{ij} に識別の観点から最適な値を設定することにより、入力ごとに非負最小自乗法を解く必要がなく、CRS に比べ識別コストを大幅に削減することができる。また他のクラスの決定関数の値と等しくなる領域以外で未分類領域は存在しないため、より良い識別を行うことができる。ここで、射影長を計算するベクトル数を削減するために式 (15) を各クラスの部分空間を張る N_i ($i = 1, \dots, n$) 個の一次独立なデータ \mathbf{x}_{ik_j} ($j = 1, \dots, N_i$) のみを用いて次式のように変更できる。

$$D_i(\mathbf{x}) = \sum_{j=1}^{N_i} \frac{w_{ij} H(\mathbf{x}, \mathbf{x}_{ik_j})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_j})\|} \quad (17)$$

このとき、類似度を

$$\mathbf{f}_i(\mathbf{x}) = \left(\frac{H(\mathbf{x}, \mathbf{x}_{ik_1})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_1})\|}, \dots, \frac{H(\mathbf{x}, \mathbf{x}_{ik_{N_i}})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_{N_i}})\|} \right)^T \quad (18)$$

と表すことができ、求める重みの数が少なくなるため、学習コストを削減できる。ここで、ある入力 \mathbf{x}_j の決定関数 $D_{y_j}(\mathbf{x}_j)$ が他のクラスの決定関数 $D_i(\mathbf{x}_j)$ ($i \neq y_j$) と比べて大きくなるような制約条件を設け、マージン最大化と誤認識最小化の概念の基に SVM を用いて SS-SVM と同様に最適化を行うことができる。ここで、SS-SVM と同様に用いる SVM のモデルとして LS-SVM と LP-SVM を用い、それぞれの手法を ESLS-SVM, ESLP-SVM とよぶ。

5章では前章に提案した特徴選択を行った ESLS-SVM と非負制約を設ける ESLP-SVM と非負制約を省いた ESLP-SVM において実験を行い比較・評価を行う。

5. 実験

本章では、2クラス識別問題においてベンチマークデータを用いた計算機実験を行い、従来の特徴選択と提案した特徴選択を用いた SS-SVM と ES-SVM との比較・評価を行った。また、すべての重みを等しく設定したときの従来のカーネル部分空間法 (KSM) と汎化能力の非常に高い識別器である SVM と比較し、それぞれの提案手法の汎化能力について評価する。

以下、従来の寄与率により軸を決定するそれぞれの SS-SVM には (A)、非負制約を設けたそれぞれの手法には (P)、重みの絶対値を基にした特徴選択を用いたそれぞれの手法には (Ab)、非負制約を省いたそれぞれの手法には (N) を記した。

5.1 ベンチマークデータ

表 1 に示す 2 クラスのベンチマークデータ^{12),13)} と表 2 に示す多クラスのベンチマークデータ¹⁴⁾ を用いた。各データセットは複数の教師データセットとテストデータセットの組から構成されており、表 1, 2 中の Inputs は入力次元数, Train は 1 つの教師データセットに含まれる教師データ数, Test は同様のテストデータ数, Sets はその組の個数, Classes はクラス数を示す。

5.2 重みに基づく特徴選択と負の重みの重要性の評価

従来の SS-SVM である SSLS (A), SSLP (P) と提案手法である SSLS (P), SSLS (Ab), SSLP (N) を比較・評価する。

5.2.1 パラメータ設定

それぞれの手法において、必要となるパラメータ κ, γ, C, T を $\kappa = \{80, 85, 90, 95, 99, 99.9\}$, $\gamma = \{0.01, 0.05, 0.1, 0.5, 1, 1.5, 3, 5, 10, 15\}$, $C = \{10^{-4}, 10^{-3}, 0.01, 0.1, 1, 10, 50, 100, 500, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4, 10^5\}$, $T = \{10^{-5}, 10^{-6}\}$ から最適な値を五分割交差検定を用いて決定する。また、多クラス問題において γ が大きい値を選択しやすい傾向があったため、さらに $\gamma = \{20, 30, 50, 100, 200, 500, 10^3\}$ を含めて五分割交差検定を用いて決定する。

表 3 に従来の SSLS-SVM, SSLP-SVM である SSLS (A), SSLP (P) のパラメータを示す。また、表 4 に SSLS (P), SSLS (Ab), SSLP (N) のパラメータを示す。ただし、多クラスデータにおいて SSLP-SVM において用いている計算機のパフォーマンス上, Numeral

表 1 2 クラスベンチマークデータ
Table 1 Two-class benchmark data.

Data	Inputs	Training	Test	Sets
Banana	2	400	4,900	100
B. cancer	9	200	77	100
Diabetes	8	468	300	100
German	20	700	300	100
Heart	13	170	100	100
Image	18	1,300	1,010	20
Ringnorm	20	400	7,000	100
F. solar	9	666	400	100
Splice	60	1,000	2,175	20
Thyroid	5	140	75	100
Titanic	3	150	2,051	100
Twonorm	20	400	7,000	100
Waveform	21	400	4,600	100

表 2 多クラスベンチマークデータ
Table 2 Multi-class benchmark data.

Data	Classes	Inputs	Training	Test
Iris	3	4	75	75
Numeral	10	12	810	820

表 3 従来の SS-SVM におけるパラメータ
Table 3 Parameter values for conventional SS-SVMs.

