



部分空間法に基づくサポートベクトルマシンに関する研究

北村, 拓也

(Degree)

博士 (工学)

(Date of Degree)

2011-03-25

(Date of Publication)

2015-11-27

(Resource Type)

doctoral thesis

(Report Number)

甲5256

(URL)

<https://hdl.handle.net/20.500.14094/D1005256>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



博士論文

部分空間法に基づくサポートベクトル
マシンに関する研究

平成23年1月

神戸大学大学院工学研究科

北村 拓也

概要

パターン認識において代表的な識別器として部分空間法やサポートベクトルマシン (SVM: Support Vector Machine) などがある。部分空間法はクラスごとに部分空間を生成することで情報量を大幅に削減し、同時に識別処理を行える。しかしながら、識別性能が極めて低い。SVM ではマージン最大化を基準に識別ルールを決定することで識別性能が極めて高いが、識別部がブラックボックスである。これらのため、近年、部分空間法と SVM の改良に関する研究が活発になされている。

本論文では、部分空間法に SVM のマージン最大化の概念を取り入れた部分空間法に基づく SVM (SS-SVM: Subspace-Based Support Vector Machine) と SS-SVM の改良手法を提案する。SS-SVM はそれらの部分空間を構成する軸に対して、それぞれクラスの部分空間の分離、すなわちマージンが最大となるような重みを決定し、重み付けすることにより、識別性能を向上させる。これにより、情報量を大幅に削減でき、識別性能が高く、解析可能な識別手法といえる。また、SS-SVM に対して以下の3つの観点から改良を加える。

- (1) SS-SVM により識別の観点から最適化された重みを用いた特徴選択によるさらなる識別性能の向上
- (2) 部分空間生成における解くべき問題を縮小することによる学習コストの削減
- (3) モデル選択に要する学習コストの削減

第1章では、研究背景、研究目的、本論文の構成を述べる。第2章では、従来手法である部分空間法と SVM の理論の概略と SVM の多クラス問題への拡張方式、またそれに広く用いられているカーネル法と代表的なモデル選択である k 分割交差検定法について述べる。

第3章では SS-SVM を提案する。ここで、一次の連立方程式を解く最小自乗サポートベクトルマシン (LS-SVM: Least Squares Support Vector Machine) を適用した部分空間法に基づく LS-SVM (SSLS-SVM: Subspace-Based Least Squares Support Vector Machine)、線形計画問題を解く線形計画サポートベクトルマシン (LP-SVM: Linear Programming Support Vector Machine) を適用した部分空間法に基づく LP-SVM (SSLP-SVM: Subspace-Based Linear Programming Support Vector Machine)、アクティブ集合を用いた学習法が適用可能である L2 サポートベクトルマシン (L2-SVM: L2 Support Vector Machine) を適用した部分空間法に基づく L2-SVM (SSL2-SVM: Subspace-Based L2 Support Vector Machine) の3つのモデルを提案する。SSLS-SVM では部分空間の次元数が少なければ学習コストが非常に小さくなる。SSLP-SVM では

識別と同時に特徴選択が行える。SSL2-SVM ではサポートベクトル数が少なければ学習コストが非常に小さくなる。計算機実験により、通常の部分空間と SVM と比較することで SS-SVM の有効性を示し、また、それら 3 つのモデル同士で比較・評価を行う。最後に、SS-SVM の識別部の解析を行う。

第 4 章では、SS-SVM における負の重みをもつ軸の識別への影響の確認と SS-SVM における識別性能をさらに向上させるために重みに基づく特徴選択手法を提案する。SS-SVM では通常の部分空間法とは異なり、重みの値が負となりうるため、非負制約を設けた SS-SVM と比較することによりそれらの軸が識別において重要となりうることを確認する。また、SS-SVM において各クラスの部分空間を構成する軸の選択を他のクラスの分離を考慮して決定された重みの値に基づいて行う。これにより識別性能を向上させる。計算機実験では、はじめに非負制約を設ける場合と設けない場合の SS-SVM を比較することにより負の重みをもつ軸が重要となりうることを確認する。また、重みの絶対値に基づく軸の選択処理を行う SS-SVM を通常の部分空間法と SS-SVM と SVM との識別性能を比較することにより、有効性を示す。その際に軸がどれほど削減されているかの確認と最適化された重みと固有値の関連性の解析を行う。

第 5 章では、SS-SVM における学習コスト削減のため、標本特徴空間の縮小と標本部分空間法に基づく SVM (ES-SVM: Empirical Subspace-Based Support Vector Machine) を提案する。1 つ目の手法では、SS-SVM における部分空間生成の前処理としてあらかじめ写像する標本特徴空間の生成をクラスごとに別々に行う。これにより、その次元数を削減し、部分空間生成の際に用いるカーネル主成分分析 (KPCA: Kernel Principal Component Analysis) の学習コストを大幅に削減できる。2 つ目の手法では、各クラスの教師ベクトルそのものを部分空間を構成する軸として取り扱うことにより KPCA の処理を省き、学習コストを削減する。計算機実験により、これらの学習の高速化を示し、また識別性能も劣化しないことを示すことにより有効性を示す。

第 6 章では、複数のカーネルパラメータにおける部分空間を考慮することによる部分空間法の識別性能を向上させ、SS-SVM と比べてモデル選択に要する学習コストを削減できる多重非線形部分空間法に基づく SVM (MNS-SVM: Multiple Nonlinear Subspace-Based Support Vector Machine) を提案する。MNS-SVM では、あらかじめ五分割交差検定法によりカーネルパラメータをひとつ決定するのではなく、複数のカーネルパラメータを選択する。次にそれぞれのカーネルパラメータにおけるクラス部分空間を生成する。そして、それらを構成する軸に対して重み付けを行うのではなく、それらの部分空間への射影長をひとつの特徴量とし、カーネルパラメータ数個の特徴量に対して SVM により重み付けを行うことにより従来の KSM に比べ汎化能力を向上させる。これにより、複数のカーネルパラメータにおける部分空間を考慮でき、どれかひとつのみを選択する必要がない。さらに、あらかじめ選択すべきパラメータの

中からカーネルパラメータを削除できるため、モデル選択における学習コストも大幅に削減できる。計算機実験により、通常の部分空間法と識別性能の比較、また SS-SVM とのモデル選択における学習時間の SS-SVM との比較を行うことにより MNS-SVM の有効性を示す。

第 7 章では、本論文のまとめと今後の課題を述べる。

目次

	4
第 1 章 緒論	1
1.1 研究の背景	1
1.2 研究の目的	3
1.3 本論文の構成	4
	114
第 2 章 パターン認識における識別手法	7
2.1 緒言	7
2.2 カーネル法	7
2.2.1 線形カーネル	8
2.2.2 多項式カーネル	9
2.2.3 ラジアル基底関数カーネル	9
2.3 部分空間法	9
2.3.1 線形問題における部分空間の生成方法	9
2.3.2 非線形問題における部分空間の生成方法	11
2.3.3 類似度の定義	14
2.3.4 カーネル部分空間法による識別	14
2.4 サポートベクトルマシン	16
2.4.1 ハードマージン SVM	16
2.4.2 ソフトマージン SVM	16
2.4.3 最小自乗サポートベクトルマシン	19
2.4.4 線形計画サポートベクトルマシン	20
2.4.5 SVM の多クラス問題への拡張	21
2.5 モデル選択	23
2.6 結言	24

第 3 章	部分空間法に基づくサポートベクトルマシン	25
3.1	緒言	25
3.2	部分空間法の問題点と解決方法	25
3.3	部分空間法に基づく最小自乗サポートベクトルマシン	27
3.3.1	同時定式化方式を適用した SSLS-SVM	28
3.3.2	一対他方式を適用した SSLS-SVM	30
3.4	部分空間法に基づく線形計画サポートベクトルマシン	32
3.4.1	同時定式化方式を適用した SSLP-SVM	32
3.4.2	一対他方式を適用した SSLP-SVM	34
3.5	部分空間法に基づく L2 サポートベクトルマシン	35
3.5.1	同時定式化方式を適用した SSL2-SVM	35
3.5.2	一対他方式を適用した SSL2-SVM	38
3.6	計算機実験	40
3.6.1	計算機実験における環境	41
3.6.2	KSM のパラメータ設定	42
3.6.3	SSLs-SVM の評価 (KSM との比較)	42
3.6.4	独自でモデル選択を行った SSLs-SVM の評価	47
3.6.5	SSLP-SVM の評価 (KSM との比較)	51
3.6.6	独自でモデル選択を行った SSLP-SVM の評価	55
3.6.7	SSL2-SVM の評価 (KSM との比較)	59
3.6.8	SS-SVM における各モデルの性能比較	62
3.6.9	識別部の解析	65
3.7	結言	66
第 4 章	重みに基づく特徴選択法	67
4.1	緒言	67
4.2	重みに基づいた軸の選択	68
4.2.1	非負制約を設けた SSLs-SVM	68
4.2.2	SSLP-SVM における負の重みを許す最適化問題	70
4.2.3	重みの絶対値による選択処理	70
4.3	計算機実験	72
4.3.1	識別における負の重みをもつ軸の重要性の評価	72
4.3.2	重みの絶対値に基づいた選択手法の評価	75
4.3.3	部分空間を構成する各軸に対応する固有値と重みの関係	77

目次	iii
4.3.4 特徴選択による軸の削減割合	77
4.3.5 正規直交系の標本特徴空間上での識別との評価	78
4.3.6 KSM と SVM との比較・評価	79
4.4 結言	81
第 5 章 学習の高速化	83
5.1 緒言	83
5.2 クラス標本特徴空間の生成による学習の高速化	84
5.3 錐制約部分空間法	86
5.4 標本部分空間法に基づくサポートベクトルマシン	88
5.5 計算機実験	89
5.5.1 クラス標本特徴空間による高速化の評価	90
5.5.2 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	91
5.5.3 ES-SVM の学習時間の比較・評価	93
5.6 結言	95
	114
第 6 章 多重非線形部分空間法の導入	97
6.1 緒言	97
6.2 多重非線形部分空間法に基づくサポートベクトルマシン	98
6.3 計算機実験	101
6.3.1 パラメータ設定	101
6.3.2 汎化能力の評価	101
6.3.3 学習時間の評価	102
6.4 結言	104
第 7 章 結論	105
謝辞	109
	114
参考文献	111
	114
関連発表論文	115

図目次

2.1	高次元特徴空間への写像	8
2.2	(a) 通常の PCA による特徴抽出 (b) 部分空間法における PCA による 特徴抽出	10
2.3	標本特徴空間への写像	12
2.4	部分空間法による識別	15
2.5	ハードマージン SVM による識別	17
2.6	ソフトマージン SVM による識別	18
2.7	一対他方式を用いた SVM による識別処理	22
2.8	五分割交差検定法	24
3.1	通常の部分空間法による誤認識	26
3.2	重みの最適化	27
3.3	各軸に対する類似度	65
4.1	負の重みと軸の関係	69
4.2	重みの絶対値と軸の関係	71
4.3	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	77
5.1	標本特徴空間の縮小	85
5.2	CRSM による識別	87
5.3	CRSM の問題点	88
6.1	多重非線形部分空間法	99

表目次

3.1	2クラスベンチマークデータセット	41
3.2	多クラスベンチマークデータ	41
3.3	KSM (1) と KSM (E) におけるパラメータ値	42
3.4	パラメータ値 (SSLS-SVM)	43
3.5	バリデーションセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (1) の比較)	44
3.6	テストデータセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (1) の比較)	45
3.7	バリデーションセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (E) の比較)	46
3.8	テストデータセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (E) の比較)	47
3.9	パラメータ値 (独自にモデル選択を行った SSLS-SVM)	48
3.10	テストデータセットにおける平均認識率 (%) と標準偏差 (独自にモデル選択を行った SSLS-SVM)	49
3.11	SSLS-SVM における学習時間 (s)	50
3.12	パラメータ値 (SSLP-SVM)	51
3.13	バリデーションセットを用いたときの平均認識率 (%) と標準偏差 (SSLP-SVM と KSM (1) の比較)	52
3.14	テストデータセットを用いたときの平均認識率 (%) と標準偏差 (SSLP-SVM と KSM (1) の比較)	53
3.15	バリデーションセットを用いたときの平均認識率 (%) と標準偏差 (SSLP-SVM と KSM (E) の比較)	54
3.16	テストデータセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (E) の比較)	55
3.17	パラメータ値 (独自でモデル選択を行った SSLP-SVM)	56
3.18	テストデータに対する平均認識率 (%) 及び標準偏差	57
3.19	特徴選択による特徴量の削減数 (SSLP-SVM)	58
3.20	SSLP-SVM における学習時間 (s)	59
3.21	パラメータ値 (SSL2-SVM)	60

3.22	テストデータセットにおける平均認識率 (%) と標準偏差 (SSL2-SVM と KSM の比較)	61
3.23	SSL2-SVM における学習時間 (s)	62
3.24	テストデータセットを用いたときの平均認識率 (%) と標準偏差 (各 SS-SVM の比較)	64
4.1	SSLS-SVM (P) と SSLP-SVM (N) におけるパラメータ値	74
4.2	テストデータにおける認識率 (%) と標準偏差の比較 (負の重みの重要性の評価)	74
4.3	SSLS-SVM (Ab) におけるパラメータ値	75
4.4	テストデータにおける認識率 (%) と標準偏差の比較 (SSLS-SVM (P) と SSLS-SVM (Ab) の比較)	76
4.5	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 (%)	78
4.6	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	79
4.7	テストデータセットにおける平均認識率 (%) と標準偏差 (KSM と SVM との比較)	81
5.1	学習時間 (s) の比較	91
5.2	ES-SVM におけるパラメータ	92
5.3	ES-SVM のテストデータにおける平均認識率 (%) と標準偏差	93
5.4	SS-SVM と ES-SVM の学習時間 (s) の比較	94
6.1	MNSLS-SVM のパラメータ値	102
6.2	テストデータにおける平均認識率 (%) と標準偏差の比較 (MNSLS-SVM)	103
6.3	五分割交差検定法の学習時間 (s) の比較 (SSLS-SVM (Ab) と MNSLS-SVM)	103

第1章

緒論

1.1 研究の背景

近年の科学技術の発展に伴い、従来、人間が処理してきた動作をロボットやコンピュータなどに行わせる人工知能の分野が非常に注目されている。その中でもパターン認識¹⁾は人工知能において、最も重要視される技術のひとつとして挙げられる。通常パターン認識とは、ある未知の入力に対して、あるひとつのクラスに分類する処理をよぶ。例として、カメラから入力された画像からその画像に写されているものは何なのかを特定するなどの画像認識や声を入力として与え、誰の声であるかを特定する話者認識などもそれぞれパターン認識のひとつである。これらのような技術は一般的に目にする防犯カメラなどによるセキュリティや個人認証技術だけでなく、自動車における標識認識や産業における製品の検査技術²⁾や医療における診断など一般の人間の身近な所から産業や医療現場など幅広く用いられている。このため、パターン認識は人工知能技術の発展には必要不可欠なものである。それらのような処理を高精度に行うためには識別処理部分における理論が非常に重要となっているのは明らかである。ここにおける理論とは、パターン認識の処理を行う際に、あらかじめクラスの与えられているサンプルをシステムに記憶させ、システムに識別ルールを学習させるための学習理論を指す。現在、パターン認識において、世界的に広く用いられている識別器として部分空間法^{3, 4)}やサポートベクトルマシン (SVM: Support Vector Machine)^{5, 6, 7)}やニューラルネットワーク^{1, 8)}などがある。現在までにこれらを基に様々な学習理論の研究がなされており、非常に高い精度を持った識別器が多く存在している。しかしながら、日々発展を続ける科学技術に対応させるため、現在においても新しい学習理論の提案は非常に重要となる。そこで本研究では、これらの識別器の中で、SVMと部分空間法に着目した。

通常、画像認識などでは画像のような情報の非常に多いサンプルを学習する必要がある。以下、これらの情報を特徴量とよぶ。これらの特徴量をすべて用いて学習すると計算機に非常に大きい負荷がかかるため、実用的ではない。そこで、学習サンプルからそれらを表現するため、または、よりよい識別するための必要最低限の特徴量を抽出または選択する。これらの処理をそれぞれ特徴抽出、特徴選択⁹⁾とよぶ。パターン認識では通常、特徴抽出・選択処理と識別処理を区分して行う。前述の部分空間法

はこれらを同時に処理する特殊な識別器である。また、30年以上前から文字認識や顔認識などの様々な応用技術に適用され続けている非常に実用的な日本発の技術であり、パターン認識において極めて重要な理論である。通常の部分空間法では、まず、クラスごとに主成分分析 (PCA: Principal Component Analysis)^{10, 11)} を用いて特徴抽出・選択を行い、部分空間を生成する。通常の特徴抽出・選択とは異なり、全学習サンプルではなくクラスごとに行うため、非常に低次元の部分空間を生成できる。次に、未知のサンプルが入力された際、各クラスの部分空間に射影することにより射影長を求め、その射影長を基に類似度を決定する。これにより、クラスごとに PCA を用いるのみで識別ルールを構成できる非常に単純なモデルとなっている。また、通常のパターン認識における実問題では入力空間において線形分離可能な問題はほとんどなく、非線形問題への適用を考えなければならない。このように線形問題だけでなく非線形問題に対しても適用するためにカーネル法¹²⁾ を適用したカーネル主成分分析 (KPCA: Kernel Principal Component Analysis)^{10, 11)} 用いたカーネル部分空間法 (KSM: Kernel Subspace Method)^{13, 14)} が提案されている。これらの部分空間法を改良した手法として、学習部分空間法^{3, 4)}、直交部分空間法 (OSM: Orthogonal Subspace Method)⁴⁾、相互部分空間法 (MSM: Mutual Subspace Method)¹⁵⁾、カーネル直交相互部分空間法 (KOMSM: Kernel Orthogonal Mutual Subspace Method)¹⁶⁾、カーネル制約相互部分空間法 (KCMSM: Kernel Constrained Mutual Subspace Method)¹⁷⁾、部分空間生成に PCA、および KPCA を用いなくてすむ錐制約部分空間法 (CRSM: Cone Restricted Subspace Method)¹⁸⁾ などが提案されており、現在でも部分空間法に関する様々な研究が日本を中心に活発に行われている。しかしながら、通常の部分空間法では、クラスごとに独立して識別ルールを決定するために識別を考慮しておらず、識別性能が SVM やニューラルネットワークなどの識別能力の高い識別器に比べて劣るといった問題点をもつ。

