

# 部分空間 SVM のための交差部分空間学習

## Cross Subspace Learning for Subspace Support Vector Machine

井之上 直矢\*  
Naoya INOUE

山下 幸彦\*  
Yukihiko YAMASHITA

### Abstract:

The support vector machine (SVM) classifies an input pattern by using a hyperplane. The SVM demonstrated high generalization ability, and is being widely researched now. We propose a new SVM-based classifier called the subspace SVM (SSVM). The SSVM has a restriction that the normal vector of its hyperplane is included in a subspace. In order to provide the subspace, we split a set of samples into two sets. One set is used to compose the normal vector, and the other set is used to train parameters in the normal vector and the threshold, similarly to the cross validation method. We call this method the cross subspace learning for the SSVM. We conducted experiments with 13 datasets in order to show the advantage of the SSVM.

**Keywords:** Pattern recognition, Support vector machine, Cross subspace learning, Cross-validation

## 1 まえがき

1990年代前半, Vapnikらがパーセプトロンをカーネル法により非線形へ拡張した Support Vector Machine (SVM) [1] を提案した。それは1960年代に Vapnik氏が提案した, マージン最大化という基準での最適な識別境界を与える線形識別機 Optimal Hyperplane Classifier (OHC) にカーネル法を用いたものである。OHC自体は, その解を求めるために凸2次計画問題を解かなくてはならないため, 提案された時代には実用的ではなかった。その上 Minskyらによって線形識別の限界が指摘され [7], 広く認知されることはなかった。しかし, 計算機の飛躍的な進化とカーネル法によりその高い汎化能力が示され [2], SVMは注目を浴びるようになり, 現在も広く研究されている。

本研究では, 部分空間 SVM (SSVM) と呼ぶ SVM を拡張した新しい識別器を提案する。SSVMは, SVMにその超平面の法線ベクトルが部分空間に含まれるという制限を導入したものである。SVMでは法線ベクトルは標本点によって張られ, そのパラメータの学習もまた同じ標本点を使って行われる。SVMのこの特性はパラメータ選択という観点からは, 汎化性を損なう可能性がある。そこで SSVM に交差検定と同様の手法を用いて

みることにする。

交差検定は汎化性能を検定する方法であり, パラメータ選択にもよく用いられる [4, 5]。機械学習において学習標本点に対して最適なパラメータを検定すると未知のデータに対する識別精度が悪化することがある。これが過学習と呼ばれる現象である。複雑な識別器では学習データに対する学習が誤差0で実現できるため過学習が起こりやすい。SVMの場合, 一般にカーネル関数のパラメータやスラック変数の重みを調節することで過学習を避けている [6]。交差検定では学習標本点の集合を学習データとテストデータに分割し, テストデータでパラメータを評価する。それにより学習データに特化した学習結果になることがなくなり, 過学習を抑えることができる。

SSVMを使えば識別境界の法線ベクトルを張る標本点とそのベクトルのパラメータを決める標本点を異なるものに行うことができる。そこで標本点の集合を2つに分けて, 片方を SSVM の制限部分空間を張るために用いる。識別境界の法線ベクトルは, このセットのベクトルの一次結合となる。そしてもう片方を, SSVM の学習と同じように, 法線ベクトルとしきい値のパラメータを学習するために用いる。この解も凸2次計画問題で解くことができる。また交差検定法と同じように標本点の分け方は変えることが可能で最終的な結果は分割それぞれでの結果を平均したものとす。この手法を SSVM のた

\*東京工業大学, 152-8553 東京都目黒区大岡山 2-12-1, tel. 03-5734-3497, e-mail {n708i,yamasita}@ide.titech.ac.jp, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo

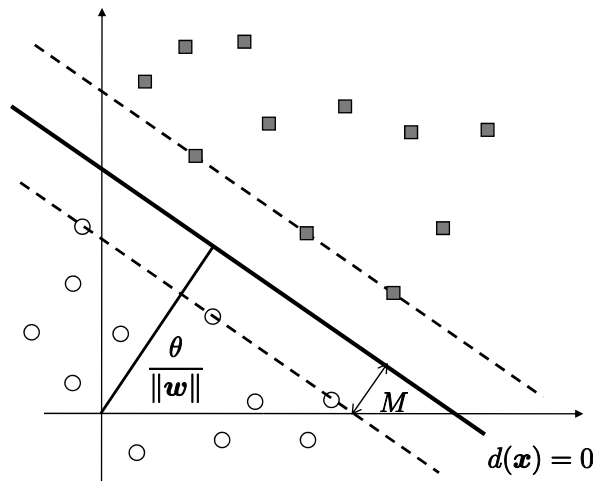


図 1: SVM とマージン

めの交差部分空間学習と名付ける。

交差部分空間学習法は、学習標本点を分割して利用するという面では、バギングなどのアンサンブル学習と類似しているようにも見える。しかしながら、アンサンブル学習では、一部分の学習標本点を使って学習した弱学習器の識別結果を、多数決などによって組み合わせるのに対して、交差部分空間学習法は、最終的な識別規則は元のサポートベクトルマシンと全く同じであり、法線ベクトルの重みやしきい値などのパラメータを決定するときに、交差検定の考え方を取り入れるために学習標本点を分割するもので、アンサンブル学習とは異なるものである。

本稿では、提案手法の有用性を示すために 2 クラス識別用の 13 のデータセットを用いて実験を行った。

## 2 Support Vector Machine

SVM は 2 クラス識別器である。クラスラベルを  $+1$  と  $-1$  で表し  $\mathbb{Y} = \{+1, -1\}$  とすることとし、ラベルが  $+1$ 、 $-1$  の母集団をそれぞれ  $\Omega_{+1}$ 、 $\Omega_{-1}$  で表す。あるベクトル  $w$  とスカラー  $\theta$  に対して、OHC、もしくは線形 SVM の識別関数は

$$d(x) = \langle w, x \rangle + \theta. \quad (1)$$

で与えられる。

未知パターン  $x$  を識別するときは、 $d(x)$  が正の時  $x \in \Omega_{+1}$ 、 $d(x)$  が負の時  $x \in \Omega_{-1}$  として識別する。

図 1 が示すように識別境界となる  $\{x \mid d(x) = 0\}$  は  $\mathbb{R}^N$  上の超平面となる。標本点を分離する超平面の中で未知の入力を最もよく分離できると考えられる超平面はマージン  $M$  を最大にするような超平面であるとする。

マージンを最大化するためには  $y_i(\langle w, x_i \rangle + \theta) \geq 1 \quad \forall_i$  の条件のもと、ノルム  $\|w\|^2$  を最小化すればよい。しかしながら、現実問題として標本点の線形分離が不可能な場合も多いため、また SVM の汎化性能を向上させるために、スラック変数  $\xi_l \geq 0$  ( $l = 1, 2, \dots, L$ ) を導入し、制約条件を

$$y_l(\langle w, x_l \rangle + \theta) \geq 1 - \xi_l. \quad (2)$$

とする。

$\xi_l = 0$  ならば、元の制約条件と同じになる。 $\xi_l > 0$  ならば、制約条件を緩めたことになり、 $\xi_l > 1$  で特に式 (2) の等号が成立する場合には、その標本点は誤認識することになる。このような標本点ばかりでは認識できなくなるため、 $\xi_l$  を小さくするような項を最小化する評価基準に加える。このような手法をソフトマージンと呼び、 $C$  を  $\xi_l$  を小さくするための項の定数とする。以上をまとめると SVM の評価基準は次のようになる。 $l = 1, 2, \dots, L$  に対する

$$y_l(\langle w, x_l \rangle + \theta) - 1 + \xi_l \geq 0, \quad (3)$$

$$\xi_l \geq 0, \quad (4)$$

の条件のもとに、

$$\frac{1}{2} \|w\|^2 + C \sum_{l=1}^L \xi_l \quad (5)$$

を最小とする。この制約付きの最適化問題を Lagrange の未定乗数法を使って変形すると

$$0 \leq \alpha_l \leq C, \quad \sum_{l=1}^L \alpha_l y_l = 0. \quad (6)$$

の条件のもとに  $\alpha_l$  ( $l = 1, 2, \dots, L$ ) について、

$$L_D = \sum_l \alpha_l - \frac{1}{2} \sum_{l=1}^L \sum_{k=1}^L \alpha_l \alpha_k y_l y_k \langle x_l, x_k \rangle \quad (7)$$