Data	SSLS (A)			SSLP (P)	
	κ (%)	γ	C	γ	C
Banana	99.9	15	50	5	10
B. cancer	85	3	0.1	0.5	1
Diabetes	95	3	10^3	0.5	100
German	80	3	10^4	5	10
Heart	80	1	1	3	0.5
Image	99	15	10^4	0.5	10^5
Ringnorm	99.9	0.1	10^4	0.1	10^3
F. solar	80	5	10^4	0.5	10^3
Splice	99	1	10^4	0.1	10^3
Thyroid	95	5	500	0.5	50
Titanic	80	5	0.1	0.5	0.5
Twonorm	80	0.1	50	0.5	1
Waveform	80	3	10^3	1	50
Iris	80	50	0.01	1.5	100
Numeral	100	100	0.01	-	-

表 4 特徴選択を用いた SSLS-SVM と重みの非負制約を省いた SSLP-SVM におけるパラメータ

Table 4 Parameter values for SSLS-SVMs using the proposed feature selection and SSLP-SVMs without non-negative weights constraints.

Data	SSLS (P)		SSLS (Ab)			SSLP (N)	
	γ	C	γ	C	T	γ	C
Banana	1	100	1	100	10^{-6}	0.5	10
B. cancer	1.5	5×10^{-4}	1.5	10^{-4}	10^{-5}	0.1	0.1
Diabetes	0.01	10	0.05	0.1	10^{-5}	0.01	1
German	3	0.01	3	0.1	10^{-6}	0.5	1
Heart	0.01	0.01	0.01	0.01	10^{-5}	0.5	10
Image	15	100	1.5	100	10^{-6}	0.5	100
Ringnorm	0.5	1	0.5	1	10^{-5}	0.1	0.01
F. solar	0.1	10	0.1	1	10^{-5}	0.5	1
Splice	1	1	1	1	10^{-6}	0.5	0.5
Thyroid	1.5	0.1	0.5	50	10^{-5}	0.5	1
Titanic	3	10^{-4}	1.5	10^{-4}	10^{-6}	0.01	1
Twonorm	0.1	0.1	1	0.01	10^{-5}	0.1	0.1
Waveform	1	0.1	1	0.1	10^{-5}	1	0.1
Iris	15	100	10	100	10^{-5}	0.01	10^4
Numeral	20	10	10	1,000	10^{-6}	-	-

データセットの計算不可能のため、Iris データセットのみの結果を示しています。

5.2.2 認識性能の評価

表 5 にテストデータセットにおける SSLS (A), SSLS (P), SSLS (Ab), SSLP (P), SSLP (N) の平均認識率と標準偏差を示す。ただし、各多クラスデータはそれぞれ組が 1 つのみであるため、その認識率のみを示す。ここで、各データセットにおいて最大となる平均認識率を太字で示し、2 クラス問題においてのみ従来手法である SSLS (A) または SSLP (P) と比べ、Welch の t 検定 (有意水準: 5%) により有意性を示した各データセットにおける提案手法の結果の前にアスタリスク (*), 有意性を示していないが劣っているとも判定されなかった結果の前に三角印 (Δ) を記す。

2 クラス問題

表 5 より、2 クラス問題において SSLS (P) は SSLS (A) と比べ 7 個のデータセットにおいて有意性を示しており、3 個のデータセットにおいて同等の結果を示している。また、SSLS (Ab) は SSLS (A) と比べ 9 個のデータセットにおいて有意性を示し、4 個のデータセットにおいて同等の結果を示している。これにより、従来の各クラスの各軸における分散を基にする特徴選択手法と比べ、重みに基づく特徴選択手法が識別において優れているとい

表 5 SS-SVM のテストデータにおける認識率 (%) と標準偏差の比較

Table 5 Comparison of average recognition rates (%) and their standard deviations of test datasets for SS-SVMs.

Data	SSLS (A)	SSLS (P)	SSLS (Ab)	SSLP (P)	SSLP (N)
Banana	88.9±0.6	*89.3±0.4	* 89.5 ±0.4	89.0±0.6	*89.3±0.5
B. cancer	75.1 ±4.4	Δ 75.1 ±4.2	Δ 75.0±4.1	73.3±4.6	68.6±4.7
Diabetes	72.0±2.3	* 75.2 ±2.0	*75.1±2.2	73.5±2.0	*74.7±2.0
German	73.8±2.2	* 75.2 ±2.3	Δ 74.4±2.1	70.9±6.6	Δ 71.1±2.3
Heart	82.6±3.7	*83.9±3.3	*83.9±3.2	83.1±3.8	Δ 84.0 ±3.4
Image	95.4±0.6	*96.3±0.7	* 97.0 ±0.6	95.1±1.0	Δ 95.7±0.9
Ringnorm	97.6±0.3	96.6±0.4	*98.2±0.2	98.2±0.2	* 98.5 ±0.1
F. solar	66.9 ±1.6	65.8±1.8	Δ 66.5±1.6	64.7±1.9	*66.7±1.8
Splice	86.2±1.0	*88.9±0.6	* 89.0 ±0.5	88.2±0.6	*88.9±0.5
Thyroid	95.4±2.1	Δ 95.3±2.3	*96.2±1.9	96.3 ±2.2	95.4±2.7
Titanic	76.9±1.0	Δ 77.0±1.5	* 77.3 ±0.3	76.8±1.1	Δ 76.9±1.0
Twonorm	97.7 ±0.1	97.6±0.1	Δ 97.7 ±0.1	97.6±0.2	Δ 97.6±0.1
Waveform	89.1±0.8	*89.8±0.5	* 90.0 ±0.5	89.3±0.7	Δ 89.5±0.6
Iris	97.3	96.0	97.3	90.7	90.7
Numeral	99.0	99.3	99.4	-	-