前述にあるように SVM は識別能力が非常に高く、近年、SVM に関する研究は世界的に活発的に行われている。SVM では、クラス間における分離を大きくするマージン最大化の概念の基で特徴空間を構成する軸に対して重み付けを行い、分離超平面を決定する。また、SVM はカーネル法を用いることに適した識別器のひとつである。SVM は大きくふたつに分類できる。1 つ目は完全に線形分離可能な問題のみを考えるハードマージン SVM、2 つ目はある程度の誤認識を許すソフトマージン SVM である。通常、SVM とはソフトマージン SVM を指し、本論文においても同様とする。SVM の代表的なモデルとして、L1 サポートベクトルマシン (L1-SVM: L1 Support Vector Machine)、L2 サポートベクトルマシン (L2-SVM: L2 Support Vector Machine)^{7, 19, 20)}、最小自乗サポートベクトルマシン (LS-SVM: Least Squares Support Vector Machine)^{7, 21)}、線形計画サポートベクトルマシン (LP-SVM: Linear Programming Support Vector

Machine)^{7, 22)} などがある。これらの手法はカーネルトリックを用いることにより識別の基準となる決定関数を求めて識別を行う。しかしながら、識別部の詳細が不明確でブラックボックスとなる。そのため識別において何が良くて、何が悪かったなどの解析が行いにくいという問題点がある。

1.2 研究の目的

本研究では、画像認識や音声認識などの特定した識別対象を考慮するのではなく、どのような識別対象に対しても安定して高い識別性能をもち、かつ識別における解析が容易となる、すなわち識別部がブラックボックスでない識別手法の提案を目的としている。そのために、前述の部分空間法のように識別部がブラックボックスではなく、かつ取り扱う空間の次元が低い点と SVM のように非常に高い識別能力をもつ点に着目し、それぞれの利点を活かした以下の新しい手法を提案する。

- (1) 部分空間法に基づくサポートベクトルマシン (SS-SVM: Subspace-Based Support Vector Machine)^{23, 24)}
- (2) クラスごとの部分空間生成における特徴選択手法の改良^{25, 26)}
- (3) SS-SVM の高速学習手法^{25, 26)}
- (4) 多重非線形部分空間法に基づくサポートベクトルマシン (MNS-SVM: Multiple Nonlinear Subspace-Based Support Vector Machine)

(1) では、部分空間法と同様にクラスごとに特徴抽出・選択処理を行い、部分空間を生成し、部分空間を構成する各軸に対して SVM によりマージン最大化を基準に重み付けを行う。すなわち他のクラスを考慮して重み付けを行うことにより識別の観点からよりよい識別ルールを決定できる。これにより識別部が解析可能かつ部分空間法に比べて識別能力の向上が期待できる。非線形問題に対応させるために KSM を用い、重みの最適化に用いる SVM のモデルは LS-SVM, LP-SVM, L2-SVM とし、それぞれのモデルにおいて評価する。特に部分空間法に基づく L2-SVM (SSL2-SVM: Subspace-Based L2 Support Vector Machine) では、計算コストの削減のためにアクティブ集合を用いた学習法^{19, 20)} を適用する。(2) では、特徴選択部において KPCA の処理によって得られた基準ではなく、SVM により求められた重みを基に特徴選択²⁸⁾ を行う。これにより識別部の結果から識別の観点から重要となりうる特徴量のみを選択することができるため識別能力の向上が期待できる。(3) では、SS-SVM の特徴抽出部において KPCA の処理を単純化することによる高速学習手法と KPCA の処理を行わずに特徴抽出することにより高速化する標本部分空間法に基づくサポートベクトルマシン (ES-SVM: Empirical Subspace-Based Support Vector Machine) である。(4) では、複数のカーネルパラメータにおけるそれぞれの部分空間を考慮し、それぞれの部分空間全体に対

して重み付けに SVM を用いて決定する。これらの手法を様々な識別対象を用いて計算機実験により、従来の手法との比較・評価を行い、解析可能かつ高精度の識別処理のできる識別手法であるか確認する。

1.3 本論文の構成

本論文は、7章から構成される。第1章では、研究の背景と目的を述べる。

第2章では、パターン認識における既知の識別手法について説明する。ここで、既知の識別手法として部分空間法と SVM を取り上げる。まず、様々な非線形問題に用いられるカーネル法についての理論を説明する。次に部分空間法における部分空間生成法と類似度の定義と識別方法を述べる。次に SVM を L1-SVM, L2-SVM, LS-SVM, LP-SVM に分けて説明し、それぞれの最適化問題の解き方などの詳細を述べる。またそれぞれの手法における問題点を結論として述べる。

第3章では、まず第2章で述べた部分空間法と SVM についての問題点を明確にし、それらの問題点を解決するために SS-SVM を提案する。次に異なるモデルの SVM を用いたときのそれぞれの SS-SVM における性能の比較・評価を行う。SVM のモデルとして連立一次方程式を解くことにより最適解を求めることのできる LS-SVM, 線形計画問題を解くことにより最適解を求めることができ、さらに識別と同時に特徴選択を行える LP-SVM, アクティブ集合を用いた学習法を適用できる L2-SVM を用いる。次に計算機実験により、各 SS-SVM の識別精度を KSM と比較することで、本手法の有効性を示す。また各 SS-SVM においても比較を行い、どのモデルにおいて最も有効性を示すかを検証する。さらに、各軸における類似度と重みをそれぞれ明確にすることにより、識別部の解析を行う。このように識別部を解析できるポイントが SVM に対する SS-SVM の大きな利点である。最後にこれらの手法についてのまとめを述べる。

第4章では、通常の部分空間法や第2章における SS-SVM における部分空間生成の際の特徴選択手法における問題点を明確にし、それらを解消するための軸の重みを用いた特徴選択手法を提案する。SS-SVM により決定された重みと対応する軸の識別における関連性について述べ、重みに基づく特徴選択手法の適用した際の学習アルゴリズムなどをそれぞれのモデルを分けて説明する。複数のベンチマークデータセットを用いた計算機実験により、重みに基づく特徴選択手法の有効性を示す。また、従来の特徴選択手法によって削除されていた特徴と重みに基づく特徴選択手法によって削除されていた特徴を解析・比較することにより、本手法が有効であることを検証する。最後にこれらの手法についてのまとめを述べる。

第5章では、SS-SVM における学習の高速化手法について述べる。まず、部分空間生成の際に必要な固有値問題のサイズを小さくすることで高速化する手法につ

いて述べる。すなわち、固有値は標本特徴空間の次元数に依存することを利用し、部分空間だけでなく標本特徴空間もクラスごとに生成することにより標本特徴空間の次元数を削減し、同時に学習コストも削減できる高速化手法である。また、固有値を行わずにクラスごとの一次独立な教師ベクトル自身を部分空間を構成する軸とみなすことによって学習コストの削減を図る標本部分空間法について述べ、標本部分空間法をSS-SVMに取り入れたES-SVMについて述べる。計算機実験により、各高速化手法を通常のKSMとSS-SVMと比較することにより有効性を評価する。最後にこれらの手法についてのまとめを述べる。

第6章では、複数のカーネルパラメータによる部分空間をすべて考慮することのできるKNS-SVMを提案する。すなわち、カーネルパラメータを五分割交差検定法により1つ選択するのではなく、複数のカーネルパラメータを用いてそれぞれに対してクラスごとに部分空間を生成する。入力のそれぞれのカーネルパラメータを用いて生成した部分空間への射影長をひとつの特徴量とする。それらに対してSVMにより重み付けすることにより複数の部分空間を考慮することができる。これにより、複数のカーネルパラメータにおける部分空間を考慮できるためKSMと比べ識別性能の向上が期待でき、さらにカーネルパラメータを決定する必要がないためモデル選択における学習コストをSS-SVMに比べ大幅に削減できる。計算機実験により、MNS-SVMを通常のKSMとSS-SVMとES-SVMと比較することで有効性を評価する。最後にこれらの手法についてのまとめを述べる。

第7章では、本論文における提案された手法やその考察などについてのまとめと今後の課題を述べる。

第2章

パターン認識における識別手法

2.1 緒言

パターン認識¹⁾は未知の入力に対してあるひとつのクラスに分類する処理であり、画像認識や音声認識など様々な分野に近年広く用いられている。そのため、パターン認識における様々な学習理論が提案されている。

本章では、まず非線形のパターン認識問題に対応させるためのカーネル法について述べる。次に、代表的な識別手法のひとつである部分空間法^{3, 4)}について説明する。このとき、主成分分析 (PCA: Principal Component Analysis)¹⁰⁾ を適用した部分空間法とカーネル主成分分析 (KPCA: Kernel Principal Component Analysis)^{10, 11)} を適用したカーネル部分空間法 (KSM: Kernel Subspace Method)^{13, 14)} と分けて説明する。次に、代表的な識別手法のひとつであるサポートベクトルマシン (SVM: Support Vector Machine)^{5, 6, 7)} について説明する。このとき、代表的な SVM のモデルである L1 サポートベクトルマシン (L1-SVM: L1 Support Vector Machine)^{5, 6, 7)}, L2 サポートベクトルマシン (L2-SVM: L2 Support Vector Machine)^{7, 19, 20)}, 最小自乗サポートベクトルマシン (LS-SVM: Least Squares Support Vector Machine)^{7, 21)}, 線形計画サポートベクトルマシン (LP-SVM: Linear Programming Support Vector Machine)^{7, 22)} に分けて説明する。次に、識別性能を大きく左右するモデル選択の代表的な手法のひとつである k -分割交差検定法について述べる。

2.2 カーネル法

パターン認識における実問題のほとんどは非線形問題であり、入力空間上で線形分離不可能となる。そこで、図 2.1 のように入力空間で与えられるデータ \mathbf{x} を線形分離度を高めるためにある非線形の写像関数 $g(\mathbf{x})$ を用いて高次元特徴空間に写像する。図 2.1 では 2 次元 2 クラス問題において 2 次元の入力空間から l 次元の高次元特徴空間へ写像しており、 x_i ($i = 1, 2$) と $g_i(\mathbf{x})$ ($i = 1, \dots, l$) はそれぞれ入力空間と高次元特徴空間の要素を示している。また \mathbf{x}_1 と \mathbf{x}_2 はそれぞれクラス 1 のデータ、クラス 2 のデータを示す。このように、入力空間では線形分離不可能なデータが高次元特徴空間上では分離性が高まり線形分離可能となる。

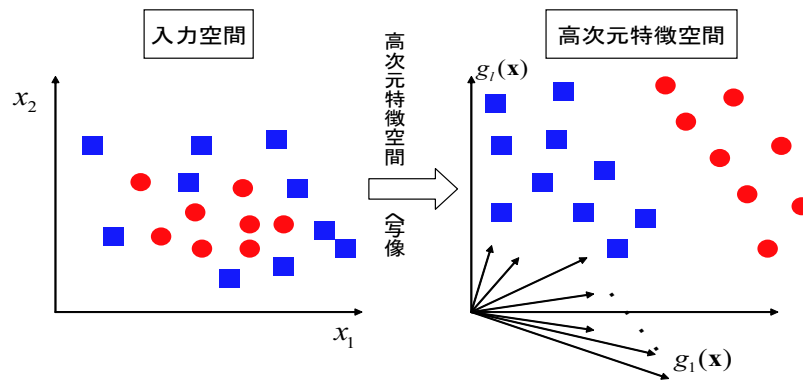


図 2.1 高次元特徴空間への写像

この高次元特徴空間上で線形問題を解くことは、入力空間において非線形の問題を解くことと等価である。一般に、非線形写像によって変換した特徴空間の次元は非常に大きくなるため、高次元特徴空間を取り扱うためには膨大な計算コストがかかってしまう。しかしながら、写像関数同士の内積はカーネルを用いることにより求めることができる。つまり、 \mathbf{x} を写像関数 $g(\mathbf{x})$ により非線形写像した写像空間での 2 つの要素 $g(\mathbf{x})$ と $g(\mathbf{x}')$ の内積を

$$H(\mathbf{x}, \mathbf{x}') = \mathbf{g}^T(\mathbf{x}) \mathbf{g}(\mathbf{x}') \quad (2.1)$$

と定義して、その値を入力データ \mathbf{x} と \mathbf{x}' とすれば、非線形写像により写像された特徴空間での特徴ベクトル $g(\mathbf{x})$ や $g(\mathbf{x}')$ を計算する過程を省き、関数 $H(\mathbf{x}, \mathbf{x}')$ の計算のみで特徴空間を扱うことができる。ここで、関数 H はカーネルと呼んでいる。このように高次元に写像しながら、高次元特徴空間での計算を避けカーネルの計算のみですませるテクニックのことをカーネルトリックという。

以下に、一般的に用いられるカーネルについて述べる。

2.2.1 線形カーネル

分類問題が入力空間において線形分離可能であるとき、入力空間を高次元空間に写像する必要は無く、その場合、次のような線形カーネルが用いられる。

$$H(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}' \quad (2.2)$$

この場合は、通常の内積と同じ形となる。

2.2.2 多項式カーネル

次数 d の多項式カーネルは

$$H(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^d \quad (2.3)$$

で与えられる。ここで、1 は次元 d 以下に含まれる全ての項を表すために加えられる。 $d = 1$ の場合、定数項のある線形カーネルを用いることと等しくなる。

入力変数の数が多くなると、カーネル関数の値は非常に大きく、または非常に小さくなってしまいうため、学習が困難となる。そこで、カーネル関数の正規化を行う。入力変数の数を m とすると、次数 d の多項式カーネル関数において、入力範囲 $[0,1]$ における最大値は $(m + 1)^d$ であるから、以下のように正規化を行う。

$$H(\mathbf{x}, \mathbf{x}') = \frac{(\mathbf{x}^T \mathbf{x}' + 1)^d}{(m + 1)^d} \quad (2.4)$$

2.2.3 ラジアル基底関数カーネル

ラジアル基底関数 (RBF: Radial Basis Function) カーネルは

$$H(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \quad (2.5)$$

で与えられる。ただし、 γ は傾きを制御する正の変数である。

RBF カーネル関数についても、多項式カーネル関数同様正規化を行う。入力範囲 $[0,1]$ における $\|\mathbf{x} - \mathbf{x}'\|^2$ の最大値は m であるから、以下のように正規化を行う。

$$H(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\gamma}{m} \|\mathbf{x} - \mathbf{x}'\|^2\right) \quad (2.6)$$

次節にこれらのカーネルを用いて非線形問題に拡張した KSM によるクラス部分空間の生成を述べる。

2.3 部分空間法

2.3.1 線形問題における部分空間の生成方法

通常、特徴量を多く持つサンプルにおいて、多くの冗長な特徴量が存在する可能性が高い。このため、これらのサンプルを用いて必要最低限の特徴量を抽出する特徴抽出法が前処理として広く用いられている。代表的な特徴抽出法のひとつである PCA では、与えられたサンプルの分散が大きくなるような、すなわち、それらのサンプルを最もよく表現できるような特徴量を順番に抽出する。このとき、サンプルの分布を表現するのに必要最低限の空間を生成することができる。

通常の部分空間法においても、各クラスの部分空間の生成には PCA が用いられるが、特徴量の求め方が通常の PCA と少々異なる。例として、図 2.2 に 2 次元入力空間から通常の PCA を用いた特徴抽出と部分空間法における PCA を用いた特徴抽出によって 1 次元部分空間を生成する様子を示す。通常の PCA では、図 2.2(a) のように全教師データの平均ベクトルに原点を移動し、写像したときのデータの分散が最大となるベクトルを 1 次元部分空間を構成する軸とする。これにより、そのデータの分布を最もよく表す特徴を抽出できる。また、ここでは 1 次元部分空間を生成しているが、2 次元以上の部分空間を生成する場合、2 つ目以降の抽出する軸はそれまでに生成された部分空間に対して直交する空間上において写像したときの分散が最大となる軸をとる。この処理を繰り返すことにより部分空間を生成する。一方、部分空間法では、図 2.2(b) のようにクラスごとに特徴抽出処理を行う。また、原点移動せず、対応するクラスのデータを写像することにより生ずる誤差 (矢印) の自乗平均が最小となるベクトルを 1 次元部分空間を構成する軸とする。また 2 次元以上の部分空間を生成する場合、同様の処理をすでに抽出された軸と直交した空間上で繰り返し行う。すなわち、通常の PCA の処理と部分空間法における PCA の処理との違いは、クラスごとに部分空間を生成する点と原点移動しない点のみである。

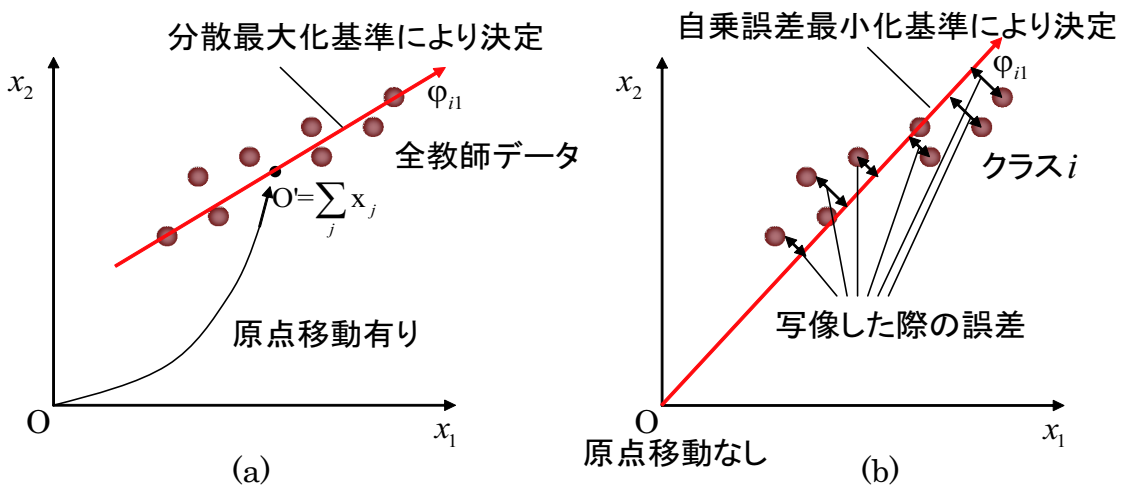


図 2.2 (a) 通常の PCA による特徴抽出 (b) 部分空間法における PCA による特徴抽出

データ数が M でクラス数が n の m 次元問題において、クラス i ($i = 1, \dots, n$) に属する教師データを \mathbf{x}_j ($j = 1, \dots, M$) とするとき、部分空間法において、各クラスの r_i 次元部分空間を構成する軸 φ_{ik} ($k = 1, \dots, r_i$) は次式の固有値問題を解くことにより得られる。

$$\frac{1}{|X_i|} \sum_{j \in X_i} \mathbf{x}_j \mathbf{x}_j^T \varphi_{ik} = \lambda_{ik} \varphi_{ik} \quad \text{for } i = 1, \dots, n, k = 1, \dots, r_i \quad (2.7)$$

ここで, X_i はクラス i に属する教師データ \mathbf{x}_j の添え字集合, $|X_i|$ はクラス i における教師データ数, λ_{ik} は軸 φ_{ik} に対応する固有値である. また, 固有値は $\lambda_{i1} > \dots > \lambda_{ir_i}$ となるように並べ替える.