を最大化すればよいことになる。

上の手法はカーネル法によって非線形へと拡張される。つまり非線形写像  $\Phi$  によって高次元の特徴空間に写像してから識別を行うことになる。2 つのベクトル  $x$  と  $z$  に対して

$$k(x, z) = \langle \Phi(x), \Phi(z) \rangle. \quad (8)$$

が成立するため、上で現れる内積をカーネル関数に置き換えればよい。サポートベクトルの集合  $SV$  を

$$SV = \{x_l \mid 0 < \alpha_l \leq C\}. \quad (9)$$

によって与える。このとき識別関数は

$$d(x) = \sum_{x_l \in SV} \alpha_l y_l k(x_l, x) + \theta. \quad (10)$$

となる。

### 3 部分空間SVMのための交差部分空間学習

#### 3.1 部分空間SVM (SSVM)

この節では識別超平面の法線ベクトル  $w$  が部分空間に属するという制限を加え, SVM を拡張した SSVM について説明する.

$\{z_n\}_{n=1}^N$  をその制約部分空間を張るためのベクトルとする. すると実数の組  $\{\beta_n\}_{n=1}^N$  に対して,

$$w = \sum_{n=1}^N \beta_n z_n. \quad (11)$$

と置くことができ,

$$\|w\|^2 = \sum_{m=1}^N \sum_{n=1}^N \beta_m \beta_n \langle z_m, z_n \rangle. \quad (12)$$

となる.

$\{(x_l, y_l)\}_{l=1}^L$  を学習標本点とする. SSVM の学習は  $l = 1, 2, \dots, L$  に対して

$$y_l \left( \sum_{n=1}^N \beta_n \langle z_n, x_l \rangle + \theta \right) - 1 + \xi_l \geq 0, \quad \xi_l \geq 0 \quad (13)$$

の条件のもとに,

$$\frac{1}{2} \sum_{m=1}^N \sum_{n=1}^N \beta_m \beta_n \langle z_m, z_n \rangle + C \sum_{l=1}^L \xi_l \quad (14)$$

を最小化すればよい.

ラグランジュの未定係数  $\alpha_l, \mu_l$  ( $l = 1, 2, \dots, L$ ) を導入して, ラグランジュ関数は

$$\begin{aligned} L_P^S = & \frac{1}{2} \sum_{m=1}^N \sum_{n=1}^N \beta_m \beta_n \langle z_m, z_n \rangle \\ & + C \sum_{l=1}^L \xi_l - \sum_{l=1}^L \mu_l \xi_l \\ & - \sum_{l=1}^L \alpha_l \left[ y_l \left( \sum_{n=1}^N \beta_n \langle z_n, x_l \rangle + \theta \right) - 1 \right] \end{aligned} \quad (15)$$

となる.

KKT 条件より,  $l = 1, 2, \dots, L$ ,  $n = 1, 2, \dots, N$  に対して,

$$\frac{\partial L_P^S}{\partial \beta_n} = \sum_{m=1}^N \beta_m \langle z_n, z_m \rangle - \sum_{l=1}^L \alpha_l y_l \langle z_n, x_l \rangle = 0, \quad (16)$$

$$\frac{\partial L_P^S}{\partial \theta} = - \sum_{k=1}^L \alpha_k y_k = 0, \quad (17)$$

$$\frac{\partial L_P^S}{\partial \xi_l} = C - \alpha_l - \mu_l = 0, \quad (18)$$

$$y_l \left( \sum_{k=1}^N \beta_k \langle z_k, x_l \rangle + \theta \right) - 1 + \xi_l \geq 0, \quad (19)$$

$$\xi_l \geq 0, \quad \alpha_l \geq 0, \quad \mu_l \geq 0. \quad (20)$$

$$\alpha_l \left[ y_l \left( \sum_{n=1}^N \beta_n \langle z_n, x_l \rangle + \theta \right) - 1 + \xi_l \right] = 0, \quad (21)$$

$$\mu_l \xi_l = 0. \quad (22)$$

が成立する.

ベクトルまたは行列  $A$  の転置を  $A^T$  で表すことにし, 次のベクトルを定義する.