える。特に SSLS (Ab) は全データセットにおいて SSLS (A) と比べ同等以上の結果を示しており、また SSLS (P) と比べても優れているといえる。よって、SSLS-SVM において重みが負となる軸は識別の観点から必要であるといえる。また、SSLP (N) は SSLP (P) と比べ 5 個のデータセットにおいて有意性を示し、6 個のデータセットにおいて同等の結果を示している。これにより非負制約を設けた SSLS (P) よりも非負制約を省いた SSLP (N) が識別において優れているといえ、SSLP-SVM においても重みが負となる軸は識別の観点から必要であるといえる。

多クラス問題

表 5 より、多クラス問題において SSLS (P) は SSLS (A) と比べ、Iris データセットにおいては 1.3 (%) の認識率の劣化が見られるが、テストデータ数が 75 個であるため、1 データ多く誤認識しているのみのため、識別性能の劣化とはいえない。SSLP (N) は SSLS (P) と比べ、等しい認識率を示した。また SSLS (P) は SSLS (A) と比べ Numeral データセットにおいて高い認識率を示した。これより、SSLS (P) と SSLS (A) または SSLP (N) と SSLP (P) は本実験で用いた多クラス問題においても従来手法と比べ同等以上の識別性能であるといえる。また、SSLS (Ab) は SSLS (A) と比べ、Iris データセットにおいては等しい認識率を示し、Numeral データセットにおいては高い認識率を示し、多クラス問題におい

表 6 重みに基づく特徴選択による部分空間の次元数と軸の削減割合 (%)

Table 6 Dimension of each subspace and the deletion rate of dictionaries (%) by feature selection based on the associated weights.

Data	Class 1		Class 2		Deletion Rate (%)
	Before	After	Before	After	
Banana	60.48	57.64	63.01	60.19	4.6
B. cancer	135.91	74.21	56.81	54.32	33.3
Diabetes	157.96	78.87	141.71	78.18	47.6
German	489.46	478.96	210.54	210.51	1.5
Heart	81.52	43.66	72.33	39.94	45.7
Image	508.45	303.55	645.7	327.2	45.3
Ringnorm	202.04	42.29	199.96	58.79	74.7
F. solar	34.33	26.51	62.65	44.91	26.4
Splice	517	425.75	459.9	363.9	19.2
Thyroid	70.64	58.06	42.69	38.98	14.4
Titanic	7.81	6.71	10.19	9.05	22.4
Twonorm	197.92	113.7	202.08	115.35	42.7
Waveform	267.6	115.14	132.4	98.82	46.5

でも優れていることが分かる。これより、重みに基づく特徴選択法は多クラス問題においても従来手法と比べて優れているといえる。

5.2.3 特徴選択による軸の削減割合

本実験において、SSLS (Ab) で用いる閾値 T を $T = \{10^{-5}, 10^{-6}\}$ と 2 つの非常に小さい値から選択している。本項では、このように閾値が小さくても軸の削減が行われていることを確認し、また各軸に対応する固有値と重みの関係について述べる。

表 6 に 2 クラスベンチマークデータにおける SSLS (Ab) を用いたときの各クラス部分空間の次元数と全クラス部分空間の軸の削減割合を示す。ここで、Before は対応するクラスにおける寄与率の閾値 κ を $\kappa = 99.9$ (%) としたときの部分空間の次元数、After はその部分空間から重みの絶対値に基づいて特徴選択を行った最終的な次元数、“Deletion Rate” はこの特徴選択による全クラスの部分空間の軸の削減割合のそれぞれの平均を示す。

表 6 より、各データセットによって軸の削減数が異なるが、削減割合が 0-1 割となるデータセットが 2 個、1-2 割が 2 個、2-4 割が 3 個、4-6 割が 5 個、6-8 割が 1 個となっており、軸の削減をほとんどなされていないデータセットがあるものの大半のデータセットにおいて軸の削減がなされていることが分かる。特に Ringnorm データセットにおいては表 3 より SSLS (A) では $\kappa = 99.9$ (%) ではまったく軸の削減が行われていないが、SSLS (Ab) により 74.7 (%) も軸が削減され、さらに表 5 より認識率が向上していることから特徴選択