次に, 部分空間を構成する r_i 個の軸の選択方法を考える. これらの軸を選択するため, まず以下に示す累積寄与率 $a(r_i)$ を定義する.

$$a(r_i) = \frac{\sum_{j=1}^{r_i} \lambda_{ij}}{\sum_{j=1}^m \lambda_{ij}} \times 100 (\%) \quad (2.8)$$

寄与率 $a(r_i)$ は, 選択した r_i 個の固有ベクトルにより構成される空間が原空間と比べ, どの程度の情報量を維持しているかを示す指標となる. あらかじめ累積寄与率の閾値 κ を設定し, $a(r_i - 1) < \kappa \leq a(r_i)$ となるような r_i 個の部分空間の軸を決定する. ここで, 部分空間に原空間における情報量を多く維持させるために通常 κ は十分大きな値に設定する. 本論文では, あらかじめいくつかの κ を用意しておき, 五分割交差検定法により選択する. 五分割交差検定については2.5節にて説明する.

次節でカーネル法を適用したKPCAを用いたKSMにおける部分空間の生成について述べる.

2.3.2 非線形問題における部分空間の生成方法

非線形問題において部分空間法を適用する場合, 各クラスの部分空間を生成するためにKPCAを用いる. 通常のKPCAと異なり, 共分散行列を計算する際, 前節のように各クラスの部分空間は原点移動はしない.

高次元特徴空間上への写像関数 $g(\mathbf{x})$ により \mathbf{x} を l 次元特徴空間に写像する. このとき, クラス i における部分空間を構成する r_i 本の軸 $\varphi_{ik} (k = 1, \dots, r_i)$ は次式のクラス i のデータの共分散行列の固有値問題の固有ベクトルとなる.

$$\frac{1}{|X_i|} \sum_{j \in X_i} g(\mathbf{x}_j) g^T(\mathbf{x}_j) \varphi_{ik} = \lambda_{ik} \varphi_{ik} \quad \text{for } i = 1, \dots, n, k = 1, \dots, r_i \quad (2.9)$$

ここで, 前節と同様に固有値は $\lambda_{i1} > \dots > \lambda_{ir_i}$ となるように並べ替え, 式 (2.8) を用いて, クラス i の部分空間の次元数 r_i を決定する.

しかしながら, $g(\mathbf{x}_j)$ は無限次元になりうるため, 式 (2.9) を直接解くことは不可能である. そこで, この問題点を解消するために, 標本特徴空間^{32, 33)}を用いる. 標本特徴空間は特徴空間上に写像された一次独立な教師データによって構成する. このとき, 標本特徴空間上でのベクトルの内積は高次元特徴空間上でのカーネル値と等しい. これにより, 図 2.3 のように高次元特徴空間を次元数が一次独立なデータ数である低次元の空間で表せる.

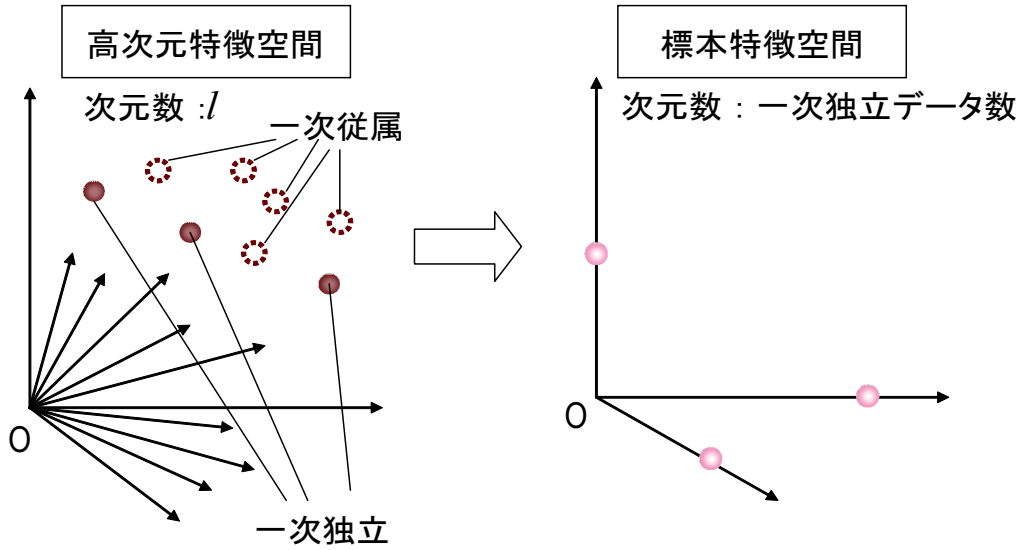


図 2.3 標本特徴空間への写像

$\mathbf{x}_j (j = 1, \dots, M)$ において, $M \times M$ のカーネル行列 $H = \{H(\mathbf{x}_j, \mathbf{x}_k)\} (j, k = 1, \dots, M)$ は対称行列であり準正定となる. ここで, H は, 次式のように表すことができる.

$$H = USU^T \quad (2.10)$$

ただし, U はカーネル行列 H の固有ベクトルを列とした行列であり, この行列は正規直交系となる. また S は次式で与えられる.

$$S = \left(\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0_N \times (M - N) \\ 0 & & \sigma_N & \\ \hline & 0_{(M-N) \times N} & & 0_{(M-N) \times (M-N)} \end{array} \right) \quad (2.11)$$

ここで, N は H の正の固有値の数であり, $\sigma_j (> 0)$ は U の j 列目の固有ベクトルに対応する固有値を示す.

N 個の U の列ベクトルを並べた $M \times N$ 行列を P , σ_j を対角成分とする $N \times N$ の対角行列を Λ とすると, 式 (2.10) は次式のように書き換えることができる.

$$H = P\Lambda P^T \quad (2.12)$$

N 次元の標本特徴空間への写像関数 $\mathbf{h}(\mathbf{x})$ は次式のように表すことができる.

$$\mathbf{h}(\mathbf{x}) = \Lambda^{-1/2} P^T (H(\mathbf{x}_1, \mathbf{x}), \dots, H(\mathbf{x}_M, \mathbf{x})) \quad (2.13)$$

標本特徴空間への写像関数 $\mathbf{h}(\mathbf{x})$ のカーネルと高次元特徴空間への写像関数 $\mathbf{g}(\mathbf{x})$ のカーネルは等しくなる. すなわち 式 (2.9) は次式のように書き換えることができる.

$$\frac{1}{|X_i|} \sum_{i \in X_i} \mathbf{h}(\mathbf{x}_j) \mathbf{h}^T(\mathbf{x}_j) \varphi_{ik} = \lambda_{ik} \varphi_{ik} \quad (2.14)$$

式 (2.9) の左辺の係数行列は RBF カーネルを用いると無限次元になりうるが, 式 (2.14) では有限次元, すなわち, N 次元となる. ここで, 式 (2.13) の固有ベクトルと固有値の計算を避けるために, N 個の一次独立なデータを求めて式 (2.13) の代わりに次式を用いる.

$$\mathbf{h}(\mathbf{x}) = (H(\mathbf{x}_{k_1}, \mathbf{x}), \dots, H(\mathbf{x}_{k_N}, \mathbf{x}))^T \quad (2.15)$$

ここで, \mathbf{x}_{k_j} は特徴空間上の一次独立なデータであり, k_j はそのときのデータの添え字である. カーネル行列 H のコレスキー分解によって一次従属なデータを削除することで一次独立なデータをデータを求めることができる. 式 (2.15) を用いて式 (2.14) を解くことにより, クラス i 部分空間の次元と軸を決定することができる. この写像関数は式 (2.13) と同一の空間に写像するが, 座標軸が異なる. そのため, 標本特徴空間上での内積はカーネルと完全に一致しない. このことによる識別への影響があるかどうかを4章の計算機実験にて述べる.

コレスキー分解による一次独立なデータの選択

ある行列 A が正定で対称行列であると仮定したとき, コレスキー分解により A は対称な下三角行列と上三角行列の積で表すことができる. すなわち, $A = LL^T$ と表すことができる. ただし, L は以下のような下三角行列である.

$$L = \begin{bmatrix} l_{11} & & 0 \\ \vdots & \ddots & \\ l_{M1} & \cdots & l_{MM} \end{bmatrix} \quad (2.16)$$

ここで L の各成分は, 以下の式により求まる.

$$l_{ii} = \sqrt{A_{ii} - \sum_{n=1}^{i-1} l_{in}^2} \quad (2.17)$$

$$l_{ij} = \left(A_{ij} - \sum_{n=1}^{j-1} l_{jn} l_{in} \right) / l_{jj} \quad \text{for } i = 1, \dots, M \quad (2.18)$$

正定である行列 A が, もし準正定に近いときにはコレスキー分解をうまくできない場合がある. その問題点を改善するために閾値 η を設定し, もし

$$A_{ii} - \sum_{n=1}^{i-1} l_{in}^2 \leq \eta \quad (2.19)$$

ならば,

$$l_{ii} = 1, l_{ij} = l_{ji} = 0 \quad \text{for } j = 1, \dots, M \quad (2.20)$$

とする. これにより一次従属になる i 番目のデータを除外して, 一次独立なデータのみを選択することができる.

2.3.3 類似度の定義

部分空間法において識別するための基準として類似度を用いる. 類似度とはある入力があるクラスの部分空間に写像した際, 元の入力と比べてどの程度近似できるかを示す指標を意味する. 部分空間法における類似度の定義を以下に述べる.

まず入力ベクトル \mathbf{x} を前述の N 次元の標本特徴空間へ写像する. このとき求められた $\mathbf{h}(\mathbf{x})$ をクラス i 部分空間に射影したときの類似度を次式のように定義する.

$$S_i(\mathbf{x}) = \sum_{k=1}^{r_i} \frac{w_{ik} (\boldsymbol{\varphi}_{ik}^T \mathbf{h}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{ik}\|^2 \|\mathbf{h}(\mathbf{x})\|^2} \quad (2.21)$$

ただし, w_{ik} はクラス i 部分空間の k 番目の軸に対する非負の重みである.

また,

$$\mathbf{f}_i(\mathbf{x}) = \left(\frac{(\boldsymbol{\varphi}_{i1}^T \mathbf{h}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{i1}\|^2 \|\mathbf{h}(\mathbf{x})\|^2}, \dots, \frac{(\boldsymbol{\varphi}_{ir_i}^T \mathbf{h}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{ir_i}\|^2 \|\mathbf{h}(\mathbf{x})\|^2} \right)^T \quad (2.22)$$

と定義する. このとき, 式 (2.21) は次式のように書き換えることができる.

$$S_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}) \quad (2.23)$$

ただし, $\mathbf{w}_i = (w_{i1}, \dots, w_{ir_i})^T$ である. このとき, $S_1(\mathbf{x}) = S_2(\mathbf{x})$ となる領域が分離超平面となる.

また, 各部分空間の次元 r_i を決定するための寄与率 κ とカーネルパラメータを設定する必要があり, 最適なパラメータの値を決定するために五分割交差検定法を用いる.

2.3.4 カーネル部分空間法による識別

KSM では, 式 (2.23) の類似度を用いて識別を行う. すなわち n クラス問題のとき入力ベクトル \mathbf{x} を

$$\arg \max_{i=1, \dots, n} S_i(\mathbf{x}) \quad (2.24)$$

となるクラス i に識別する. また各クラス部分空間を構成する軸の重み $w_{ij} (j = 1, \dots, r_i)$ を設定する必要があり, 通常の KSM ではすべての重みを等しく設定するか, または式 (2.14) により導かれる固有値を設定する.

次に図 2.4 を用いて従来の部分空間法の処理の例を説明する. 重みはすべて等しく設定する. クラス 1, 2 に対して 2 次元空間からそれぞれの 1 次元クラス部分空間を生

成るとし, φ_1, φ_2 はそれぞれクラス1, クラス2の1次元部分空間を形成する軸とする. このとき, 入力ベクトル x を各々のクラス部分空間に射影し, その射影長の長い方のクラスに分類する. この場合, クラス1に対する射影長が長いので, x はクラス1に分類される. このときのクラスの境界は, φ_1, φ_2 が成す角の二等分線上となる.

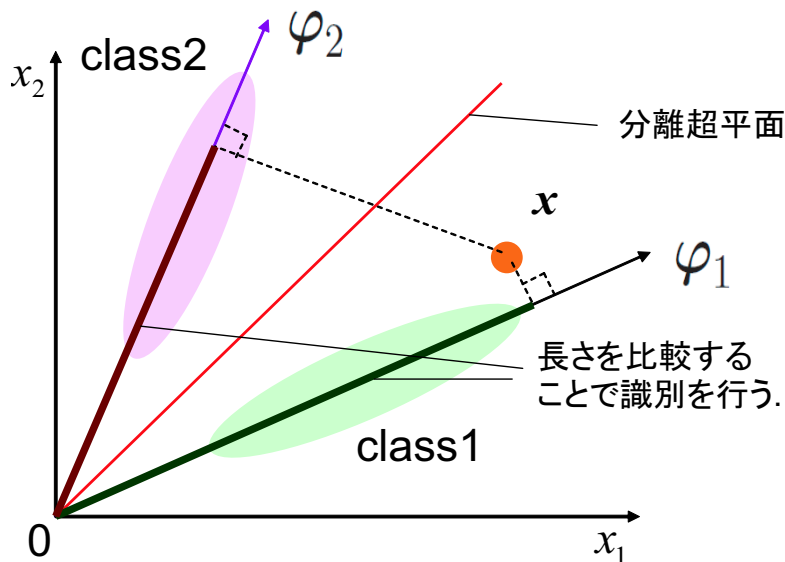


図 2.4 部分空間法による識別

KSM の学習アルゴリズムを以下に示す.

アルゴリズム 1

Step 1 五分割交差検定法を用いて, カーネル部分空間法の γ と κ の値をそれぞれ決定する.

Step 2 Step 1 で決定したパラメータの値を用いてコレスキー分解によって教師データから一次独立なデータを選択する.

Step 3 Step 2 で求めた一次独立なデータで構成した標本特徴空間への写像関数を生成する. また, 式 (2.14) よりクラス i ($i = 1, \dots, n$) の固有ベクトル φ_{ik} と固有値 λ_{ik} を求める.

Step 4 Step 1 で決定した κ の値と Step 2 で決定した λ_{ik} を用いて, クラス i 部分空間の次元数 r_i を決定する.

Step 5 式 (2.22) より $f_i(x_j)$ を計算する. ($i = 1, \dots, n, j = 1, \dots, M$)

Step 6 すべての重みを等しく設定するか, 各軸の対応する固有値 λ_{ik} と設定し, 式 (2.23) より類似度 $S_i(x)$ を求める.