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_L)^T, \quad (23)$$

$$\beta = (\beta_1, \beta_2, \dots, \beta_N)^T, \quad (24)$$

$$y = (y_1, y_2, \dots, y_L)^T, \quad (25)$$

$$\mathbf{1} = (1, 1, \dots, 1)^T. \quad (26)$$

また,  $(N, N)$ -行列  $Z$  を,

$$(Z)_{mn} = \langle z_m, z_n \rangle. \quad (27)$$

で定義する.  $Z$  は可逆とする. 少なくとも非負であるので, 可逆でない場合は  $Z$  の代わりに,  $Z + \varepsilon I$  によって正則化する. ここで,  $\varepsilon > 0$  は正則化パラメータであり,  $I$  は単位行列である. さらに  $(N, L)$ -matrix  $D$  を,

$$(D)_{nl} = y_l \langle z_n, x_l \rangle. \quad (28)$$

で定義する. 式 (16), (17) はそれぞれ

$$Z\beta - D\alpha = 0, \quad (29)$$

$$y^T \alpha = \langle y, \alpha \rangle = 0, \quad (30)$$

となる.

SSVM の学習は SVM と同様に  $\alpha$  に関する凸 2 次計画問題となり, 計算機によって解くことができる.

$l = 1, 2, \dots, L$  に対して,

$$0 \leq \alpha_l \leq C, \quad (31)$$

$$y^T \alpha = 0, \quad (32)$$

の制約条件のもと,  $\alpha$  に関する二次形式

$$L_D^S = \mathbf{1}^T \alpha - \frac{1}{2} \alpha^T D^T Z^{-1} D \alpha \quad (33)$$

を最大化する.

この2次計画問題を解いて、 $\alpha$  を求めれば、

$$\beta = Z^{-1}D\alpha. \quad (34)$$

から  $w$  が計算できる。

$SV^S = \{x_l | 0 < \alpha_l < C, 1 \leq l \leq L\}$  とおけば、

$$\theta = \frac{1}{|SV^S|} \sum_{x_l \in SV^S} \left[ y_l - \sum_{n=1}^N \beta_n k(z_n, x_l) \right] \quad (35)$$

によって、 $\theta$  を求めることができる。

これにカーネル法を用いた場合、識別関数は

$$d(x) = \sum_{n=1}^N \beta_n k(z_n, x) + \theta. \quad (36)$$

となる。

### 3.2 交差部分空間学習

式 (7) を見ると判るように、SVM の法線ベクトルは学習標本点によって張られる。そして同じ標本点によって法線ベクトル内のパラメータ  $\alpha_k$  が学習される。

SVM のこの特性は汎化性能を悪くする可能性がある。そこで SSVM に交差検定法と同様の手法を用いることを提案する。

交差検定は本来、任意の統計量の推定法として考案されたものであり、次のような手順で行われる。まず標本点の集合を  $K$  個の標本群に分割する。そのうちの1つをテスト事例とし、残る  $K-1$  個を訓練事例とする。その後、 $K$  個に分割した標本群それぞれをテスト事例として  $K$  回検定を行う。 $K$  回の実験を繰り返すことですべてのデータをテスト事例として用いることとなる。このようにして得られた  $K$  回の結果を平均して1つの推定を得る。この方法を  $K$ -fold 交差検定と呼ぶ [4, 5]。

SVM を部分空間制約で拡張したことにより法線ベクトルを構成するデータと学習を行うデータを分離することが可能となった。そこでその SSVM に交差検定のアルゴリズムを組み込む。学習標本点をいくつかの標本群に分割し、そのうちの一部を部分空間を張るために、その他を識別関数のパラメータの学習のために利用する。この手法を交差部分空間学習法と呼ぶことにする。

交差部分空間学習について詳細に説明する。

まず、 $L$  個の標本点  $T = \{x_i\}_{i=1}^L$  を  $S_1, S_2, \dots, S_K$  の  $K$  個の標本群に分割する。

$$T = S_1 \cup S_2 \cup S_3 \cup \dots \cup S_K. \quad (37)$$

$U$  と  $V$  をそれぞれ部分空間を張るための標本点と学習標本点とする。SSVM であるため、 $U, V$  の元は式

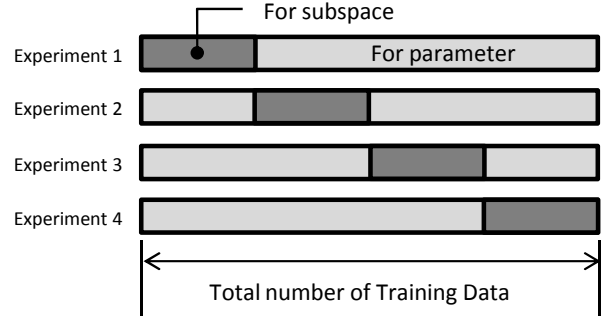


図 2: 交差部分空間学習のデータ分割例 ( $K = 4 (1,3)$ )