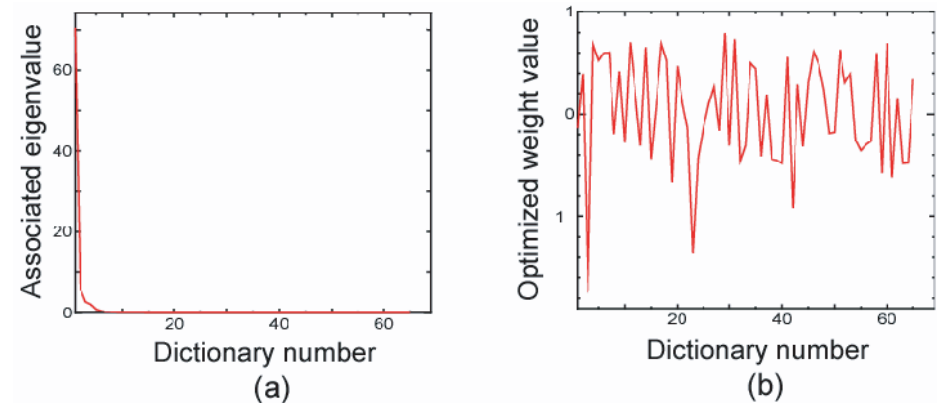


図 1 Thyroid データにおける各軸に対応する固有値と重みの関係 (a) 各軸に対する固有値 (b) 各軸に対する重み
Fig.1 The relation between eigenvalues and weights associated with dictionaries for the thyroid data (a) eigenvalues associated with dictionaries (b) weights associated with dictionaries.

によって識別性能が向上していることが分かる。削減割合は非常に小さいが Banana データセットにおいても同様のことがいえる。これにより、閾値 T の値を $T = \{10^{-5}, 10^{-6}\}$ の値に設定したとしても特徴選択は大半のデータセットにおいて行われ、さらに重みの絶対値による特徴選択により識別性能が向上していることが分かる。

図 1 では、各軸に対応する固有値と重みの関係を示すために、各軸の固有値を大きい順に並べ、Thyroid データセットにおけるそれぞれの固有値と重みを表した。これら 2 つの図を見比べると、固有値が徐々に小さくなっていくなかで、重みはそれに関係のない値となっている。固有値が非常に大きくても実際の重みは非常に小さい場合もあることから、固有値と重みに相関関係はほとんどないことが見てとれる。

5.2.4 正規直交系の標本特徴空間上での識別との評価

SS-SVM の全手法において、標本特徴空間を一次独立な教師ベクトルを基底ベクトルとして定義していたため、これらの基底ベクトルはそれぞれ直交していない。 $H_{ij} = H(\mathbf{x}_i, \mathbf{x}_j)$ とし、カーネル行列 H を $H = \{H_{ij}\} (i, j = 1, \dots, M)$ とすると正規直交系の標本特徴空間への写像関数 $\mathbf{h}(\mathbf{x})$ を以下のように再定義できる^{9),10)}。

$$\mathbf{h}(\mathbf{x}) = \Lambda^{-1/2} P^T (H(\mathbf{x}_1, \mathbf{x}), \dots, H(\mathbf{x}_M, \mathbf{x})) \quad (19)$$

ここで、 Λ は対角成分がカーネル行列 H における固有値となる $M \times M$ 対角行列であり、 P

20 標本部分空間法に基づくサポートベクトルマシンの開発

表 7 直交系と非直交系とのテストデータにおける認識率 (%) と標準偏差の比較

Table 7 Comparison of average recognition rates (%) and their standard deviations of test datasets for orthogonal and non-orthogonal systems.

Data	SSLS (P)	SSLS (P, O)	SSLS (Ab)	SSLS (Ab, O)
Banana	89.3±0.4	89.3±0.5	89.5±0.4	89.5±0.4
B. cancer	75.1±4.2	74.7±4.4	75.0±4.1	74.6±4.4
Diabetes	75.2±2.0	75.1±2.0	75.1±2.2	75.0±2.0
German	75.2±2.3	74.9±2.2	74.4±2.1	74.2±2.3
Heart	83.9±3.3	84.1±3.2	83.9±3.2	83.8±3.2
Image	96.3±0.7	95.8±0.8	97.0±0.6	97.0±0.7
Ringnorm	96.6±0.3	96.3±0.5	98.2±0.2	98.3±0.2
F. solar	65.8±1.8	66.3±1.8	66.5±1.6	66.3±1.9
Splice	88.9±0.6	88.2±0.7	89.0±0.5	88.5±0.6
Thyroid	95.3±2.3	95.1±2.4	96.2±1.9	96.2±1.8
Titanic	77.0±1.5	76.7±1.7	77.3±0.3	77.1±0.9
Twonorm	97.6±0.1	97.6±0.1	97.7±0.1	97.6±0.1
Waveform	89.8±0.5	89.8±0.6	90.0±0.5	89.9±0.6
Iris	96.0	96.0	97.3	97.3
Numeral	99.3	99.3	99.4	99.4

はそれらの固有値に対応する固有ベクトルの列ベクトルで構成される $M \times M$ 行列である。この写像関数を用いることにより、カーネル値が同じという意味で特徴空間と等価な標本特徴空間が構成できることが知られている^{9),10)}。写像関数の違いによる識別に対する影響はないかを式 (1) に式 (19) を代入した場合の手法との認識率の比較・評価をすることにより確認する。ここで、SS-SVM の中で全データセットにおいて高い認識率を示した SSLS (P) と SSLS (Ab) を用いて評価する。以下でそれぞれの手法に対応する標本特徴空間の基底ベクトルをそれぞれ直交化する手法を SSLS (P, O), SSLS (Ab, O) とよぶ。また、マージンパラメータ C 以外は対応するそれぞれの手法のパラメータと等しい値を設定し、 C のみ五分割交差検定で決定した。