2.4 サポートベクトルマシン

2.4.1 ハードマージン SVM

ここで、線形分離可能である2クラス問題において M 個の m 次元教師データ \mathbf{x}_j ($j = 1, \dots, M$) が与えられ、クラス1のデータには $y_j = 1$ 、クラス2のデータには $y_j = -1$ とラベル付けされているとする。ここで、これらの教師データを写像関数 $\mathbf{g}(\mathbf{x})$ により l 次元の高次元特徴空間に写像する。次に識別に用いる決定関数を

$$D(\mathbf{x}_j) = \mathbf{w}^T \mathbf{g}(\mathbf{x}_j) + b \quad (2.25)$$

と定義する。ただし、 \mathbf{w} は l 次元の重みベクトル、 b はバイアス項である。次に、クラス1、クラス2それぞれのデータに対して

$$\mathbf{w}^T \mathbf{g}(\mathbf{x}_j) + b \begin{cases} \geq 1 & \text{for } y_j = 1, \\ \leq -1 & \text{for } y_j = -1 \end{cases} \quad (2.26)$$

とするとき、全教師データは、

$$y_i(\mathbf{w}^T \mathbf{g}(\mathbf{x}_j) + b) \geq 1 \quad \text{for } j = 1, \dots, M \quad (2.27)$$

を満たす。このとき、分離超平面は

$$D(\mathbf{x}) = \mathbf{w}^T \mathbf{g}(\mathbf{x}) + b = c \quad \text{for } -1 < c < 1 \quad (2.28)$$

となる。分離超平面と分離超平面の最近傍ベクトルとの距離をマージンという。SVMにおいて、このマージンを最大にするマージン最大化の概念のもとで分離超平面を決定する。 $D(\mathbf{x}) = 1$ 及び $D(\mathbf{x}) = -1$ に、少なくとも1つの教師データが含まれているとすると、 $D(\mathbf{x}) = 0$ は $-1 < c < 1$ において最大となるマージンをもつ。これを最適分離超平面と呼ぶ。また、最適分離超平面の最近傍、つまり $D(\mathbf{x}) = \pm 1$ の超平面に存在するベクトルをサポートベクトルという。未知サンプルデータ \mathbf{x} は、

$$\begin{cases} D(\mathbf{x}) > 0 & \implies \text{クラス1} \\ D(\mathbf{x}) < 0 & \implies \text{クラス2} \end{cases} \quad (2.29)$$

というように決定関数の値の正負によって各クラスに識別される。図2.5に2次元2クラス問題におけるハードマージン SVM によって決定された分離超平面とマージンを示す。

2.4.2 ソフトマージン SVM

前節のハードマージン SVM は、教師データが線形分離可能な場合のみを考えた。しかしながら、実問題では線形分離可能な問題はほとんど存在しない。そこで、線形分

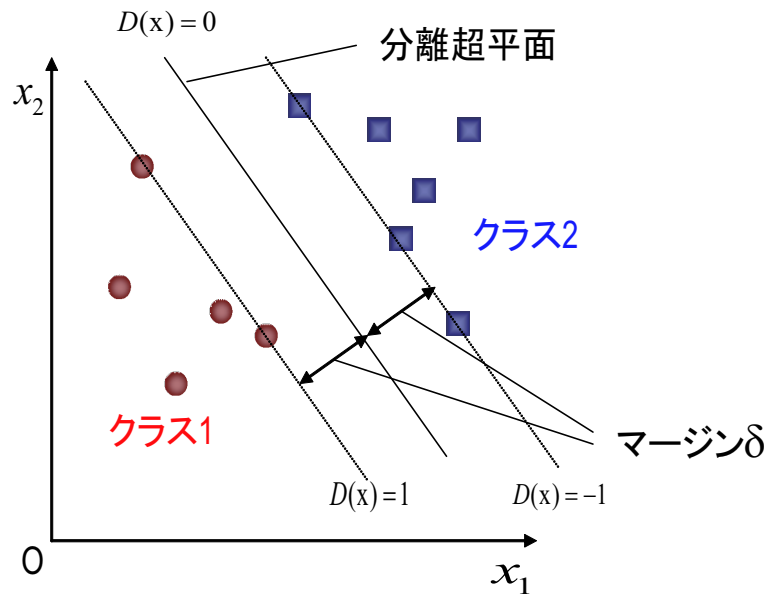


図 2.5 ハードマージン SVM による識別

線形分離不可能な問題に対応させるために、多少の誤差を許すソフトマージン SVM について説明する。図 2.6 に 2 次元 2 クラス問題における SVM によるソフトマージン SVM によって決定された分離超平面とマージンと誤差を許すための非負のスラック変数 ξ を示す。このようにハードマージン SVM とは異なり、クラス 1 のデータにおいては $D(x) < 1$ 、クラス 2 のデータにおいては $D(x) > -1$ となるデータを考慮しているため線形分離不可能な問題に対しても適用可能である。通常、SVM と呼ばれる識別器はソフトマージン SVM を示すため、以降 SVM とよぶ。

SVM では非負のスラック変数 $\xi_j (\geq 0)$ を式 (2.27) に導入することにより、図 2.6 のように線形分離不可能な問題に対して識別が可能となっている。このとき、制約条件は次式で表せる。

$$y_i(\mathbf{w}^T \mathbf{g}(\mathbf{x}_j) + b) \geq 1 - \xi_j \quad \text{for } j = 1, \dots, M \quad (2.30)$$

このスラック変数の導入により、実行可能解が常に存在する。 $0 < \xi_j < 1$ である教師データは最大マージンをもたないが、正しく分類される。しかしながら、 $\xi_j \geq 1$ であるデータは最適分離超平面によって誤分類される。最大マージンを持たない教師データの数を最小とする最適分離超平面を求めるため、以下の最小化問題を考える。

$$\text{minimize } Q(\mathbf{w}) = \sum_{i=1}^M \theta(\xi_j) \quad (2.31)$$

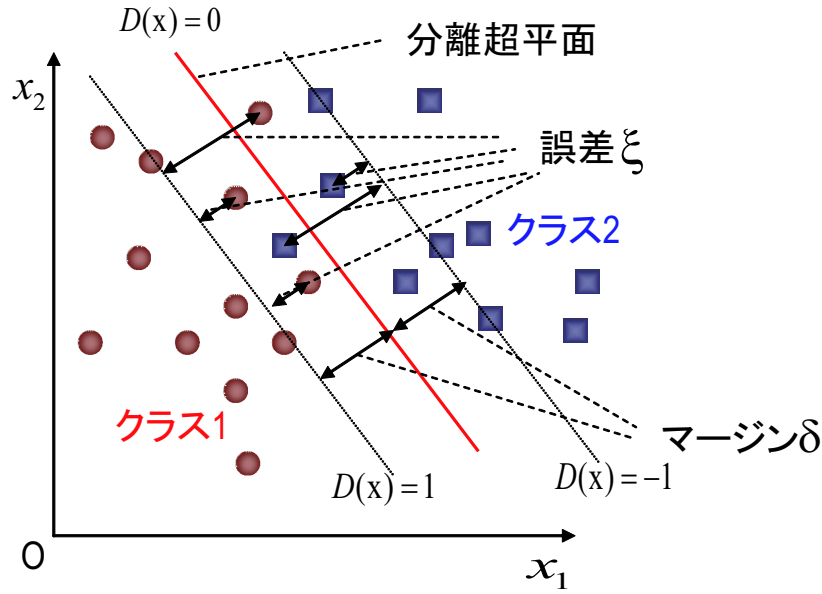


図 2.6 ソフトマージン SVM による識別

ここで

$$\theta(\xi_i) = \begin{cases} 1 & \text{for } \xi_j > 0, \\ 0 & \text{for } \xi_j = 0 \end{cases} \quad (2.32)$$

である．しかしながら，これは組み合わせ最適化問題であり，解を求めるには困難である．そこで，以下のように等価な最小化問題を解く．

$$\text{minimize} \quad \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{p} \sum_{j=1}^M \xi_j^p \quad (2.33)$$

$$\text{subject to} \quad y_i(\mathbf{w}^T \mathbf{g}(\mathbf{x}_j) + b) \geq 1 - \xi_j \quad \text{for } j = 1, \dots, M \quad (2.34)$$

ここで， C はマージン最大化と誤認識率の最小化のトレードオフを決定するマージンパラメータである．また p の値は 1 もしくは 2 から選ばれ， $p = 1$ の場合の SVM を L1-SVM， $p = 2$ の場合の SVM を L2-SVM とよぶ．SVM では入力空間から高次元特徴空間への写像を行うため，この最小化問題は解くことができない．そこで，式 (2.33)，(2.34) を教師データと同数の変数を持つ等価双対問題に変換する．

L1-SVM の場合，式 (2.33)，式 (2.34) からなる最小化問題に非負のラグランジュ乗数 $\alpha = (\alpha_1, \dots, \alpha_M)^T$ を導入し，以下の α に関する最大化問題を得る．

$$\text{maximize} \quad Q(\boldsymbol{\alpha}) = \sum_{j=1}^M \alpha_j - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j \mathbf{g}^T(\mathbf{x}_i) \mathbf{g}(\mathbf{x}_j) \quad (2.35)$$

$$\text{subject to} \quad \sum_{j=1}^M y_j \alpha_j = 0, \quad C \geq \alpha_j \geq 0 \quad \text{for } j = 1, \dots, M \quad (2.36)$$

一方, L2-SVM の場合, 式 (2.33), 式 (2.34) からなる最小化問題に非負のラグランジュ乗数 $\alpha = (\alpha_1, \dots, \alpha_M)^T$ を導入し, 以下の α に関する最大化問題を得る.

$$\text{maximize} \quad Q(\alpha) = \sum_{j=1}^M \alpha_j - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j (\mathbf{g}^T(\mathbf{x}_i) \mathbf{g}(\mathbf{x}_j) + \frac{\delta_{ij}}{C}) \quad (2.37)$$

$$\text{subject to} \quad \sum_{j=1}^M y_j \alpha_j = 0, \quad C \geq \alpha_j \geq 0 \quad \text{for } j = 1, \dots, M \quad (2.38)$$

また, これらの最適化問題を解くことにより決定関数を以下のように変換できる.

$$D(\mathbf{x}_j) = \sum_{i=1}^M y_i \alpha_i \mathbf{g}^T(\mathbf{x}_i) \mathbf{g}(\mathbf{x}_j) + b \quad (2.39)$$

ここで L1-SVM における決定関数も上式と同じになる. これにより, 最適な α_i と b を求めれば, l 次元の重みベクトル \mathbf{w} を直接求めずに決定関数はカーネル関数を用いることにより容易に導くことができる. しかしながら, 各入力において決定関数の値のみしか得られないことから識別部がブラックボックスとなる.

次節以降にその他の SVM のモデルについて説明する.

2.4.3 最小自乗サポートベクトルマシン

前述の SVM では, 2 次計画問題を解くことによって最適化を行うが, 線形の連立方程式を解くことによって最適化を行う LS-SVM について述べる. LS-SVM では計算が容易になり, データ量が膨大でないデータセットにおいて高速な学習が行える. 最小自乗サポートベクトルマシン (LS-SVM) は, 以下のような等号制約のもとで最適化問題を解くことによって学習を行う.

$$\text{minimize} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{C}{2} \sum_{i=1}^M \xi_i^2 \quad (2.40)$$

$$\text{subject to} \quad y_j (\mathbf{w}^T \mathbf{g}(\mathbf{x}_j) + b) = 1 - \xi_j \quad \text{for } j = 1, \dots, M \quad (2.41)$$

このように LS-SVM における目的関数は L2-SVM の目的関数と同じであるが, 制約条件は, SVM における制約条件の不等式を等式に変換した形式になっている. これにより, その学習は二次計画問題ではなく, 線形の連立方程式を解くことによって行うことができる. また, 通常 SVM では, スラック変数 ξ_j は正でなければならないが, LS-SVM では負の値を許している. ソフトマージン SVM と同様に式 (2.40), (2.41) にラグランジュ乗数 α_j を導入すると次のような無制約の目的関数が得られる.

$$Q(\mathbf{w}, b, \alpha, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{C}{2} \sum_{j=1}^M \xi_j^2 - \sum_{j=1}^M \alpha_j \{y_j (\mathbf{w}^T \mathbf{g}(\mathbf{x}_j) + b) - 1 + \xi_j\} \quad (2.42)$$

最小化問題であるため、式(2.42)の勾配が0となる時に最適解が得られる。したがって、式(2.42)を w, b, α, ξ で偏微分し、それぞれが0または0をとるとすると、以下のような最適条件が得られる。

$$\mathbf{w} = \sum_{j=1}^M \alpha_j y_j \mathbf{g}(\mathbf{x}_j) \quad (2.43)$$

$$\sum_{i=1}^M \alpha_j y_j = 0 \quad (2.44)$$

$$\alpha_j = C \xi_j \quad (2.45)$$

$$y_i(\mathbf{w}^T \mathbf{g}(\mathbf{x}_j) + b) - 1 + \xi_j = 0 \quad (2.46)$$

これらをKKT (Karush Kuhn Tucker) 条件とよぶ。式(2.43)–(2.46)をまとめて以下のように表す。

$$\begin{bmatrix} \Omega & \mathbf{y} \\ \mathbf{y}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ 0 \end{bmatrix} \quad (2.47)$$

ここで、

$$\Omega_{ij} = y_i y_j H(\mathbf{x}_i, \mathbf{x}_j) + \frac{\delta_{ij}}{C} \quad (2.48)$$

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (2.49)$$

$$\mathbf{y} = (y_1, \dots, y_M)^T \quad (2.50)$$

$$\mathbf{1} = (1, \dots, 1)^T \quad (2.51)$$

となる。これより、

$$b = (\mathbf{y}^T \Omega^{-1} \mathbf{y})^{-1} \mathbf{y}^T \Omega^{-1} \mathbf{1} \quad (2.52)$$

が得られ、求められた b を式(2.47)に代入することで以下のように $\boldsymbol{\alpha}$ が得られる。

$$\boldsymbol{\alpha} = \Omega^{-1}(\mathbf{1} - \mathbf{y}b) \quad (2.53)$$

よってLS-SVMでは最適化問題を1次の連立方程式を解くことによって解を得られることがわかる。

しかしながら、決定関数はL1-SVM, L2-SVMと同様に式(2.39)となるためLS-SVMにおいても識別部はブラックボックスとなる。

2.4.4 線形計画サポートベクトルマシン

本節では前節で述べたLS-SVMとは異なり2次の目的関数を線形関数に置き換えることにより線形計画問題として定式化するLP-SVMについて述べる。

SVMの目的関数である式(2.33)中のL2ノルム $\|\mathbf{w}\|_2^2 = w_1^2 + w_2^2 + \dots + w_m^2$ をL1ノルム $\|\mathbf{w}\|_1 = |w_1| + |w_2| + \dots + |w_m|$ に置き換えることにより, LP-SVM とすることができる. LP-SVMの最適化問題は,

$$\text{minimize} \quad Q(\mathbf{w}, \boldsymbol{\xi}) = \sum_{i=1}^l |w_i| + C \sum_{i=1}^M \xi_i \quad (2.54)$$

$$\text{subject to} \quad y_i(\mathbf{w}^T \mathbf{g}(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad \text{for } i = 1, \dots, M \quad (2.55)$$

のように目的関数, 制約条件ともに1次関数となる. ここで, l は次元数である. 式(2.54), 式(2.55)を単体法で解くためには, 変数が非負である必要がある. そこで $w_i^+ \geq 0, w_i^- \geq 0, b^+ \geq 0, b^- \geq 0$ を用いると $w_i = w_i^+ - w_i^-, b = b^+ - b^-$ とすることができ, 変数が $w_i^+, w_i^-, b^+, b^-, \xi$ である次式の線形計画問題となる.

$$\text{minimize} \quad Q(\mathbf{w}^+, \mathbf{w}^-, b^+, b^-, \boldsymbol{\xi}) = \sum_{i=1}^l (w_i^+ + w_i^-) + C \sum_{i=1}^M \xi_i \quad (2.56)$$

$$\text{subject to} \quad y_i(\mathbf{w}^T \mathbf{g}(\mathbf{x}_i) + b^+ - b^-) + \xi_i \geq 1 \quad \text{for } i = 1, \dots, M \quad (2.57)$$

ここで, 変数の総数は $2l + M + 2$ 個, 制約条件は M 本となる.

また, これらの最適化問題を解くことにより決定関数を以下のように変換できる.

$$D(\mathbf{x}_j) = \sum_{i=1}^M \alpha_i \mathbf{g}^T(\mathbf{x}_i) \mathbf{g}(\mathbf{x}_j) + b \quad (2.58)$$

これより, LP-SVMにおいても識別部がブラックボックスとなることが明らかである.

2.4.5 SVMの多クラス問題への拡張

前述のSVMは2クラス問題を前提として分離超平面を決定した. しかしながら, パターン認識においてクラス数が3個以上の問題は多く存在し, そのためにSVMの多クラス問題への拡張が必須となる. SVMの多クラスへの様々な拡張方式が現在までに提案されている. それらの中で本論文では, (1) 一対他方式, (2) 同時定式化方式の2つの拡張方式について述べる.

一対他方式

クラス数が n ($n > 2$) である識別問題における一対他方式を用いたSVMを説明する. まず, あるクラス i に属するデータ \mathbf{x}_j のクラスラベル y_j を1とおき, その他のデータのクラスラベルを -1 とする. 次に2クラス問題のときと同様の最適化問題を解くことにより, クラス i における重みベクトル \mathbf{w}_i , バイアス項 b_i を導く. そして, この処理を全クラスにおいて行うことにより, 以下のような n クラス分の決定関数を導ける.

$$D_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{g}(\mathbf{x}) + b_i \quad (2.59)$$

通常のSVMでは $D_i(\mathbf{x}) > 0$ ならばクラス i に分類し, $D_i(\mathbf{x}) < 0$ ならばクラス i 以外のクラスに分類し, $D_i(\mathbf{x}) = 0$ となる領域が分離超平面となる. しかしながら, このままでは未分類領域が存在してしまう.

例として, 図 2.7 に 3 クラス問題における一対他方式を用いた SVM による識別処理を示す. 図中の直線がそれぞれのクラスにおける決定関数が 0 となる領域であり, 2 クラス問題においてはこれらを分離超平面としていた. ただし, 灰色で表されている領域は $D_i(\mathbf{x}) > 0$ となるような i が 2 つ存在するか, または 1 つも存在しない領域である. そのため, これらの領域では分類が不可能であることは明らかである. そこで, これらの未分類領域を解消するために, 入力 \mathbf{x} を

$$\arg \max_{i=1, \dots, n} D_i(\mathbf{x}) \quad (2.60)$$

となるクラス i に分類する. すなわち, クラスごとに定義した決定関数の中で最大となるクラスに分類する. これにより, 図 2.7 では, 破線で描かれた直線が分離超平面となり, 分離超平面以外で未分類領域は存在しない.