(27), (28) の  $z_n$  と  $x_l$  となる。そして凸2次問題を解き、識別関数  $d_{U,V}(x)$  を得る。

$U$  と  $V$  の決め方は複数存在する。 $U$  を  $S_1, S_2, \dots, S_k$  とし、そして残った  $T - S_1, T - S_2, \dots, T - S_k$  を  $V$  とする。するとそれぞれの識別関数を  $K$  個得ることになる。最終的な識別関数  $d(x)$  はそれらを加えて

$$d(x) = d_{S_1, T-S_1}(x) + d_{S_2, T-S_2}(x) + \dots + d_{S_K, T-S_K}(x). \quad (38)$$

となる。図2でこの場合の  $K = 4$  での標本点の分け方を示す。

また、逆の方法も考えることができる。つまり  $T - S_1, T - S_2, \dots, T - S_k$  を  $U$  とし、 $S_1, S_2, \dots, S_k$  を  $V$  とする。この場合も最終的な識別関数  $d(x)$  は

$$d(x) = d_{T-S_1, S_1}(x) + d_{T-S_2, S_2}(x) + \dots + d_{T-S_K, S_K}(x). \quad (39)$$

となる。図3でこの場合の  $K = 4$  での標本点の分け方を示す。

交差検定では学習用とテスト用に標本点を分けていた。実験のすべての段階でテストデータの情報が学習系に入り込まないようにしておかないと実験結果が信頼できるものとならないため、1つのデータを学習にもテストにも利用することはなかった。しかし提案手法は、SVM では元々同じデータを使っていたものを部分空間で制限し分けたものであるため、1つの標本群を  $U$  か  $V$  に決めることなく、両方に使うことも可能である。またパラメータを訓練する標本点の数が不十分であると良い結果が出ない。そこで標本点全てを法線ベクトルのパラメータを訓練するための  $V$  に使い、部分空間はいくつかの標本群で張ることにする。図4でこの場合の  $K = 4$  での標本点の分け方を示す。

## 4 実験

2クラス識別用の13のデータセットで実験を行った。データセットの特性は表1で示す。

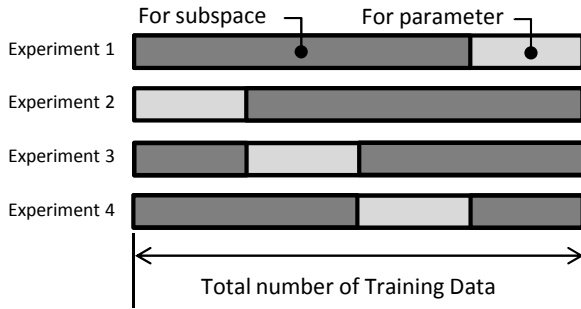


図 3: 交差部分空間学習のデータ分割例 ( $K = 4 (3,1)$ )

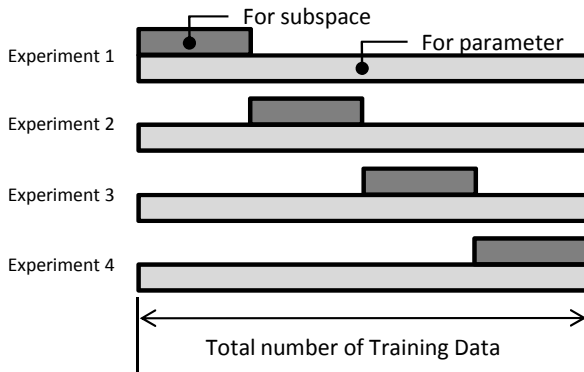


図 4: 交差部分空間学習のデータ分割例 ( $K = 4 (1,4)$ )