表 7 より、SSLS (P), SSLS (Ab) の 2 つの手法ともそれぞれ SSLS (P, O), SSLS (Ab, O) と比べ、全データセットにおいて認識率はほぼ同等な値を示しており、それぞれ対応する識別手法の識別性能はあまり変わらないといえる。また、いくつかのデータセットにおいて若干の差が見られるのは、本実験において SSLS (P) と SSLS (Ab) では一次独立なデータの選定にコレスキー分解⁹⁾を用いており、コレスキー分解の対角成分の値の平方根が 10^{-6} 以下のとき 0 とおくため、厳密には一次独立なデータではないデータも選択されることも原因の 1 つであると考えられるが、その差の小ささから問題にする必要はない。この実験よ

表 8 ES-SVM におけるパラメータ

Table 8 Parameter values for ES-SVMs.

Data	ESLS (P)		ESLS (Ab)			ESLP (P)		ESLP (N)	
	γ	C	γ	C	T	γ	C	γ	C
Banana	3	10	1.5	500	10^{-5}	0.5	100	0.5	100
B. cancer	1	10^{-3}	0.05	10^{-3}	10^{-5}	0.05	1	0.1	1
Diabetes	0.01	500	0.01	500	10^{-6}	0.05	0.1	0.01	10
German	5	1	5	10	10^{-6}	0.1	1	0.05	1
Heart	0.1	1	0.01	100	10^{-5}	0.1	1	0.1	1
Image	5	5×10^3	5	500	10^{-6}	0.5	500	0.5	1000
Ringnorm	1	10	1	10	10^{-6}	1	1	0.1	1
F. solar	0.1	10	0.1	10	10^{-6}	0.1	10	0.5	1
Splice	1.5	100	1	10	10^{-6}	0.05	500	0.5	50
Thyroid	1	10	1	10	10^{-6}	3	1	1	10
Titanic	1.5	10^{-3}	1.5	10^{-3}	10^{-6}	1	1	1	1
Twonorm	1.5	0.1	0.5	1	10^{-5}	0.05	1	0.05	1
Waveform	1.5	1	1.5	1	10^{-6}	0.1	10	1.5	1
Iris	15	1	3	100	10^{-5}	0.5	500	0.5	500
Numeral	50	1	15	1000	10^{-6}	-	-	-	-

り、標本特徴空間の基底ベクトルを直交化しなくても同等の識別性能を示すという結果が得られた。

5.3 標本部分空間法に基づくサポートベクトルマシンの評価

SS-SVM と ES-SVM の認識性能と学習時間の比較・評価を行う。

5.3.1 パラメータ設定

ESLS (P), ESLS (Ab), ESLP (P), ESLP (N) に用いるパラメータを 5.2.1 項のパラメータ設定と同じ手法で決定する。表 8 に各手法におけるパラメータを示す。ただし、多クラスデータにおいて ESLP-SVM において用いている計算機のパフォーマンス上、Numerical データセットの計算不可能のため、Iris データセットのみの結果を示している。

5.3.2 テストデータにおける認識率の比較・評価

表 9 に ESLS (P), ESLS (Ab), ESLP (P), ESLP (N) のテストデータセットにおける平均認識率と標準偏差を示す。ただし、各多クラスデータはそれぞれ組が 1 つのみであるため、その認識率のみを示す。ここで、各データセットにおいて最大となる平均認識率を太字で示し、2 クラス問題においてのみ表 5 の SSLS (P), SSLS (Ab), SSLP (P), SSLP (N) のそれぞれの結果に比べ、Welch の t 検定 (有意水準: 5%) により有意性を示した各データセットにおけるそれぞれの用いている特徴選択手法に対応する ESLS (P), ESLS (Ab),

21 標本部分空間法に基づくサポートベクトルマシンの開発

表 9 ES-SVM のテストデータにおける平均認識率 (%) と標準偏差

Table 9 Average recognition rates (%) and their standard deviations of test datasets for ES-SVMs.

Data	ESLS (P)	ESLS (Ab)	ESLP (P)	ESLP (N)
Banana	* 89.5 ±0.4	△ 89.5 ±0.4	△89.1±0.6	89.1±0.5
B. cancer	△73.9±4.7	△ 74.1 ±4.5	69.7±5.1	△69.0±4.5
Diabetes	△ 75.0 ±2.0	△ 75.0 ±2.0	*74.3±2.8	△75.0±2.1
German	74.3±2.1	△ 74.7 ±2.5	△71.6±2.6	△71.6±2.5
Heart	△84.0±3.3	△84.0±3.2	* 84.5 ±3.1	△ 84.5 ±3.1
Image	*96.8±0.5	△ 97.0 ±0.6	*96.2±0.5	△95.8±0.6
Ringnorm	*96.8±0.4	96.7±0.4	* 98.3 ±0.2	98.0±0.3
F. solar	*66.5±1.5	△66.5±1.6	* 67.3 ±1.8	*67.2±1.7
Splice	△ 89.1 ±0.7	△ 89.1 ±0.6	87.7±0.9	87.7±0.9
Thyroid	*95.9±1.9	△ 96.0 ±2.0	94.5±2.8	△94.9±2.3
Titanic	* 77.3 ±0.4	△ 77.3 ±0.3	76.3±2.1	76.3±2.1
Twonorm	△ 97.6 ±0.1	97.6 ±0.1	97.3±0.2	97.3±0.2
Waveform	89.7±0.5	89.8 ±0.5	88.6±0.5	88.8±0.7
Iris	96.0	96.0	93.3	93.3
Numeral	99.3	99.3	-	-

ESLP (P), ESLP (N) の結果の前にアスタリスク (*), 有意性を示していないが劣っているとも判定されなかった結果の前に三角印 (△) を記す.