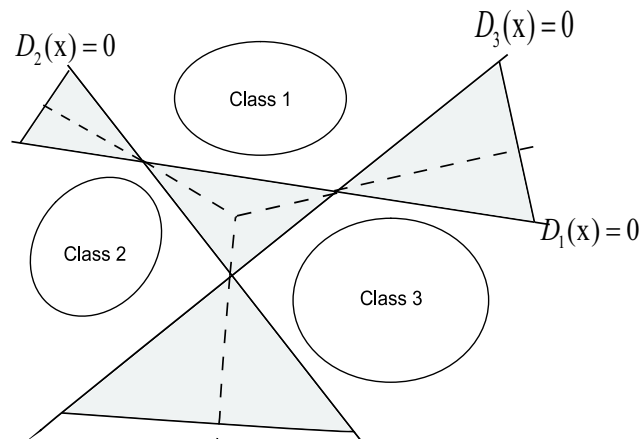


図 2.7 一対他方式を用いた SVM による識別処理

同時定式化方式

一対他方式を用いた SVM では, クラスごとの決定関数 $D_i(\mathbf{x})$ をそれぞれ別々に決定しているため, 他のクラスにおける決定関数を考慮していない. そこで他のクラスにおける決定関数を考慮し, 同時に決定する方式として同時定式化方式がある. 同時定式化方式を用いた SVM ではクラス y_j に属する教師ベクトル \mathbf{x}_j の決定関数 $D_{y_j}(\mathbf{x}_j)$ が他のクラスにおける決定関数 $D_i(\mathbf{x}_j)$ ($i \neq y_j$) より大きくなるように最適化する. また, すべての決定関数における重みベクトル \mathbf{w}_i ($i = 1, \dots, n$) とスラック変数

ξ_{ij} ($i = 1, \dots, n, j = 1, \dots, M$) を同時に決定する必要があるため、最適化問題は以下のように示される.

$$\text{minimize} \quad \frac{1}{2} \sum_{i=1}^n \|\mathbf{w}_i\|^2 + \frac{C}{p} \sum_{i=1}^n \sum_{\substack{j=1 \\ y_j \neq i}}^M \xi_j^p \quad (2.61)$$

$$\text{subject to} \quad (\mathbf{w}_{y_j} - \mathbf{w}_i)^T \mathbf{g}(\mathbf{x}_j) + b_{y_j} - b_i \geq 1 - \xi_{ij} \\ \text{for } y_j \neq i, j = 1, \dots, M, i = 1, \dots, n \quad (2.62)$$

これにより、他のクラスを考慮して最適化できるため高い汎化能力が期待できる。ただし、大規模データセットにおいて、学習コストが膨大となりうる。

2.5 モデル選択

多くの識別手法では、あらかじめ人為的に設定する必要のあるパラメータが存在する。通常、あらかじめいくつかのパラメータ値を用意しておいて、その中から最もよい汎化性能を示すようなパラメータ値を推定・選択しなければならない。このパラメータ選択により識別器の汎化能力は大きく左右される。汎化能力の高い識別器を得るためには、あらかじめ最適なパラメータ値を決めなくてはならない。本論文における汎化能力とは教師データを用いて決定された識別ルールによって未知のデータに対する識別精度を意味する。このようにパラメータ値を決定し、識別器のモデルを選択することをモデル選択と呼ぶ。モデル選択は通常、選択すべき数種類のパラメータからなる空間上の格子点における汎化能力を推定することで行う。モデル選択に広く用いられている手法のひとつとして k -分割交差検定法³¹⁾がある。

$k = 5$ のとき、すなわち、五分割交差検定法を用いる際の各モデルにおける評価を図 2.8 に示す処理を行うことにより決定する。まず、あらかじめ用意された教師データを等分に 5 分割する。次にこれらの 5 分割されたうちの 4 つのデータ集合を学習用データ、残り 1 つのデータ集合をテスト用データとして、全 5 通りの組み合わせになるように分ける。ここで、これらのテスト用データとして扱う教師データセットをバリデーションセットとよぶ。次に組み合わせごとに学習用データを用いて学習を行い、バリデーションセットを用いてテストする。そして、全組み合わせのバリデーションセットの中で正しく分類された割合を求めることにより認識率を導き出し、そのモデルにおける汎化能力を推定する。各モデルにおいて同様の処理を行い、最も高い汎化能力を示したモデルを選択する。

五分割交差検定法では各モデルに対して 1 ファイルにつき 5 回学習を行う必要がある。そのため、計算コストが大きくなる傾向があるが、最も信頼できる汎化能力の評価方法のひとつとして広く用いられている。本論文におけるすべての識別器のモデル選択の手法として五分割交差検定法を用いる。

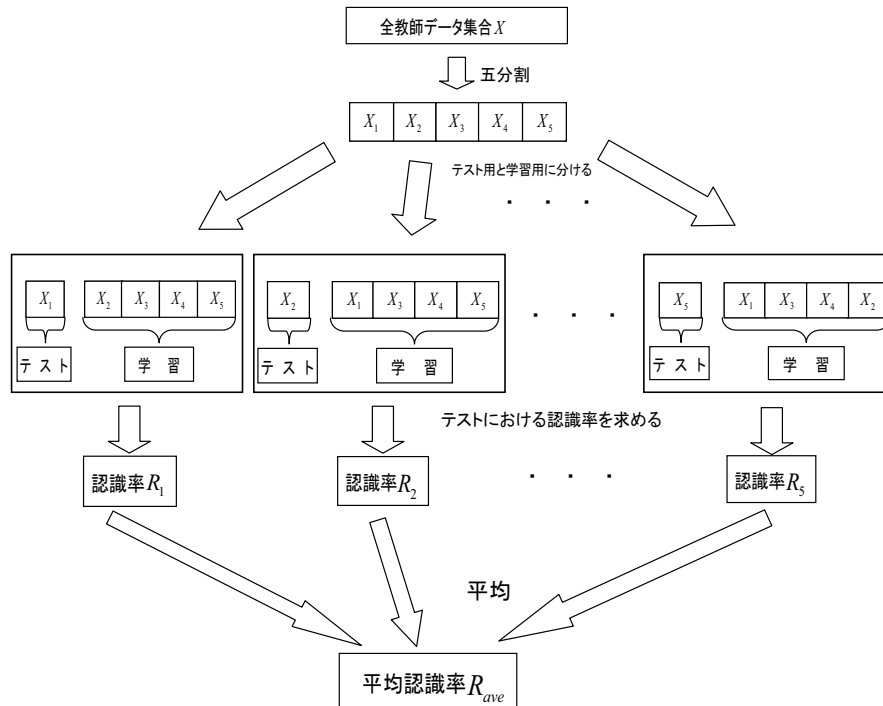


図 2.8 五分割交差検定法

2.6 結言

本章では、まず非線形問題に対応するための手法としてカーネル法を述べ、パターン認識における代表的なカーネル法を適用できる識別手法として部分空間法と SVM について説明した。部分空間法はクラスごとに PCA または KPCA を用いて特徴抽出・選択を行うことにより、低次元の部分空間を生成し、それらの射影長を基に類似度を定義することにより識別処理を行う。しかしながら、識別部を解析できるが汎化能力の低い手法である。SVM はクラス間のマージンの最大化と誤認識の最小化を基準に最適な分離超平面を決定する識別性能の高い手法である。SVM においては L1-SVM, L2-SVM, LS-SVM, LP-SVM の 4 つのモデルを紹介し、これらすべてのモデルにおいて識別部がブラックボックスであることを示した。また SVM を多クラス問題に拡張するための一対他方式と同時定式化方式を述べた。さらに、汎化能力を大きく左右するモデル選択の手法のひとつとして k -分割交差検定法について詳述した。

次章に、部分空間法の汎化能力の低い原因とその問題点の解決方法について述べる。

第3章

部分空間法に基づくサポートベクトルマシン

3.1 緒言

部分空間法^{3, 4)} またはカーネル部分空間法 (KSM: Kernel Subspace Method)^{13, 14)} は一般的にサポートベクトルマシン (SVM: Support Vector Machine)^{5, 6, 7)} やニューラルネットワーク^{1, 8)} と比べると汎化能力の低い識別器である。その原因として、部分空間を構成する軸をすべて同等に扱っているか、またはその軸の軸に対応する固有値を重み付けして扱っているためである。これより、クラス間の分離度を可能な限り大きくするという識別の観点から必ずしも最適な重みとはいえない。一方、SVMでは汎化能力は非常に高いが、2章で述べたように識別部がブラックボックスである。

そこで本章では、各クラス部分空間を構成する各軸に対して識別の観点から最適な重みを決定するために SVM のマージン最大化の概念を KSM に取り入れた部分空間法に基づく SVM (SS-SVM: Subspace-Based Support Vector Machine)^{23, 24)} を提案する。まず、KSM と同様にカーネル主成分分析 (KPCA: Kernel Principal Component Analysis)^{10, 11)} によりクラスごとに部分空間を生成し、それらに対する教師ベクトルの射影長を基にそれぞれの類似度を決定する。それらの類似度を SVM における決定関数とし、マージンが最大となるように部分空間を構成する軸に対して識別の観点から最適な重みを決定する。これより他のクラスにおける部分空間との分離を考慮できるため、KSM と比べて汎化能力の向上が期待できる。これらの方式を用いた SS-SVM に用いるモデルとして連立一次方程式を解くことにより最適化を行う最小自乗サポートベクトルマシン (LS-SVM: Least Squares Support Vector Machine)^{7, 21)}、線形計画問題を解くことにより最適化を行い、同時に特徴抽出を行える線形計画サポートベクトルマシン (LP-SVM: Linear Programming Support Vector Machine)^{7, 22)}、アクティブ集合による高速学習法を適用できる L2 サポートベクトルマシン (L2-SVM: L2 Support Vector Machine)^{7, 19, 20)} を適用する。

3.2 部分空間法の問題点と解決方法

部分空間の問題点として、他のクラスとの分離を考慮せずに識別ルールを決定するため、汎化能力が低い点が挙げられる。例として、図 3.1 に各クラス部分空間を構成す

るすべての軸の重みを等しく設定した場合を示す。ここで、2次元2クラスの線形分離可能な問題とし、クラスごとに1次元の部分空間を生成する。 φ_1, φ_2 はそれぞれのクラスの部分空間を構成する軸を表す。すべての重みを等しく設定するため、各クラスにおける類似度は射影長と等しくなり、分離超平面は各クラスの部分空間を二等分する超平面となる。この例では各クラスとも1次元部分空間であるため、図3.1のように分離超平面は二等分線となり、オーバーラップする領域が存在する。このように、通常の部分空間法では、重みの決定に他のクラスとの分離を考慮していないため識別の観点から最適な分離超平面を決定できていない。そこで、図3.2のように部分空間の各軸に対して識別の観点から最適な重みを決定し、新しい分離超平面を決定することによりこの問題点を解消する。ここでは、クラス1の部分空間の軸の重みをクラス2の部分空間の軸の重みより大きく設定することで、分離超平面をクラス2側に傾かせることができ、オーバーラップする領域を解消することができる。

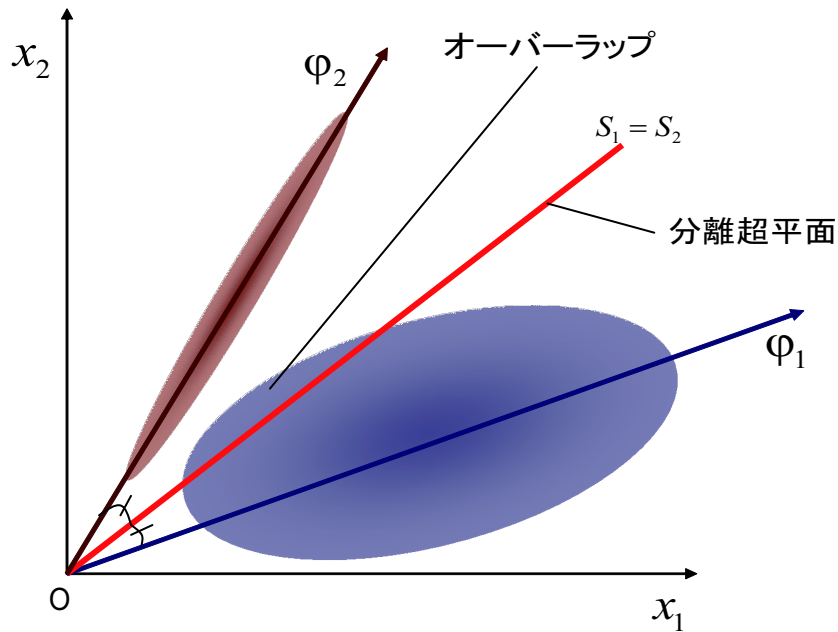


図 3.1 通常の部分空間法による誤認識

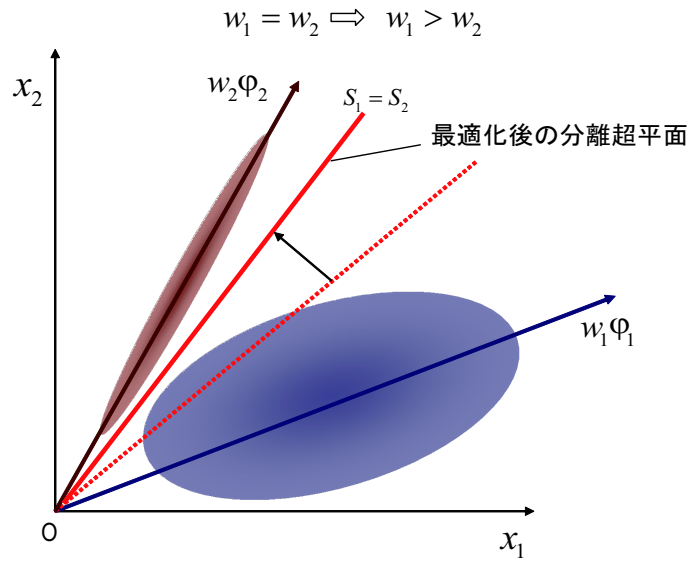


図 3.2 重みの最適化

3.3 部分空間法に基づく最小自乗サポートベクトルマシン

部分空間法に適用する SVM のモデルとして LS-SVM を用いた部分空間法に基づく LS-SVM (SSLS-SVM: Subspace-Based Least Squares Support Vector Machine) について述べる.

M 個の教師データのペアを $\{(x_1, y_1), \dots, (x_M, y_M)\}$ とおき, n クラス問題とする. ここで, \mathbf{x} と y_i はそれぞれ m 次元入力ベクトル, クラスラベルとし, $y_i \in \{1, \dots, n\}$ とする. まず式 (2.23) は $f_i(x_j)$ より導かれるクラス i の次式のような決定関数とみなすことができる.

$$D_i(\mathbf{x}) = S_i(\mathbf{x}) \quad \text{for } i = 1, \dots, n \quad (3.1)$$

また, 次式のようにデータの対応するクラスにおける類似度が全クラスにおける類似度の中で最大となるように重みを決定するための制約を設ける.

$$D_{y_j}(\mathbf{x}_j) > D_i(\mathbf{x}_j) \quad \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (3.2)$$

このように, 全クラスにおける類似度を考慮して同時にすべての重みを決定する必要があるため, 同時定式化方式を適用する. このとき, $D_{y_j}(\mathbf{x}) = D_i(\mathbf{x})$ で表される超平面を分離超平面とよび, 分離超平面と $D_{y_j}(\mathbf{x}_j) - D_i(\mathbf{x}_j) = 1$ の距離をマージンと呼ぶ. 類似度 $f_{y_j}(\mathbf{x}_j)$ と分離超平面との距離は $|D_{y_j}(\mathbf{x}_j) - D_i(\mathbf{x}_j)| / \|\mathbf{w}_{y_j}\|$ で与えられるので, マージン σ は

$$\sigma_{y_j} = \frac{1}{\|\mathbf{w}_{y_j}\|} \quad (3.3)$$

として定義される.

ここで, SS-SVM では最適化問題において同時定式化方式または一対他方式を用いた2つのモデルが考えられる.

3.3.1 同時定式化方式を適用した SSLS-SVM

ここで SS-SVM におけるモデルのひとつである SSLS-SVM について述べる. 各教師データ $\mathbf{x}_j (j = 1, \dots, M)$ に対して, 式 (3.2) の不等号制約をスラック変数を ξ_{ji} を導入することにより次式のような等号制約とする.

$$D_{y_j}(\mathbf{x}_j) - D_i(\mathbf{x}_j) = 1 - \xi_{ji} \quad \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (3.4)$$

このとき, 決定関数 (3.1) を得るために, 以下の最適化問題を解く.

$$\text{minimize } Q(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2} \sum_{i=1}^n \|\mathbf{w}_i\|^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ y_j \neq i}}^M \frac{CM}{2n |X_{y_j}|} \xi_{ji}^2 \quad (3.5)$$

$$\begin{aligned} \text{subject to } & \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) = 1 - \xi_{ji} \\ & \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \end{aligned} \quad (3.6)$$

ここで $M/(n|X_{y_j}|)$ は不均衡クラスデータによって偏ることを回避するための係数である. KSM では, 重みは非負と仮定しているが, SSLS-SVM における最適化の際, 単純に非負制約を設けることは困難となるため, ここでは設けない.