制限部分空間を張る標本点  $U$  とパラメータ学習標本点  $V$  のそれぞれの標本群の数を  $(k, l)$  で表すことにする。すなわち  $l = K$  のときは、パラメータの学習に全ての標本点を使っていることになる。通常の交差検定ではテストデータは分割した標本群 1 つであるが交差部分空間学習では  $l = 1, 2, \dots, K$  と変えて実験を行った。

スラック変数のパラメータ  $C$  は 5-fold 交差検定で決定した。カーネル関数のパラメータは [10] の通常の SVM と同じものを用いた。

部分空間用とパラメータ学習用に違う標本点を用いた SSVM の実験結果を表 1 に示す。この表の中の SVM の結果は [8, 9] を参照した。太字は各データセットでの最良の結果である。表 2 から、13 データセットのうち 9 つのデータセットで従来手法の SVM の識別率を上回り、'banana', 'ringnorm', 'thyroid', 'twonorm' では t-検定においても有意水準 5% の有用性が明確に示された。

制限部分空間用の標本群数を変えた場合、最良の結果はほとんど全てのデータセットにおいて  $K = 3, K = 4$  に対してそれぞれ  $(1, 2), (1, 3)$  となっている。これはパラメータ学習のための標本点数が多い方が良かったためであると考えられる。

分割数  $K$  を変えた場合は、ほとんど全てのデータセッ

トで  $K = 3$  もしくは  $K = 4$  の場合で最も認識率が高くなった。 $K = 2$  と  $K = 3$  の場合を比べてみても 13 種類中 11 のデータセットで  $K = 3$  の場合の認識率が高い。しかし  $K = 4$  の結果は  $K = 3$  の場合と比べてもその違いは大きくなく、一概に分割数を増やしていけば認識率が高くなるとは言えない。しかしながら交差部分空間学習する上で 2 分割では足りず、3、もしくはそれ以上に分割したほうが良いと思われる。

パラメータ学習用に全ての標本点を使用した SSVM の実験結果を表 3 に示す。この結果を見ると 13 データセット中 11 種類で SSVM の有用性が示された。制限部分空間を張る標本点の割合の変化で見ると表 2 のときとは違い、あまり変化がない。これはパラメータ学習の標本点数が十分であるからと考えられる。 $K$  に関しては表 2 と同様に  $K = 2$  では不十分と見える。

表 3 と表 2 を比較すると 'germen', 'image', 'titanic' の SVM よりも SSVM の方が誤認識率が高いデータセットに対して、パラメータの学習に標本点をすべて用いた方が良い結果が得られる。SSVM の結果が SVM より優れているとき、表 2 での結果は表 3 での結果よりも誤認識率が低い。従って、交差部分空間学習を用いた SSVM を使う場合  $U$  と  $V$  で同じ標本点を使うことを避けた方が良いと思われる。

'splice' に対しては、全ての場合で SSVM が SVM に優位性が示せなかった。この理由に関して調査する必要がある。

## 5 おわりに

SVM を元にした新たな識別器、SSVM を提案した。この識別器は識別境界の法線ベクトルを予め与えた部分空間に制限している。この SSVM で法線ベクトルを構成する標本点と法線ベクトルのパラメータを学習する標本点に、それぞれ違うものを使用する交差部分空間学習を提案し、計算機実験によって SSVM の有用性を示した。

今後の課題として可変カーネル SVM [3] の過学習の問題を解消するために交差部分空間学習を適用する。可変カーネル SVM は非対称カーネルを使用することで自由にそのカーネルパラメータが選択できる。しかし、自由度が高すぎるために、簡単な問題では性能を向上させることができるが、現実的な問題では過学習のため、十分な結果は得られなかった。そこで交差部分空間学習を使用することで過学習の問題を軽減することが必要である。

DN	Dataset name	# of training patterns	# of test patterns	# of realizations	dimension
1	banana	400	4900	100	2
2	breast cancer	200	77	100	9
3	diabetis	468	300	100	8
4	flare solar	666	400	100	9
5	german	700	300	100	20
6	heart	170	100	100	13
7	image	1300	1010	20	18
8	ringnorm	400	7000	100	20
9	splice	1000	2175	20	60
10	thyroid	140	75	100	5
11	titanic	150	2051	100	3
12	twonorm	400	7000	100	20
13	waveform	400	4600	100	21