2 クラス問題

表 9 より ESLS (P) は SSLS (P) と比べ 6 個のデータセットにおいて有意性を示し, 5 個のデータセットにおいて同等の結果を示した. これにより, 非負制約を設けた場合, ESLS-SVM が SSLS-SVM よりも高い汎化能力を示すという結果が得られた. また表 9 より ESLS (Ab) は SSLS (Ab) と比べ, 10 個のデータセットにおいて有意性の観点から同等の結果を示しており, 有意性を示したデータセットはないが, 劣る結果を示した 3 個のデータセットのうち Twonorm データセットと Waveform データセットにおいては平均認識率, 標準偏差ともにそれほど差がなくほぼ同等であるといえる. これより ESLS-SVM の汎化能力は SSLS-SVM と比べ, 若干の劣化はあるものの同等であるといえる. また, ESLS (Ab) は 10 個のデータセットにおいて平均認識率が最大となり, その他の ES-SVM と比べて最も汎化能力が良いといえる.

また, ESLP (P) は SSLP (P) と比べると 5 個のデータセットにおいて有意性を示し, 2 個のデータセットにおいて同等の結果を示している. これより, ESLP (P) は SSLP (P) とほぼ同等の識別能力を持つといえる. ESLP (N) は SSLP (N) と比べると 1 個のデータセットにおいて有意性を示しており, 6 個のデータセットにおいて同等の結果を示している. こ

表 10 学習時間 (s) の比較

Table 10 Comparison of training time (s).

Data	SSLs (P)	SSLs (Ab)	SSLP (N)	ESLS (P)	ESLS (Ab)	ESLP (N)
Banana	0.39	0.39	4.59	0.29	0.29	4.21
B. cancer	0.23	0.23	0.87	0.13	0.12	0.61
Diabetes	0.64	1.33	8.59	0.39	0.57	4.49
German	9.38	9.67	65.4	6.45	6.51	32.2
Heart	0.13	0.13	0.55	0.09	0.09	0.46
Image	53.1	46.6	406	36.1	39.8	624
Ringnorm	1.76	1.78	11.9	1.25	1.26	4.5
F. solar	1.24	1.25	19.9	0.87	0.98	14.5
Splice	26.7	25.9	208	18.7	19.2	538
Thyroid	0.08	0.06	0.19	0.05	0.05	0.0.9
Titanic	0.02	0.02	0.18	0.02	0.02	0.09
Twonorm	1.73	1.73	9.34	1.24	1.25	2.79
Waveform	1.92	1.86	11.2	1.31	1.21	5.91
Numeral	3.11	3.25	-	2.31	2.25	-

れより, ESLP (N) は SSLP (N) と比べ, 識別能力が若干劣るといえる.

ESLS (P) と ESLS (Ab), または ESLP (P) と ESLP (N) は大半のデータセットにおいてほぼ同等の識別能力を示した. この原因として, カーネル法を用いることによって部分空間どうしの重なりが小さくなり, 重みが負となる軸における他のクラスに属する入力の種類もそれほど大きくなり, 識別にあまり影響を与えていないことが考えられる.

多クラス問題

表 9 より, 多クラス問題において ESLS (P) は SSLS (P) と比べ, Iris データセット, Numeral データセットともに認識率が等しい. これにより, これらのデータセットにおいて識別性能は同等であるといえる. また, ESLS (Ab) は SSLS (Ab) と比べ, Iris データセットにおいて 1.3 (%), Numeral データセットにおいて 0.1 (%) の劣化が見られた. しかしながら, Numral データセットにおいてはほぼ同等といえ, また Iris データセットにおいてもテストデータ数が 75 個であることから 1 個多くのデータを誤認識しているのみであるため同等といえる. これより, 多クラス問題においても ESLS (Ab) は SSLS (Ab) と比べ, 同等の識別性能といえる. また, ESLP (P), ESLP (N) は SSLP (P), SSLP (N) のそれぞれと比べ, Iris データセットにおいて, 高い認識率を示した.

5.3.3 学習時間の比較・評価

表 10 に SSLS (P), SSLS (Ab), ESLS (P), ESLS (Ab), SSLP (N), ESLP (N) の各

22 標本部分空間法に基づくサポートベクトルマシンの開発

表 11 テストデータセットにおける KSM と SVM と SSLs (Ab) と ESLS (Ab) の平均認識率 (%) と標準偏差

Table 11 Average recognition rates (%) and their standard deviations of test datasets for KSM, SVM, SSLs (Ab), and ESLS (Ab).