また, 各クラスにおける部分空間の次元数は, それぞれ最大でクラスに属する教師ベクトル数となる. また, LS-SVM における主問題では用いる空間の次元数に学習コストが依存する. 一方, 双対問題ではサポートベクトル数に学習コストが依存するが, $\xi_{ji} < 0$ となりうるため, すべての教師ベクトルがサポートベクトルに必ずなる. そのため, 高次元特徴空間上での最適化問題を解く通常の LS-SVM と異なり, SSLS-SVM では双対問題より主問題を解くことの方が学習コストが小さくなると推測できる. そこで SSLS-SVM では最適化問題の主問題を解く. 式 (3.5) に式 (3.6) を代入すると, 次式のように表せる.

$$Q(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^n \|\mathbf{w}_i\|^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ y_j \neq i}}^M \frac{CM}{2n |X_{y_j}|} \times (1 - (\mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j)))^2 \quad (3.7)$$

また, 最小化問題なので式 (3.7) の \mathbf{w}_i の勾配が 0 となる, すなわち

$$\frac{\partial Q(\mathbf{w}_i)}{\partial \mathbf{w}_i} = \left\{ \frac{n}{MC} I_{r_i} + \sum_{\substack{j=1, \\ y_j \neq i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_i^T(\mathbf{x}_j) \right\} \quad (3.8)$$

$$\begin{aligned}
& + (n-1) \sum_{\substack{j=1, \\ y_j=i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_i^\top(\mathbf{x}_j) \} \mathbf{w}_i \\
& - \left\{ \sum_{\substack{j=1, \\ y_j \neq i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_{y_j}^\top(\mathbf{x}_j) \right\} \mathbf{w}_{y_j} - (n-1) \sum_{\substack{j=1, \\ y_j=i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) \\
& + \sum_{\substack{j=1, \\ y_j \neq i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) = \mathbf{0} \quad \text{for } i \neq y_j, i = 1, \dots, n
\end{aligned}$$

となる。ここで

$$\begin{aligned}
A_i &= \frac{n}{MC} I_{r_i} + \sum_{\substack{j=1, \\ y_j \neq i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_i^\top(\mathbf{x}_j) + (n-1) \sum_{\substack{j=1, \\ y_j=i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_i^\top(\mathbf{x}_j) \\
& \text{for } i = 1, \dots, n
\end{aligned} \tag{3.9}$$

$$B_{iy_j} = - \sum_{\substack{j=1, \\ y_j \neq i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_{y_j}^\top(\mathbf{x}_j) \quad \text{for } i \neq y_j, i = 1, \dots, n \tag{3.10}$$

$$\mathbf{a}_i = (n-1) \sum_{\substack{j=1, \\ y_j=i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) - \sum_{\substack{j=1, \\ y_j \neq i}}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) \quad \text{for } i = 1, \dots, n \tag{3.11}$$

と定義すると解くべき最適問題は以下ようになる。

$$\mathbf{w} = \begin{pmatrix} A_1 & B_{12} & \cdots & B_{1n} \\ B_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ B_{n1} & \cdots & \cdots & A_n \end{pmatrix}^{-1} \mathbf{a} \tag{3.12}$$

これにより、全クラスの部分空間を構成する軸に対する重みを同時に決定することができる。

評価実験ではラジアル基底関数 (RBF: Radial Basis Function) カーネルを用いる。このとき、SLS-SVM で学習を行う前に γ と寄与率 κ とマージンパラメータ C の値を決定する必要がある。この決定には前述の五分割交差検定法を用いる。同時定式化方式を用いた SLS-SVM の学習アルゴリズムは以下ようになる。

アルゴリズム 2

Step 1 五分割交差検定法を用いて、 γ , κ , C 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って教師データから一次独立なデータを選択する。

Step 3 Step 2 で求めた一次独立なデータで構成した標本特徴空間への写像関数を生成する. また, 式 (2.14) よりクラス i ($i = 1, \dots, n$) の固有ベクトル φ_{ik} と固有値 λ_{ik} を導く.

Step 4 Step 3 で求めた固有値を各クラスそれぞれ大きい値から順に並び替える. そして固有値の総和との比率が Step 1 で決定した κ より大きくなるまで順番に固有値を足していく. 足された固有値に対応する固有ベクトルをクラス i ($i = 1, \dots, n$) 部分空間を形成する軸とし, 次元数 r_i を決定する.

Step 5 式 (2.22) より $f_i(\mathbf{x}_j)$ ($i = 1, \dots, n, j = 1, \dots, M$) を計算する.

Step 6 式 (3.12) を用いて重み \mathbf{w}_i ($i = 1, \dots, n$) を導く.

上記の方式では, 同時定式化方式を用いているため, 前述のように $\sum_{i=1}^n r_i \times \sum_{i=1}^n r_i$ 行列を解く必要がある. そのため, 多クラス問題や各部分空間の次元数が大きくなるデータセットにおいて重みの最適化における学習コストが膨大となると推測できる. そこで, クラスごとの部分空間の重みを別々に最適化する一対他方式を適用した部分空間法に基づく最小自乗サポートベクトルマシンを考える. 以下, この手法を SSSL-SVM (O) とよぶ. 次節に SSSL-SVM (O) の詳細を述べる.

3.3.2 一対他方式を適用した SSSL-SVM

本節では SSSL-SVM にそれぞれのクラスの類似度の重みを別々に最適化する SSSL-SVM (O) を考える.

一対他方式では, 他のクラスの部分空間を構成する軸に対する重みを考慮せずにクラスごとに重みを決定する. そのため, 汎化能力が同時定式化方式を用いた SSSL-SVM と比べ劣化すると推測できる. そこで, その劣化を抑えるため, 式 (3.1) に代えて, 次式のようなバイアス項 b_i を含む決定関数を定義する.

$$D_i(\mathbf{x}) = S_i(\mathbf{x}) + b_i \quad (3.13)$$

また, $D_{y_j}(\mathbf{x}_j) = D_i(\mathbf{x}_j)$ ($i \neq y_j$) で表される超平面は分離超平面となる. 類似度 $f_{y_j}(\mathbf{x}_j)$ と分離超平面との距離は $|D_{y_j}(\mathbf{x}_j)| / \|\mathbf{w}_{y_j}\|$ で与えられるため, マージン σ は同時定式化と同様に式 (3.3) として定義される. また各教師データ \mathbf{x}_j ($j = 1, \dots, M$) に対して, スラック変数 ξ_{ji} を以下のようにする.

$$\xi_{ji} = \begin{cases} 1 - D_i(\mathbf{x}_j) & \text{if } y_j = i \\ -1 + D_i(\mathbf{x}_j) & \text{if } y_j \neq i \end{cases} \quad (3.14)$$

このとき, クラス i ($i = 1, \dots, n$) において SSSL-SVM (O) の定式化は次式のように表せる.

$$\text{minimize } Q(\mathbf{w}, \mathbf{b}, \boldsymbol{\xi}) = \frac{1}{2} \|\mathbf{w}_i\|^2 + \sum_{j=1}^M \frac{CM}{2n |X_{z_j}|} \xi_{ij}^2 \quad (3.15)$$

$$\text{subject to } z_j(\mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) + b_i) = 1 - \xi_{ij} \quad \text{for } j = 1, \dots, M \quad (3.16)$$

ここで, z_j は $y_j = i$ のとき 1 とおき, $y_j \neq i$ のとき -1 とする.

また, 同時定式化と同様に最適化問題の主問題を解くため, 式 (3.15) に式 (3.16) を代入すると次式のように表せる.

$$Q(\mathbf{w}_i, b_i) = \frac{1}{2} \|\mathbf{w}_i\|^2 + \sum_{j=1}^M \frac{CM}{2n|X_{z_j}|} (1 - z_j(\mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) + b_i))^2 \quad (3.17)$$

最小化問題なので式 (3.7) の \mathbf{w}_i と b_i の勾配がそれぞれ 0, 0 となる. すなわち

$$\frac{\partial Q(\mathbf{w}_i, b_i)}{\partial \mathbf{w}_i} = \frac{n}{CM} I_{r_i} \mathbf{w}_i + \sum_{j=1}^M \frac{1}{|X_{z_j}|} \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_i^T(\mathbf{x}_j) \mathbf{w}_i \quad (3.18)$$

$$- \sum_{j=1}^M \frac{1}{|X_{z_j}|} \mathbf{f}_i(\mathbf{x}_j) \sum_{j=1}^M \frac{1}{|X_{z_j}|} \mathbf{f}_i^T(\mathbf{x}_j) \mathbf{w}_i - \sum_{j=1}^M \frac{1}{|X_{z_j}|} z_j \mathbf{f}_i(\mathbf{x}_j) = \mathbf{0}$$

$$\frac{\partial Q(\mathbf{w}_i, b_i)}{\partial b_i} = b_i + \mathbf{w}_i^T \sum_{j=1}^M \frac{1}{|X_{y_j}|} \mathbf{f}_i(\mathbf{x}_j) = 0 \quad (3.19)$$

となる. ここで

$$\Omega_i = \frac{n}{CM} I_{r_i} + \sum_{j=1}^M \frac{1}{|X_{z_j}|} \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_i^T(\mathbf{x}_j) - \sum_{j=1}^M \frac{1}{|X_{z_j}|} \mathbf{f}_i(\mathbf{x}_j) \sum_{j=1}^M \frac{1}{|X_{z_j}|} \mathbf{f}_i^T(\mathbf{x}_j) \quad (3.20)$$

$$\mathbf{a}'_i = \sum_{j=1}^M \frac{1}{|X_{z_j}|} z_j \mathbf{f}_i(\mathbf{x}_j) \quad (3.21)$$

と定義すると, 解くべき最適化問題は以下ようになる.

$$\mathbf{w}_i = \Omega_i^{-1} \mathbf{a}'_i \quad (3.22)$$

$$b_i = -\mathbf{w}_i^T \sum_{j=1}^M \frac{1}{|X_{z_j}|} \mathbf{f}_i(\mathbf{x}_j) \quad (3.23)$$

$r_i \times r_i$ 行列を計算するので, 行列演算数はオーダー $\sum_i^n r_i^3$ となる. また, 同時定式化ではオーダー $(\sum_{i=1}^n r_i)^3$ となるので, 一対他方式は計算量が同時定式化に比べ非常に少なくなる. SSLS-SVM (O) の学習アルゴリズムを次に示す.

アルゴリズム 3

Step 1 五分割交差検定法を用いて, γ, κ, C の値をそれぞれ決定する.

Step 2 Step 1 で決定したパラメータの値を用いてコレスキー分解によって教師データから一次独立なデータを選択する. $i = 1$ とおく.

Step 3 Step 2 で求めた一次独立なデータで構成した標本特徴空間への写像関数を生成する. また, 式 (2.14) よりクラス i の固有ベクトル φ_{ik} と固有値 λ_{ik} を導く.

- Step 4** Step 3 で求めた固有値をそれぞれ大きい値から順に並び替える．そして固有値の総和との比率が Step 1 で決定した κ より大きくなるまで順番に固有値を足していく．足された固有値に対応する固有ベクトルをクラス i 部分空間を形成する軸とし，次元数 r_i を決定する．
- Step 5** 各クラス部分空間の軸を用いて式 (2.22) より $\mathbf{f}_i(\mathbf{x}_j)$ ($j = 1, \dots, M$) を計算する．
- Step 6** 式 (3.22), (3.23) より重みベクトル \mathbf{w}_i とバイアス項 b_i をそれぞれ求める．
- Step 7** もし $i \neq n$ ならば, $i = i + 1$ とおき, Step 3 に戻る． $i = n$ ならば, 学習を終了する．

3.4 部分空間法に基づく線形計画サポートベクトルマシン

本節では, SS-SVM に適用する SVM のモデルとして LP-SVM を用いた部分空間法に基づく LP-SVM (SSLP-SVM: Subspace-Based Linear Programming Support Vector Machine) について述べる．

3.4.1 同時定式化方式を適用した SSLP-SVM

SSLP-SVM も SSSL-SVM と同様に同時定式化方式または一対他方式の 2 つのモデルが存在し, まず同時定式化方式を適用した SSLP-SVM について本節で述べる．

通常の LP-SVM では, 汎化能力が他のモデルの SVM に比べて劣る．汎化能力を高めるため, 同時定式化手法を用いた SSSL-SVM とは異なり SSLP-SVM ではバイアス項を決定関数に導入する．SSLP-SVM では以下の最適化問題を解く．

$$\text{minimize } Q(\mathbf{w}^+, \mathbf{w}^-, \mathbf{b}^+, \mathbf{b}^-, \boldsymbol{\xi}) = \sum_{i=1}^n \sum_{k=1}^{r_k} (w_{ik}^+ + w_{ik}^-) + \sum_{j=1}^M \sum_{\substack{i \neq y_j \\ i=1}}^n \frac{CM}{n |X_{y_j}|} \xi_{ji} \quad (3.24)$$

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

$$\text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (3.25)$$

$$\xi_{ji} \geq 0 \quad \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (3.26)$$

単体法により上記の最適化問題を解くためには, 不等式制約を等式制約に変形し, 標準形にする必要がある．そこで非負のスラック変数 u_{ji} ($i = 1, \dots, n, i \neq y_j, j = 1, \dots, M$) を導入する．このとき, 式 (3.24)–(3.26) により与えられる最適化問題は, 以下のように変換できる．

$$\text{minimize } Q(\mathbf{w}^+, \mathbf{w}^-, \mathbf{b}^+, \mathbf{b}^-, \boldsymbol{\xi}, \mathbf{u}) = \sum_{i=1}^n \sum_{k=1}^{r_k} (w_{ik}^+ + w_{ik}^-) + \sum_{j=1}^M \sum_{\substack{i \neq y_j \\ i=1}}^n \frac{CM}{n N_{y_j}} \xi_{ji} \quad (3.27)$$

$$\begin{aligned} \text{subject to } & \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) + b_{y_j}^+ - b_{y_j}^- - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) - b_i^+ + b_i^- = 1 - \xi_{ji} + u_{ji} \\ & \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \end{aligned} \quad (3.28)$$

$$\xi_{ji} \geq 0 \quad \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (3.29)$$

ここで $\mathbf{u} = (\dots, u_{ji}, \dots)^T$ である。同時定式化方式では、全クラスの変数を同時に最適化するため、目的関数の第一項、及び第二項ともに各クラスの変数を足し合わせる必要があり、変数の数は $\sum_{i=1}^n 2r_i + (n-1)M + 2n$ 個となり、制約条件の数は $(n-1)M$ 個となる。

この最適化問題において、もし $w_{ik} = 0$ ならば、その重みに対応する軸は識別には用いる必要がない。すなわち、 $f_{ik}(\mathbf{x})$ の要素は削除しても識別に影響を及ぼさない。ここで $f_{ik}(\mathbf{x})$ は $f_i(\mathbf{x})$ の k 番目の要素である。つまり、LP-SVM では最適化と同時に特徴選択を行える。

SSLS-SVM とは異なり、重みの最適化の過程で特徴選択ができるため、他のクラスを考慮しない累積寄与率による特徴選択を行う必要もない。そのため、あらかじめ設定する κ は $\kappa = 99.9(\%)$ と設定する。同時定式化方式を用いた SSLP-SVM の学習アルゴリズムは以下ようになる。

アルゴリズム 4

Step 1 $\kappa = 99.9(\%)$ に設定し、五分割交差検定法を用いて γ, C 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する。

Step 3 Step 2 で求めた一次独立なデータで構成した標本特徴空間への写像関数を生成する。また、式 (2.14) よりクラス i ($i = 1, \dots, n$) の固有ベクトル φ_{ik} と固有値 λ_{ik} を導く。

Step 4 Step 3 で求めた固有値を各クラスそれぞれ大きい値から順に並び替える。そして固有値の総和との比率が Step 1 で決定した κ より大きくなるまで順番に固有値を足していく。足された固有値に対応する固有ベクトルをクラス i ($i = 1, \dots, n$) 部分空間を形成する軸とし、次元数 r_i を決定する。

Step 5 式 (2.22) から $\mathbf{f}_i(\mathbf{x}_j)$ ($i = 1, \dots, n, j = 1, \dots, M$) を計算する。

Step 6 式 (3.27)–(3.29) を解くことにより、 \mathbf{w}_i, b_i ($i = 1, \dots, n$) を得る。

2 クラス問題において、同時定式化方式では変数が制約条件の数に比べ膨大となる。すなわち、単体法では (変数の数)–(制約条件の数) 個の変数の更新は不可能であり、必ず値が 0 である。そのため、必要以上に軸の削除をしてしまう可能性がある。また、変数が多くなることにより計算コストも膨大となる。そこで、これらの問題を緩和する

ため、同時定式化の近似手法である一対他方式を次節で述べる。以下、一対他方式を適用した SSLP-SVM を SSLP-SVM (O) とよぶ。

3.4.2 一対他方式を適用した SSLP-SVM

SLP-SVM (O) の最適化問題を以下に示す。

$$\text{minimize} \quad Q(\mathbf{w}^+, \mathbf{w}^-, \mathbf{b}^+, \mathbf{b}^-, \boldsymbol{\xi}) = \sum_{k=1}^{r_k} (w_{ik}^+ + w_{ik}^-) + \sum_{j=1}^M \frac{CM}{n |X_{y_j}|} \xi_{ji} \quad (3.30)$$

$$\text{subject to} \quad y_j (\mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) + b_i) \geq 1 - \xi_{ji}, \quad \text{for } i = 1, \dots, n, j = 1, \dots, M \quad (3.31)$$

$$\xi_{ji} \geq 0 \quad \text{for } i = 1, \dots, n, j = 1, \dots, M \quad (3.32)$$

さらに、同時定式化方式と同じく単体法により最適化問題を解くため、以下のように標準形に変換する。

$$\text{minimize} \quad Q(\mathbf{w}^+, \mathbf{w}^-, \mathbf{b}^+, \mathbf{b}^-, \boldsymbol{\xi}, \mathbf{u}) = \sum_{k=1}^{r_k} (w_{ik}^+ + w_{ik}^-) + \sum_{j=1}^M \frac{CM}{n |X_{y_j}|} \xi_{ji} \quad (3.33)$$

$$\text{subject to} \quad y_j (\mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) + b_i) = 1 - \xi_{ji} + u_{ji}, \quad \text{for } i = 1, \dots, n, j = 1, \dots, M \quad (3.34)$$

$$\xi_{ji} \geq 0 \quad \text{for } i = 1, \dots, n, j = 1, \dots, M \quad (3.35)$$

ここで、変数の数は $2r_i + M + 2$ 個となり、制約条件の数は M 個となる。このように一対他方式では、データが属するクラスとそれ以外のクラスとのマージンが最大になるように分離超平面を決定し、各クラス毎の関係を考慮しなくてもよい。そのため、変数、制約条件数ともに1クラス分となるため、同時定式化方式に比べ少なくなる。

以下に SSLP-SVM (O) の学習アルゴリズムを示す。

アルゴリズム 5

Step 1 $\kappa = 99.9(\%)$ に設定し、五分割交差検定法を用いて γ, C 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する。 $i = 1$ と設定する。

Step 3 Step 2 で求めた一次独立なデータで構成した標本特徴空間への写像関数を生成する。また、式 (2.14) よりクラス i の固有ベクトル φ_{ik} と固有値 λ_{ik} を導く。

Step 4 Step 3 で求めた固有値を各クラスそれぞれ大きい値から順に並び替える。そして固有値の総和との比率が Step 1 で決定した κ より大きくなるまで順番に固有値を足していく。足された固有値に対応する固有ベクトルをクラス i 部分空間を形成する軸とし、次元数 r_i を決定する。

Step 5 式 (2.22) から $\mathbf{f}_i(\mathbf{x}_j)$ ($j = 1, \dots, M$) を計算する.