表 1: データセット特性

## 参考文献

- [1] C. Cortes and V. Vapnik “Support-vector networks”, Machine Learning, vol.20, no.3, pp.273-297, Sep. 1995.
- [2] V. Vapnik, “Statistical Learning Theory”, Wiley Interscience, New York, 1998.
- [3] 柳森, 山下幸彦, “カーネル理論の拡張と非対称カーネルサポートベクタマシン”, 2004 Workshop on Information-Based Induction Sciences (IBIS2004), Tokyo, Nov. 2004.
- [4] M. Stone, “Cross-Validatory Choice and Assessment of Statistical Predictions”, Journal of the Royal Statistical Society. Series B (Methodological), vol.36, no.2, pp.111-147, 1974.
- [5] S. Geisser, “The predictive sample reuse method with applications”, Journal of the American Statistical Association vol.70, pp.320-328, Jun. 1975.
- [6] D. C. Manning and P. Raghavan and H. Schuetze, “Introduction to Information Retrieval”, Journal of the American Statistical Association chapter 15.2, Cambridge University Press, 2008.
- [7] M. Minsky and S. Papert, “Perceptrons”, MIT Press, Cambridge, 1969.
- [8] G. Rätsch and T. Onoda and K.-R. Müller, “Soft Margins for AdaBoost”, Machine Learning, vol.42, no.3, pp.287-320, Mar. 2001.
- [9] S. Mika and G. Rätsch and J. Weston and B. Schölkopf and K.-R. Müller, “Fisher discriminant analysis with kernels”, Neural Networks for Signal Processing IX pp.41-48, Jun. 1999.
- [10] G. Rätsch and T. Onoda and K.-R. Müller, “Soft Margins for AdaBoost”, Technical Report NC-TR-1998-021, Department of Computer Science, Royal Holloway, University of London, Aug. 1998.

DN	SVM	SSVM, K=2 (1,1)	SSVM $K = 3$ (1,2)	SSVM $K = 3$ (2,1)
1	11.53 ± 0.66	10.51 ± 0.43	<b>10.38 ± 0.43</b>	10.62 ± 0.46
2	26.04 ± 4.74	25.56 ± 4.53	25.61 ± 4.55	26.12 ± 4.89
4	23.53 ± 1.73	<b>23.03 ± 1.72</b>	23.17 ± 1.69	23.53 ± 1.70
3	32.43 ± 1.82	32.43 ± 1.77	<b>32.40 ± 1.78</b>	32.57 ± 1.77
5	<b>23.61 ± 2.07</b>	23.92 ± 2.15	23.86 ± 2.33	23.97 ± 2.35
6	15.95 ± 3.26	15.70 ± 3.34	<b>15.32 ± 3.26</b>	15.60 ± 3.33
7	<b>2.96 ± 0.60</b>	3.63 ± 0.55	3.37 ± 0.43	4.03 ± 0.72
8	1.66 ± 0.12	1.50 ± 0.10	1.47 ± 0.08	1.56 ± 0.10
9	<b>10.88 ± 0.66</b>	12.75 ± 0.59	12.52 ± 0.82	13.10 ± 0.73
10	4.80 ± 2.19	4.41 ± 2.10	<b>4.16 ± 2.04</b>	4.39 ± 2.18
11	<b>22.42 ± 1.02</b>	22.62 ± 1.27	22.62 ± 1.13	23.15 ± 1.05
12	2.96 ± 0.23	2.41 ± 0.16	2.39 ± 0.15	2.39 ± 0.13
13	9.88 ± 0.43	9.91 ± 0.49	9.82 ± 0.37	9.85 ± 0.47

(a) SVM と SSVM ( $K = 2, 3$ )