Data	SSLs (Ab)	ESLS (Ab)	SVM	KSM
Banana	*89.5±0.4	*89.5±0.4	89.3±0.5	88.6±0.6
B. cancer	*75.0±4.1	*74.1±4.5	72.4±4.6	75.0±4.2
Diabetes	75.1±2.2	75.0±2.0	76.3±1.8	73.4±1.7
German	74.1±2.1	74.7±2.5	76.2±2.2	75.1±2.2
Heart	△83.9±3.2	△84.0±3.3	83.7±3.4	80.4±3.3
Image	△97.0±0.6	△97.0±0.6	97.3±0.4	96.3±0.6
Ringnorm	*98.2±0.2	96.7±0.4	97.8±0.3	76.5±11.1
F. solar	66.5±1.8	66.5±1.6	67.6±1.7	65.1±1.8
Splice	△89.0±0.5	△89.1±0.6	89.2±0.7	87.6±0.8
Thyroid	△96.2±1.9	△96.0±2.0	96.1±2.0	95.6±2.0
Titanic	*77.3±0.3	*77.3±0.3	77.2±1.1	76.6±1.2
Twonorm	*97.7±0.1	△97.6±0.1	97.6±0.1	97.6±0.1
Waveform	△90.0±0.5	89.8±0.5	90.0±0.4	88.5±0.6

データセットにおける 1 ファイルあたりの平均学習時間をそれぞれ示す。ここで、Iris データセットにおいては、学習時間が非常に短く、計測不可能であったため省略する。

表 10 より、ESLS (P), ESLS (Ab) において SSLs (P), SSLs (Ab) と比べて全データセットにおいて学習時間が短縮されている。また本実験で用いるデータの中でデータ数の多い German データセット、Image データセット、Splice データセットにおいても学習時間が約 $\frac{2}{3}$ に削減できており、大規模データにおいても有効であることがいえる。また ESLS (N) においても SSLP (N) と比べて大半のデータセットにおいて学習時間を大幅に削減できている。しかしながら、Image データセットや Splice データセットの大規模データセットにおいては SSLP と比べて学習時間が増加している。この原因として重みの最適化の際に用いている単体法による計算コストが膨大になるためであると考えられる。

5.4 KSM と SVM との比較

本節では SS-SVM, ES-SVM において高い汎化能力を示した SSLs (Ab), ESLS (Ab) を RBF カーネルを用い、パラメータを五分割交差検定で決定した KSM, SVM と 2 クラス問題において比較する。表 11 にそれぞれの手法におけるテストデータセットにおける平均認識率と標準偏差を示す。ここで、各データセットにおいて最大となる平均認識率を太字で示し、SVM と比べ、Welch の t 検定 (有意水準: 5%) により有意性を示した各データセットにおける SSLs (Ab) と ESLS (Ab) の結果の前にアスタリスク (*), 有意性を示し

ていないが劣っているとも判定されなかった結果の前に三角印 (△) を記した。

5.4.1 KSM との比較

表 11 より、SSLs (Ab), ESLS (Ab) とともに 11 個のデータセットにおいて、KSM と比べて高い認識率を示している。特に、Ringnorm データセットにおいて、KSM の認識率は非常に低く、標準偏差も明らかに大きいことから KSM は不安定な識別手法であるといえる。SSLs (Ab) と ESLS (Ab) ではこのような傾向を示すデータセットはなく、安定して高い識別能力を示す識別手法であるといえる。また、KSM と比べ認識率の劣るデータセットにおいても認識率の劣化はあまり見られず同等であるといえる。このことから、部分空間法の識別性能を向上させることができているといえる。

5.4.2 SVM との比較

表 11 より、SVM と比べ SSLs (Ab) において 5 個のデータセットにおいて有意性を示し、5 個のデータセットにおいて同等の結果を示している。これより SSLs (Ab) は SVM と比べ、同等以上の高い汎化能力を持つ識別器であるといえる。

また、SVM と比べ ESLS (Ab) において 4 個のデータセットにおいて有意性を示し、5 個のデータセットにおいて同等の結果を示している。これより ESLS (Ab) は SVM と比べ、ほぼ同等の汎化能力を持つ識別器であるといえる。また SSLs (Ab) と比べても 10 個のデータセットにおいて同等の結果を示しており、劣っていると判定された Ringnorm データセット、Twonorm データセット、Waveform データセットにおいて平均認識率の差はほとんどなくほぼ同等の汎化能力を持っているといえる。さらに 5.3.3 項における学習時間の比較を考慮すると ESLS-SVM が高速かつ高い汎化能力を持つ優れた識別器であることがいえる。なお、5% で有意差の検定を行ったが、これを 1% として、SSLs (Ab) と SVM を比較すると、Waveform のみで結果が異なった。Waveform では SSLs (Ab) の標準偏差が大きいとの判定であった。しかしながら、この差は小さく、1% で検定しても大差のない結果であるといえる。

6. ま と め

本論文では、1) SSLs-SVM における最適な重みに基づく特徴選択の評価、2) SS-SVM における部分空間生成による学習コストの削減のための ES-SVM の提案を述べた。1) の手法では、SSLs-SVM により最適化された軸の重みに基づき条件を満たすまで選択処理を繰り返す。その条件として 2 つ設定した。1 つ目は非負制約であり、最適化のたびに重みが正となる軸のみを選択し続ける。2 つ目は、すべての重みの絶対値が閾値より大きくなること