Step 6 式 (3.33)–(3.35) を解くことにより最適な \mathbf{w}_i 及び, b_i の値を得る.

Step 7 $i \neq n$ ならば, $i = i + 1$ とし, Step 3 に戻る. $i = n$ ならば学習を終了する.

3.5 部分空間法に基づく L2 サポートベクトルマシン

SSLS-SVM における解くべき線形システムのサイズが同時定式化では全クラスの部分空間の次元数の総和, 一対他方式ではクラスごとの部分空間の次元数に依存する. よって, 各クラスの部分空間の次元数が低ければ, 非常に高速な学習が行える. しかしながら, 大規模データセットなどにおいては次元数が非常に多くなりうる. また, SSLS-SVM では, すべてのデータをサポートベクトルとして同等に取り扱うため, 識別面から遠く離れた明らかに正しく分類できるデータを学習で用いる. このとき, それらのデータは識別に悪影響を及ぼしうる. そこで, これらの問題点を解消する部分空間法に基づく L2-SVM (SSL2-SVM: Subspace-Based L2 Support Vector Machine) について述べる. SSL2-SVM では識別面付近または誤認識するデータのみをサポートベクトルとして取り扱うため, 識別面から遠く離れた明らかに正しく分類できるデータを学習に用いない. サポートベクトルの選択のためにアクティブ集合による学習を SSL2-SVM に適用する. アクティブ集合による学習とは, 一部のデータを用いて学習を行い, 全データが KKT (Karush Kuhn Tucker) 条件を満たすまで学習するデータの入れ替えを行う繰り返し学習法である.

3.5.1 同時定式化方式を適用した SSL2-SVM

同時定式化方式を適用した SSL2-SVM における最適化問題は以下のようになる.

$$\text{minimize} \quad Q(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2} \sum_{i=1}^n \|\mathbf{w}_i\|^2 + \sum_{i=1}^n \sum_{\substack{y_j \neq i, \\ j=1}}^M \frac{C}{2} \xi_{ji}^2 \quad (3.36)$$

$$\begin{aligned} \text{subject to} \quad & \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) \geq 1 - \xi_{ji} \quad (3.37) \\ \text{for} \quad & i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \end{aligned}$$

ここで, SSLS-SVM と異なる点は制約条件が不等号制約であることとスラック変数 ξ_{ji} が非負である点である. また SSL2-SVM では, SSLS-SVM とは異なり, 学習コストを次元数ではなくサポートベクトル数に依存させるため双対問題を解く必要がある. ここで, ラグランジュ乗数 α_{ij} を導入し最適化問題を以下の形に変換する.

$$\text{minimize} \quad Q(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2} \sum_{i=1}^n \|\mathbf{w}_i\|^2 + \sum_{i=1}^n \sum_{\substack{i \neq y_j, \\ j=1}}^M \xi_{ji}^2 \quad (3.38)$$

$$\begin{aligned}
& - \sum_{i \neq y_j}^n \sum_{j=1}^M \alpha_{ij} (\mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) - 1 + \xi_{ji}) \\
& \xi_{ji} \geq 0, \alpha_{ij} \geq 0 \quad \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (3.39)
\end{aligned}$$

この最適化解は式 (3.39) を \mathbf{w}_i, ξ_{ij} で偏微分し、それぞれ 0, 0 となる点である。したがって、次式により求めることができる。

$$\mathbf{w}_i = \sum_{k \neq i, k=1}^n \sum_{\substack{j \in S_k, \\ y_j = i}} \alpha_{kj} \mathbf{f}_i(\mathbf{x}_j) - \sum_{\substack{j \in S_i, \\ y_j \neq i}} \alpha_{ij} \mathbf{f}_i(\mathbf{x}_j) \quad (3.40)$$

$$\xi_{ji} = \frac{\alpha_{ij}}{C}, \quad (3.41)$$

ここで、 S_i は $D_i(\mathbf{x})$ におけるサポートベクトルの添え字集合を意味している。また、クラス i におけるサポートベクトル数を $|S_i|$ と定義する。そして、KKT 条件は以下のようになる。

$$\alpha_{ij} \{ \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) - 1 + \xi_{ji} \} = 0 \quad \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (3.42)$$

式 (3.40), (3.41) を式 (3.42) に代入することにより α_{ij} ($i = 1, \dots, n, j = 1, \dots, M, i \neq y_j$) を求めることができる。したがって、サポートベクトルを選択できれば、以下の $\sum_{i=1}^n |S_i|$ の線形システムを解くことにより α_{ij} を導ける。

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_n \end{pmatrix} = \begin{pmatrix} A_1 & B_{12} & \cdots & B_{1n} \\ B_{12}^T & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_{(n-1)n} \\ B_{1n}^T & \cdots & B_{(n-1)n}^T & A_n \end{pmatrix}^{-1} \mathbf{1} \quad (3.43)$$

ここで、 $\mathbf{1} = (1, \dots, 1)^T$, $\boldsymbol{\alpha}_i = (\alpha_{iS_{i1}}, \dots, \alpha_{iS_{i|S_i|}})^T$, $A_i = \{A_{ikl}\}$ ($i = 1, \dots, n, k, l = 1, \dots, |S_i|$), $B_{ij} = \{B_{ijkl}\}$ ($i = 1, \dots, n, j = i+1, \dots, n, k = 1, \dots, |S_i|, l = 1, \dots, |S_j|$) であり、

$$A_{ikl} = \begin{cases} \mathbf{f}_i^T(\mathbf{x}_{s_{ik}}) \mathbf{f}_i(\mathbf{x}_{s_{il}}) + \mathbf{f}_{y_{s_{ik}}}^T(\mathbf{x}_{s_{ik}}) \mathbf{f}_{y_{s_{ik}}}(\mathbf{x}_{s_{il}}) + \frac{\delta_{kl}}{C} & \text{for } y_{s_{ik}} = y_{s_{il}} \\ \mathbf{f}_i^T(\mathbf{x}_{s_{ik}}) \mathbf{f}_i(\mathbf{x}_{s_{il}}) & \text{otherwise} \end{cases} \quad (3.44)$$

$$B_{ijkl} = \begin{cases} \mathbf{f}_{y_{s_{ik}}}^T(\mathbf{x}_{s_{ik}}) \mathbf{f}_{y_{s_{ik}}}(\mathbf{x}_{s_{jl}}) & \text{for } y_{s_{ik}} = y_{s_{jl}} \neq i, j \\ -\mathbf{f}_{y_{s_{ik}}}^T(\mathbf{x}_{s_{ik}}) \mathbf{f}_{y_{s_{ik}}}(\mathbf{x}_{s_{jl}}) & \text{for } y_{s_{ik}} = i, j, y_{s_{jl}} \neq i, j \\ -\mathbf{f}_{y_{s_{jl}}}^T(\mathbf{x}_{s_{ik}}) \mathbf{f}_{y_{s_{jl}}}(\mathbf{x}_{s_{jl}}) & \text{for } y_{s_{ik}} \neq i, j, y_{s_{jl}} = i, j \\ -(\mathbf{f}_i^T(\mathbf{x}_{s_{ik}}) \mathbf{f}_i(\mathbf{x}_{s_{jl}}) + \mathbf{f}_j^T(\mathbf{x}_{s_{ik}}) \mathbf{f}_j(\mathbf{x}_{s_{jl}})) & \text{otherwise} \end{cases} \quad (3.45)$$

と表せる. ここで, s_{ij} ($i = 1, \dots, n, j = 1, \dots, |S_i|$) はクラス i における j 番目サポートベクトルを示す.

サポートベクトルの集合 S_i を得るために, 繰り返し学習を行う. この繰り返し学習として, アクティブ集合による学習法を用いる. しかしながら, 通常の SVM におけるアクティブ集合による学習法とは異なり, SSL2-SVM では同時定式化方式を用いる. そのため, 全クラスにおけるサポートベクトルを同時に選択してくる必要がある.

アクティブ集合による学習

まず, 全クラスから p 個の教師データをランダムにサポートベクトルとして選択し, 対応するクラスのワーキングセット W_i ($i = 1, \dots, n$) に入れる. 次に, 選択されなかった教師データを集合 F_i に入れる. そして, ワーキングセット内のデータのみを用いて, 式 (3.43) を解く事により, α_{ij} ($i = 1, \dots, n, j \in W_i$) を導き, \mathbf{w}_i ($i = 1, \dots, n$) を式 (3.40) により決定する. そして, \mathbf{w}_i ($i = 1, \dots, n$) を式 (2.23) に代入することにより決定関数を求める. 次に, ワーキングセット W_i 内の \mathbf{x}_j ($y_j \neq i$) から, $D_{y_j}(\mathbf{x}_j) - D_i(\mathbf{x}_j) > 1$ となる教師データを選択し, 集合 F_i ($i = 1, \dots, n$) に移す. また集合 F_i から $D_{y_j}(\mathbf{x}_j) - D_i(\mathbf{x}_j) < 1$ を満たす教師データ \mathbf{x}_j ($y_j \neq i$) をこの値が大きい順に選択し, 最大 p 個ワーキングセット内に入れる. これらの処理をワーキングセット W_i 内の教師データが変動しないまで続ける. 最終的に, ワーキングセット内のデータはサポートベクトルのみとなる.

同時定式化を用いた SSL2-SVM の学習アルゴリズムを以下に示す.

アルゴリズム 6

- Step 1** 五分割交差検定法を用いて, γ, κ, C 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って教師データから一次独立なデータを選択する.
- Step 3** Step 2 で求めた一次独立なデータで構成した標本特徴空間への写像関数を生成する. また, 式 (2.14) よりクラス i ($i = 1, \dots, n$) の固有ベクトル φ_{ik} と固有値 λ_{ik} を導く.
- Step 4** Step 3 で求めた固有値を各クラスそれぞれ大きい値から順に並び替える. そして固有値の総和との比率が Step 1 で決定した κ より大きくなるまで順番に固有値を足していく. 足された固有値に対応する固有ベクトルをクラス i ($i = 1, \dots, n$) 部分空間を形成する軸とし, 次元数 r_i を決定する.
- Step 5** 式 (2.22) より $\mathbf{f}_i(\mathbf{x}_j)$ ($i = 1, \dots, n, j = 1, \dots, M$) を計算する.
- Step 6** 全教師データからランダムに p 個取り出し, ワーキングセット W_i ($i = 1, \dots, n$) にそれぞれ入れ, 残りのデータは集合 F_i ($i = 1, \dots, n$) にそれぞれ入れる.

Step 7 式 (3.43) を解くことにより α_{ij} を求め、式 (3.40) より \mathbf{w}_i ($i = 1, \dots, n$) を求める。

Step 8 ワーキングセット W_i 内から $\alpha_{ij} < 0$ となる教師データを取り出し集合 F_i に入れ、集合 F_i 内のデータの決定関数を求め、 $y_j D(\mathbf{x}_j) < 1$ となるデータのうち最大 p 個を集合 W_i に入れる。 i 個のワーキングセットのうちどれかひとつでも変動があれば Step 8 に戻り、なければ学習を終了する。

以上のように、L2-SVM の最適化問題を部分空間法に取り入れることにより、各クラス部分空間の各重みを最適化できる。しかしながら、同時定式化方式を用いるため、 $\sum_{i=1}^n |S_i| \times \sum_{i=1}^n |S_i|$ 行列を解く必要があり、多クラス問題や大規模データセットにおいて計算コストが膨大となると推測できる。この問題点を解消するために各クラスごとの重みを一つずつ最適化する一対他方式を適用した部分空間法に基づく L2 サポートベクトルマシンを提案する。この手法を SSL2-SVM (O) と呼ぶ。次節に SSL2-SVM (O) の詳細を述べる。

3.5.2 一対他方式を適用した SSL2-SVM

前節と同様に式 (2.23) を一対他方式の L2-SVM の最適化問題に導入すると以下のように表せる。

$$\text{minimize } Q(\mathbf{w}, \mathbf{b}, \boldsymbol{\xi}) = \frac{1}{2} \|\mathbf{w}_i\|^2 + \frac{C}{2} \sum_{j=1}^M \xi_{ji}^2 \quad (3.46)$$

$$\text{subject to } y_j (\mathbf{w}_i \mathbf{f}_i(\mathbf{x}_j) + b_i) \geq 1 - \xi_{ji} \quad \text{for } j = 1, \dots, M \quad (3.47)$$

非負のラグランジュ乗数 α_{ij} を導入すると式 (3.46), (3.47) を次式のように変換できる。

$$\text{maximize } Q(\boldsymbol{\alpha}_i) = \sum_{j=1}^M \alpha_{ij} + \frac{1}{2} \sum_{j,k=1}^M \alpha_{ij} \alpha_{ik} y_j y_k \times \left(\mathbf{f}_i^\top(\mathbf{x}_j) \mathbf{f}_i(\mathbf{x}_k) + \frac{\delta_{jk}}{C} \right) \quad (3.48)$$

$$\text{subject to } \sum_{j=1}^M y_j \alpha_{ij} = 0 \quad (3.49)$$

$$\alpha_{ij} \geq 0 \quad \text{for } i = 1, \dots, n, j = 1, \dots, M \quad (3.50)$$

これにより、求めるべきクラス i の重みベクトル \mathbf{w}_i は

$$\mathbf{w}_i = \sum_{j=1}^M \alpha_j y_j \mathbf{f}_i(\mathbf{x}_j) \quad (3.51)$$

となる。

LS-SVM とは違い、L2-SVM では一部の教師ベクトルが SV となるため、全教師ベクトルを用いて学習を行う必要はない。クラス i の決定関数における SV の集合 S_i のみ

を用いて学習を行う場合、ある一つのSVに対応するデータの添え字 j を $j = s (s \in S)$ とおき、 $\alpha_{sv} = \{\alpha_{ij} \mid j \neq s, j \in S\}$ とすると、式 (3.49) を式 (3.48) に代入すると以下のようなになる。

$$\max \quad Q(\alpha_{sv}) = \mathbf{c}_{sv}^T \alpha_{sv} - \frac{1}{2} \alpha_{sv}^T K_{sv} \alpha_{sv} \quad (3.52)$$

$$\alpha_{sv} \geq \mathbf{0} \quad (3.53)$$

ここで、 $\mathbf{c}_{sv}, K_{sv} = \{K_{svij} \mid i, j \neq s, i, j \in S_k\}$ は以下のように表せる。

$$\mathbf{c}_{sv} = 1 - y_s y_i \quad \text{for } i \neq s, \quad i \in S_k \quad (3.54)$$

$$K_{svij} = y_i y_j (\mathbf{f}_k(\mathbf{x}_i) - \mathbf{f}_k(\mathbf{x}_s))(\mathbf{f}_k(\mathbf{x}_j) - \mathbf{f}_k(\mathbf{x}_s)) + \frac{1 + \delta_{ij}}{C} \quad (3.55)$$

$$\text{for } i, j \neq s, \quad i, j \in S_k \quad (3.56)$$

これより、式 (3.57) は以下のように変換できる。

$$\mathbf{w}_i = \sum_{j \in S_i} \alpha_{ij} y_j \mathbf{f}_i(\mathbf{x}_j) \quad (3.57)$$

式 (3.52) の解は α_{sv} の勾配が 0 となる点であるため、 α_{sv} は以下のように表せる。

$$\alpha_{sv} = K_{sv}^{-1} \mathbf{c}_{sv} \quad (3.58)$$

ただし、SV 以外の教師データを集合に入れて計算を行えば、式 (3.53) を満たさなくなる。集合外に SV が存在すれば、そのようなデータは決定関数から見つけ出すことが可能である。クラス i における決定関数 (類似度) は式 (3.57) を用いて次式のように表せる。

$$D_i = \sum_{j \in S_i} \alpha_{ij} y_j \mathbf{f}_i(\mathbf{x}_j) \mathbf{f}_i(\mathbf{x}) + b_i \quad (3.59)$$

はじめに集合 S_i 内のデータをランダムに選択し、集合 S_i 内のデータは式 (3.58) よりクラス i に対応する決定関数における SV となり得るデータのみを選択する。式 (3.59) を用いて集合 S_i 外の SV を選択し、選択されたデータを再び集合 S_i に入れ替え学習を行う。これらの処理を繰り返し行えば、最終的に集合 S_i 内は全 SV のみで構成される。その後、入れ替えは行われなため、クラス i 部分空間の重みの最適化のための学習を終了する。これをクラス数 n 回同じ処理を繰り返す。これにより、全教師データを学習に用いずに少数のデータのみで識別の観点から最適な重みを決定できる。

SSL2-SVM (O) のアルゴリズムを以下に示す。

アルゴリズム 7

Step 1 五分割交差検定法を用いて、 γ, κ, C の値を決定する。 $i = 1$ とおく。

- Step 2** Step 1 で求めた γ を用いてコレスキー分解によって教師データから一次独立なデータを選択する.
- Step 3** Step 2 で求めた一次独立なデータを用いて式 (2.15) を計算し, 標本特徴空間への写像関数 $h(\mathbf{x})$ を生成する.
- Step 4** 式 (2.14) よりクラス i の固有ベクトル φ_{ik} と λ_{ik} を求める.
- Step 5** Step 4 で求めた固有値を各クラスそれぞれ大きい値から順に並び替える. そして固有値の総和との比率が Step 1 で決定した κ より大きくなるまで順番に固有値を足していく. 足された固有値に対応する固有ベクトルをクラス i 部分空間を形成する軸とし, 次元数 r_i を決定する.
- Step 6** 式 (2.22) より $\mathbf{x}_j (j = 1, \dots, M)$ を計算する.
- Step 7** 全教師データからランダムに p 個取り出し, ワーキングセット W_i に入れ, 残りのデータは集合 F_i に入れる.
- Step 8** 式 (3.52) を解くことにより \mathbf{w}_i とバイアス項 b_i をそれぞれ求める.
- Step 9** 集合 W_i 内から $\alpha_{ij} < 0$ となる教師データを取り出し集合 F_i に入れ, 集合 F_i 内のデータの決定関数 (3.59) を求め, $y_j D(\mathbf{x}_j) < 1$ となるデータのうち最大 p 個を集合 W_i に入れる. 変動がなければ, Step 10 に進み, あれば Step 8 に戻る.
- Step 10** もし $i \neq n$ ならば, $i = i + 1$ とおき Step 2 に戻る. $i = n$ ならば, 学習を終了する.