DN	SSVM $K = 4$ (1,3)	SSVM $K = 4$ (2,2)	SSVM $K = 4$ (3,1)
1	10.39 ± 0.43	10.45 ± 0.45	10.75 ± 0.54
2	<b>25.31 ± 4.39</b>	25.65 ± 4.59	25.97 ± 4.37
4	23.16 ± 1.70	23.13 ± 1.70	23.25 ± 1.70
3	<b>32.40 ± 1.78</b>	32.41 ± 1.76	32.84 ± 1.64
5	23.80 ± 2.28	23.75 ± 2.22	23.97 ± 2.21
6	15.40 ± 3.17	15.67 ± 3.23	15.65 ± 3.18
7	3.13 ± 0.52	3.32 ± 0.57	4.38 ± 0.68
8	<b>1.46 ± 0.08</b>	1.72 ± 0.17	1.58 ± 0.12
9	12.23 ± 0.70	12.29 ± 0.76	13.02 ± 0.76
10	4.25 ± 2.11	4.28 ± 2.06	4.29 ± 1.98
11	22.46 ± 1.12	22.62 ± 1.13	22.83 ± 1.15
12	<b>2.38 ± 0.14</b>	2.39 ± 0.13	2.41 ± 0.13
13	<b>9.81 ± 0.44</b>	9.83 ± 0.42	9.85 ± 0.45

(b) SSVM ( $K = 4$ )

表 2: 実験結果 (部分空間用とパラメータ学習用の標本点が違う場合)

DN	SVM	SSVM, K=2 (1,2)	SSVM $K = 3$ (1,3)	SSVM $K = 3$ (2,3)
1	11.53 ± 0.66	10.44 ± 0.45	10.44 ± 0.45	10.44 ± 0.46
2	26.04 ± 4.74	25.42 ± 4.65	25.53 ± 4.53	<b>25.39 ± 4.64</b>
3	23.53 ± 1.73	23.21 ± 1.84	<b>23.20 ± 1.86</b>	23.22 ± 1.82
4	32.43 ± 1.82	<b>32.40 ± 1.78</b>	<b>32.40 ± 1.79</b>	32.41 ± 1.79
5	23.61 ± 2.07	24.13 ± 2.15	23.69 ± 2.28	<b>23.57 ± 2.15</b>
6	15.95 ± 3.26	15.48 ± 3.24	<b>15.40 ± 3.27</b>	15.44 ± 3.24
7	2.96 ± 0.60	<b>2.95 ± 0.47</b>	2.97 ± 0.53	2.96 ± 0.60
8	<b>1.66 ± 0.12</b>	1.82 ± 0.21	1.86 ± 0.24	1.72 ± 0.16
9	<b>10.88 ± 0.66</b>	11.58 ± 0.72	11.73 ± 0.65	11.29 ± 0.77
10	4.80 ± 2.19	4.75 ± 2.18	<b>4.45 ± 2.21</b>	4.53 ± 2.18
11	22.42 ± 1.02	22.45 ± 1.05	22.37 ± 1.05	22.40 ± 0.99
12	2.96 ± 0.23	2.40 ± 0.14	2.40 ± 0.14	2.40 ± 0.14
13	9.88 ± 0.43	9.80 ± 0.42	9.79 ± 0.43	9.81 ± 0.41

(a) SVM と SSVM ( $K = 2, 3$ )

DN	SSVM $K = 4$ (1,4)	SSVM $K = 4$ (2,4)	SSVM $K = 4$ (3,4)
1	<b>10.43 ± 0.44</b>	10.44 ± 0.46	10.44 ± 0.46
2	25.53 ± 4.47	25.44 ± 4.60	<b>25.39 ± 4.66</b>
3	23.22 ± 1.85	<b>23.20 ± 1.83</b>	23.22 ± 1.82
4	<b>32.40 ± 1.78</b>	32.41 ± 1.78	32.42 ± 1.79
5	23.66 ± 2.32	23.64 ± 2.17	23.58 ± 2.13
6	15.43 ± 3.23	15.43 ± 3.26	15.45 ± 3.28
7	2.98 ± 0.59	2.99 ± 0.51	2.98 ± 0.61
8	1.91 ± 0.26	1.79 ± 0.18	1.68 ± 0.14
9	11.72 ± 0.76	11.32 ± 0.78	11.17 ± 0.72
10	<b>4.45 ± 2.15</b>	4.52 ± 2.19	4.53 ± 2.17
11	22.46 ± 1.12	<b>22.35 ± 1.14</b>	22.39 ± 0.99
12	<b>2.39 ± 0.14</b>	<b>2.39 ± 0.14</b>	2.40 ± 0.14
13	<b>9.77 ± 0.40</b>	9.80 ± 0.42	9.83 ± 0.39

(b) SSVM ( $K = 4$ )

表 3: 実験結果 (部分空間用とパラメータ学習用の標本点と同じ場合)