である．ここでは，最適化と同時に特徴選択を行う SSLP-SVM から非負制約を省いた手法を考慮した手法も述べた．計算機実験により SSLS-SVM と SSLP-SVM とともに負の重みを許す手法と許さない手法を比較することにより重みが負となる軸が識別において必要であるかを評価した．2) の手法では，各クラスの教師ベクトルをそれぞれ各クラス部分空間を構成する軸とする CRS の考え方を取り入れ SS-SVM と同様に SVM のマージン最大化と誤識別最小化を考慮し各軸における類似度に対し識別の観点から最適な重み付けを行い，各クラス部分空間生成における学習コストの削減を図った．計算機実験では ESLS-SVM において 1) のそれぞれの特徴選択を取り入れ，ESLP-SVM では非負制約を設ける手法と設けない手法をそれぞれ用いた．

計算機実験により 1) の特徴選択法を用いた SS-SVM が従来の SS-SVM と比べ有効であることが示すことができた．さらに 1) を用いた SSLS-SVM と SSLP-SVM において非負制約を設けない手法が非負制約を設ける手法に比べ有効であり，重みが負となる軸が識別において必要な可能性があることを示した．また ES-SVM は 1) の手法を用いた SS-SVM と比べてもほぼ同等の識別性能を示し，学習コストも大半のデータセットにおいて削減することができた．さらに 1) の手法を用いた SSLS-SVM と ESLS-SVM は SVM と比べ同等以上の高い識別性能を示すことができ，有効性を示すことができた．

参 考 文 献

- 1) Watanabe, S. and Pakvasa, N.: Subspace methods of pattern recognition, *Proc. 1st IJCP*, pp.283–328 (1973).
- 2) Oja, E.: *Subspace Methods of Pattern Recognition*, Research Studies Press (1983).
- 3) Abe, S.: *Support Vector Machines for Pattern Classification*, Springer (2005).
- 4) Kitamura, T., Takeuchi, S., Abe, S. and Fukui, K.: Subspace-based support vector machines for pattern classification, *Neural Networks*, Vol.22, pp.558–567 (2009).
- 5) Shölkopf, B., Mika, S., Burges, C.J.C., Knirsch, P., Müller, K.R., Rätsch, G. and Smola, A.J.: Input space vs. feature space in kernel-based methods, *IEEE Trans. Neural Networks*, Vol.10, No.5, pp.1000–1017 (1999).
- 6) Shölkopf, B., Smola, A.J. and Müller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, Vol.10, pp.1299–1319 (1998).
- 7) Kitamura, T., Takeuchi, S. and Abe, S.: Feature selection and fast training for subspace based support vector machines, *Proc. International Joint Conference on Neural Networks (IJCNN2010)*, pp.1967–1972 (2010).
- 8) 小林 匠，大津展之：パターン認識のための錐制約部分空間法，電子情報通信学会論

文誌 (D), Vol.J92-D, No.1, pp.104–111 (2009).

- 9) Abe, S.: Sparse least squares support vector training in the reduced empirical feature space, *Pattern Analysis and Applications*, Vol.10, No.3, pp.203–214 (2007).
- 10) Xiong, H., Swamy, M.N.S. and Ahmad, M.O.: Optimizing the kernel in the empirical feature space, *IEEE Trans. Neural Networks*, Vol.16, No.2, pp.460–474 (2005).
- 11) Guyon, I., Weston, J., Barnhill, S. and Vapnik, V.: Gene selection for cancer classification using support vector machines, *Machine Learning*, Vol.46, No.1–3, pp.389–422 (2002).
- 12) Rätsch, G., Onda, T. and Müller, K.R.: Soft margins for AdaBoost, *Machine Learning*, Vol.42, No.3. pp.287–320 (2001).
- 13) <http://ida.first.fraunhofer.de/projects/bench/benchmarks.htm>
- 14) <http://cs.nyu.edu/~roweis/data.html>

(平成 22 年 2 月 4 日受付)

(平成 22 年 3 月 26 日再受付)

(平成 22 年 4 月 13 日採録)

北村 拓也

昭和 60 年生．平成 20 年神戸大学工学部電気電子工学科卒業．平成 21 年同大学院工学研究科電気電子工学専攻博士課程前期課程修了．現在，同大学院工学研究科電気電子工学専攻博士課程後期課程在籍．パターン認識の理論と応用に関する研究に従事．IEEE 会員．

竹内 翔吾

昭和 60 年生．平成 20 年神戸大学工学部電気電子工学科卒業．平成 22 年同大学院工学研究科電気電子工学専攻博士課程前期課程修了．パターン認識の理論と応用に関する研究に従事．IEEE 会員．

阿部 重夫 (正会員)

昭和 22 年生．昭和 47 年京都大学大学院工学研究科電気工学専攻修士課程修了．企業の研究所を経て，平成 9 年神戸大学工学部電気電子工学科教授．平成 19 年同大学院工学研究科教授となり現在に至る．ニューロ，ファジィ，サポートベクトルマシンの研究に従事．昭和 59 年，平成 7 年電気学会論文賞，平成 19 年日本知能情報ファジィ学会著述賞受賞．IEEE，INNS，電気学会等の各会員．