3.6 計算機実験

本節では, ベンチマークデータセット^{29, 30)}を用いて SS-SVM のそれぞれのモデルを KSM と比較することにより KSM に対して SS-SVM における各部分空間を構成する軸における重みを最適化により識別性能が向上するか評価する. ただし, SSLS-SVM, SSLP-SVM, SSL2-SVM のそれぞれのモデルに対する評価を別々に行う. また, それぞれのモデルにおける同時定式化方式と一対他方式の識別性能と学習時間の比較を行う. これらの評価と同時に, 五分割交差検定法により汎化性能の予測ができているかを確認するため, バリデーションセットにおける評価も行う. それぞれの SS-SVM の識別性能を比較することにより, 最も汎化能力の高くなるモデルを確認する. さらに通常の SVM と比較することにより, 汎化能力が高い識別手法であるといえるかを確認する. 最後に, 識別部の解析を行い, SVM に対しての SS-SVM の利点を考察する.

3.6.1 計算機実験における環境

計算機の仕様

本論文における計算機実験において用いる計算機の仕様は OS : Windows 7 (32bit), CPU : Intel (R) Core (TM) i5 3.33GHz, メモリ : 2.00GB である.

計算機実験に用いるベンチマークデータセット

各手法において, それらの有効性を評価するために計算機実験には2クラスまたは多クラスのベンチマークデータセットを用いる. 表3.1に2クラスベンチマークデータセット, 表3.2に多クラスベンチマークデータセットを示す. 表3.1, 3.2における Data はデータセット名, Inputs は次元数, Training は教師データ数, Test はテストデータ数, Sets はファイル数を示し, 表3.2の Classes はクラス数を示す.

表 3.1 2クラスベンチマークデータセット

Data	Inputs	Training	Test	Sets
Banana	2	400	4900	100
B. cancer	9	200	77	100
Diabetes	8	468	300	100
German	20	700	300	100
Heart	13	170	100	100
Image	18	1300	1010	20
Ringnorm	20	400	7000	100
F. solar	9	666	400	100
Splice	60	1000	2175	20
Thyroid	5	140	75	100
Titanic	3	150	2051	100
Twonorm	20	400	7000	100
Waveform	21	400	4600	100

表 3.2 多クラスベンチマークデータ

Data	Inputs	Classes	Trn	Tst
Iris	4	3	75	75
Numeral	12	10	810	820
Blood cell	13	12	3097	3100
Hiragana-50	50	39	4610	4610
Hiragana-105	105	38	8375	8356
Hiragana-13	13	38	8375	8356

3.6.2 KSMのパラメータ設定

表 3.3 に通常の KSM における決定すべきカーネルパラメータ γ , 寄与率の閾値 κ は $\kappa = \{80, 85, 90, 95, 99, 99.9\}$, $\gamma = \{0.01, 0.05, 0.1, 0.5, 1, 1.5, 3, 5, 10, 15, 20, 30, 50, 100, 200, 500, 10^3\}$ から五分割交差検定法を用いて決定する. ここで重みをすべて等しく設定する KSM を KSM (1), 対応する固有値を設定する KSM を KSM (E) とする.

表 3.3 KSM (1) と KSM (E) におけるパラメータ値

Data	KSM (1)		KSM (E)	
	κ	γ	κ	γ
Banana	99.9	15	99.9	15
B. cancer	99.9	0.5	85	3
Diabetes	99.9	3	80	5
German	85	3	85	10
Heart	95	1.5	80	3
Image	99	15	99.9	15
Ringnorm	99.9	0.1	99.9	15
F. solar	80	5	95	10
Splice	99	1	99	15
Thyroid	85	15	80	15
Titanic	80	0.1	95	5
Twonorm	80	0.1	80	3
Waveform	80	1.5	80	5
Iris	85	30	80	100
Numeral	95	10^3	99	10^3
Blood cell	95	10^3	99	10^3
Hiragana-50	95	200	90	500
Hiragana-105	80	500	99.9	10^3
Hiragana-13	95	10^3	80	500

3.6.3 SSLS-SVM の評価 (KSM との比較)

パラメータ設定

SSLS-SVM と KSM を比較するため, KSM の最適なカーネルパラメータ γ と累積寄与率の閾値 κ を設定し, それらのパラメータ値を用いて, SSLS-SVM の最適なマージンパラメータ C を五分割交差検定法により決定する. これにより, SSLS-SVM による重みの最適化処理の有効性を確認できる. $C = \{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 500, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4, 10^5\}$ から五分割交差検定より最適な値を決定する. ここで表 3.4 の C_1, C_2 はそれぞれ KSM (1), KSM (E) における五分割交差検定法により決定された最適なカーネルパラメータ γ ,

累積寄与率の閾値 κ を設定したときの五分割交差検定法により選択された2クラスベンチマークデータセットにおけるそれぞれのマージンパラメータの値を示す。

表 3.4 パラメータ値 (SSLS-SVM)

Data	SSLS-SVM		SSLS-SVM (O)	
	C_1	C_2	C_1	C_2
Banana	50	50	1	1
B. cancer	10	0.1	5	1
Diabetes	10^3	1	50	0.5
German	10^4	0.5	100	500
Heart	10	0.1	500	0.5
Image	10^5	10^5	500	10^3
Ringnorm	10^4	500	1	0.1
F. solar	10^4	500	1	10
Splice	10^4	5×10^3	10^3	0.1
Thyroid	500	500	5	5
Titanic	500	0.1	0.5	5
Twonorm	50	10	1	0.1
Waveform	50	5×10^3	1	0.1

KSM (1) との比較 (バリデーションセット)

表 3.5 に 2 クラスベンチマークデータセットにおける KSM (1) と SSLS-SVM と SSLS-SVM (O) を用いたときの最初の 5 個の教師データセットのバリデーションデータセットの平均認識率 (%) と標準偏差を示す。また、それぞれ KSM (1) における累積寄与率の閾値 κ , カーネルパラメータ γ を用いていることを示すためにそれぞれの提案手法を SSLS-SVM (1), SSLS-SVM (O,1) と表中に示す。各データセットに対して、最も高い平均認識率を太字で示し, SSLS-SVM (1), SSLS-SVM (O,1) において KSM (1) より高い認識率の前に (*) を付ける。

表 3.5 より, SSLS-SVM (1) は 7 個のデータセットにおいて, KSM (1) と比べ高い平均認識率を示している。また SSLS-SVM (O,1) においても KSM (1) と比べ 5 個のデータにおいて KSM (1) より高い平均認識率を示している。さらに, Ringnorm データセットにおいて, SSLS-SVM, SSLS-SVM (O,1) とともに平均認識率が 37 (%) 以上高い平均認識率を示しており, 識別性能が非常に向上しているといえる。しかしながら, Splice データセットにおいて, SSLS-SVM (O,1) によって重みを最適化することによって平均認識率が低下している。SSLS-SVM (1) と SSLS-SVM (O,1) を比較すると, Banana データを除いて SSLS-SVM がよい結果を得られ, SSLS-SVM において同時定式化方式が一対他方式に比べて有効であると推測できる。ただし, これらはモデ

ル選択における汎化能力の推測であり、実際の汎化能力は次節にて比較する。

表 3.5 バリデーショナルセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (1) の比較)

Data	KSM (1)	SSLS-SVM (1)	SSLS-SVM (O,1)
Banana	89.8 ± 3.1	89.7 ± 3.1	*90.1 ± 3.2
B. cancer	72.6 ± 4.6	*74.4 ± 4.3	*72.8 ± 4.3
Diabetes	73.5 ± 2.6	*73.8 ± 2.8	72.7 ± 3.9
German	73.9 ± 3.0	72.6 ± 3.3	68.4 ± 3.0
Heart	82.1 ± 3.9	*83.2 ± 4.4	81.7 ± 5.4
Image	95.7 ± 0.9	94.8 ± 1.0	93.7 ± 1.2
Ringnorm	52.8 ± 1.8	*97.8 ± 1.7	*89.9 ± 2.5
F. solar	65.2 ± 2.8	*66.1 ± 3.5	62.6 ± 4.6
Splice	87.4 ± 2.5	86.3 ± 2.1	57.2 ± 2.2
Thyroid	95.6 ± 2.5	*96.3 ± 2.4	*96.2 ± 2.0
Titanic	79.6 ± 7.8	79.0 ± 7.5	76.0 ± 5.9
Twonorm	97.1 ± 1.5	*97.4 ± 1.4	*97.3 ± 1.4
Waveform	89.6 ± 3.0	82.5 ± 3.2	81.9 ± 3.4

KSM (1) との比較 (テストデータ)

表 3.6 に 2 クラスベンチマークデータセット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。各データセットに対して、最も高い平均認識率を太字で示し、SSLS-SVM (1), SSLS-SVM (O,1) において KSM (1) より高い認識率の前に (*) を付ける。

表 3.6 より、SSLS-SVM (1) は 7 個のデータセットにおいて、KSM (1) と比べ高い平均認識率を示している。また SSLS-SVM (O,1) においても KSM (1) と比べ 5 個のデータにおいて KSM (1) より高い平均認識率を示している。特に、Ringnorm データセットにおいて SSLS-SVM (1), SSLS-SVM (O,1) 共に 15 (%) 以上高い平均認識率を示している。これにより Ringnorm データセットにおいて LS-SVM による重みの最適化により汎化能力が向上しているといえる。しかしながら、German, Splice データセットに対しての SSLS-SVM (O,1) が KSM (1) と比べて非常に劣る平均認識率を示している。これは、一対他方式では他のクラスにおける重みを考慮していないためであると考えられる。ほとんどの場合、SSLS-SVM (1) は SSLS-SVM (O,1) に比べよい結果を得られ、SSLS-SVM において同時定式化手法は一対他方式方式に比べ汎化能力が高いといえる。

表 3.5, 3.6 を比較すると、バリデーショナルセットにおいて最もよい結果を示すときテストデータでも最もよい結果を示し、バリデーショナルセットにおいて悪ければ、テスト

データでも悪い結果を示している。このように、五分割交差検定によりある程度の汎化能力の推測が行えていることがわかる。

また, KSM (1) と比べて非常に劣る平均認識率を示したデータセットが複数存在している。これは累積寄与率 κ , マージンパラメータ C の値を KSM (1) と等しく設定していることが原因であると考えられる。そのため, SSLS-SVM, SSLS-SVM (O) のそれぞれを用いて KSM (1) とは別に最適化問題を解いた場合において汎化能力が向上するかどうかを 3.6.4 節にて計算機実験により確認する。

表 3.6 テストデータセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (1) の比較)

Data	KSM (1)	SSLS-SVM (1)	SSLS-SVM (O,1)
Banana	88.6 ± 0.6	*88.9 ± 0.6	*88.9 ± 0.6
B. cancer	75.0 ± 4.3	73.8 ± 4.6	72.2 ± 4.6
Diabetes	73.5 ± 1.8	72.4 ± 2.2	72.5 ± 1.9
German	75.1 ± 2.2	74.0 ± 2.2	69.7 ± 2.8
Heart	80.5 ± 3.3	*82.6 ± 3.9	*82.7 ± 3.5
Image	96.3 ± 0.6	95.1 ± 0.6	94.5 ± 0.5
Ringnorm	76.6 ± 11.2	*97.6 ± 0.3	*91.5 ± 0.8
F. solar	65.1 ± 1.8	*66.9 ± 1.6	62.4 ± 2.2
Splice	87.6 ± 0.8	86.2 ± 1.0	57.4 ± 1.8
Thyroid	95.6 ± 2.1	*95.9 ± 2.1	95.5 ± 2.2
Titanic	76.6 ± 1.2	*77.2 ± 0.8	*76.9 ± 0.9
Twonorm	97.6 ± 0.1	*97.7 ± 0.1	*97.7 ± 0.1
Waveform	88.5 ± 0.6	81.2 ± 0.9	81.7 ± 0.9

KSM (E) との比較 (バリデーションセット)

表 3.7 に 2 クラスベンチマークデータセットにおける KSM (E) と SSLS-SVM と SSLS-SVM (O) を用いたときのバリデーションデータセットにおける平均認識率 (%) と標準偏差を示す。各データセットに対して、最も高い平均認識率を太字で示す。また、それぞれ KSM (E) における累積寄与率の閾値 κ , カーネルパラメータ γ を用いていることを示すためにそれぞれの提案手法を SSLS-SVM (E), SSLS-SVM (O,E) と表中に示す。各データセットに対して、最も高い平均認識率を太字で示し、SSLS-SVM (E), SSLS-SVM (O,E) において KSM (E) より高い認識率の前に (*) を付ける。

表 3.7 より, SSLS-SVM (E) は 7 個のデータセットにおいて KSM (E) と比べて高い平均認識率を示している。また, SSLS-SVM (E) においても 6 個のデータセットにおいて KSM (E) と比べて高い平均認識率を示している。Image データセットにおいて SSLS-SVM (E), SSLS-SVM (O,E) は共に KSM (E) は極めて高い平均認識率を示

している. 表 3.5, 3.7 を比較すると, KSM (E) より KSM (1), SSLS-SVM (1) より SSLS-SVM (E), SSLS-SVM (O,1) より SSLS-SVM (O,E) の識別性能が高いことがわかる. ただし, これらはモデル選択における汎化能力の推測であり, 実際の汎化能力は次節にて比較する.

表 3.7 バリデーションセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (E) の比較)

Data	KSM (E)	SSLS-SVM (E)	SSLS-SVM (O,E)
Banana	88.3 ± 2.9	*89.7 ± 3.1	*90.1 ± 3.2
B. cancer	75.7 ± 3.5	75.1 ± 3.1	71.0 ± 4.6
Diabetes	73.8 ± 3.0	70.6 ± 3.2	70.8 ± 3.0
German	71.6 ± 2.8	71.0 ± 2.7	71.0 ± 3.0
Heart	82.9 ± 3.9	*84.3 ± 4.7	82.5 ± 4.3
Image	88.1 ± 1.4	*94.9 ± 1.2	*94.1 ± 1.2
Ringnorm	62.5 ± 4.0	*63.2 ± 3.3	*63.4 ± 3.2
F. solar	64.4 ± 3.7	*65.1 ± 3.8	* 65.1 ± 3.7
Splice	72.0 ± 3.0	*72.1 ± 3.1	*72.4 ± 2.9
Thyroid	96.0 ± 2.7	95.3 ± 2.5	96.0 ± 2.4
Titanic	79.4 ± 7.0	78.4 ± 8.2	78.3 ± 8.1
Twonorm	97.2 ± 1.4	97.2 ± 1.2	*97.3 ± 1.2
Waveform	89.0 ± 2.8	*89.8 ± 2.2	84.2 ± 3.5

KSM (E) との比較 (テストデータ)

表 3.8 に 2 クラスベンチマークデータセット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. 各データセットに対して, KSM (E), SSLS-SVM (E), SSLS-SVM (O,E) の結果の中で最も高い平均認識率を太字で示す. SSLS-SVM (E), SSLS-SVM (O,E) において KSM (E) より高い認識率の前に (*) を付ける.

SSLS-SVM (E) は 9 個のデータセットにおいて, KSM (E) と比べて高い平均認識率を示している. SSLS-SVM (O,E) は 8 個のデータセットにおいて KSM (E) に比べて高い平均認識率を示している. 特に, Image データセットにおいて SSLS-SVM (E), SSLS-SVM (O,E) とともに KSM (E) と比べ非常に高い平均認識率示している. また, SSLS-SVM (E), SSLS-SVM (O,E) とともに KSM (E) と比べ低い平均認識率を示しているデータセットにおいても, ほぼ同程度の平均認識率を示している. これより, KSM (E) に比べ, 同じパラメータ値 (κ, γ) を設定した際の汎化能力が高いことが明らかである. また, SSLS-SVM (E) と SSLS-SVM (O,E) は Waveform データセット, B. cancer データセットを除けばほぼ同程度の平均認識率を示しており, 若干 SSLS-SVM (E) の方が汎化能力が高いといえる.

また、バリデーションセットを用いたときの平均認識率とほぼ近い精度が得られたことから五分割交差検定法によるモデル選択により汎化能力の推測ができていているといえる。

表 3.8 テストデータセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (E) の比較)

Data	KSM (E)	SSLS-SVM (E)	SSLS-SVM (O,E)
Banana	87.8±0.6	*88.9±0.6	*88.9±0.6
B. cancer	75.1±4.4	71.2±4.5	69.3±4.8
Diabetes	71.7±2.3	71.1±2.3	71.1±2.2
German	73.6±2.1	72.7±2.4	73.1±2.1
Heart	82.4±3.6	*82.8±3.7	*82.6±3.6
Image	88.1±1.0	*95.4±0.6	*95.1±0.6
Ringnorm	64.1±2.4	*64.3±2.3	*64.7±1.3
F. solar	63.5±4.0	*65.1±1.6	*65.5±1.8
Splice	71.8±1.5	*72.1±1.6	72.3±1.4
Thyroid	95.1±2.4	*95.7±2.1	*95.4±2.2
Titanic	77.3±0.3	77.3±0.7	*77.4±0.7
Twonorm	97.0±0.5	*97.4±0.2	*97.4±0.2
Waveform	88.0±1.1	*88.2±0.6	84.0±1.6

3.6.4 独自でモデル選択を行った SSLS-SVM の評価

前述のように、SSLS-SVM, SSLS-SVM (O) における累積寄与率の閾値 κ , カーネルパラメータ γ は KSM (1), KSM (E) のそれぞれで決定された値を選択していたが、SSLS-SVM, SSLS-SVM (O) において最適な選択ではない。そこで、すべてのパラメータ値を独自に五分割交差検定法により決定した際、それぞれの KSM と比較して汎化能力が向上しているかどうかを評価する。また本実験では、多クラス問題についても評価する。

パラメータ値の選択候補は前節までの候補と同じにしており、表 3.9 に選択された SSLS-SVM, SSLS-SVM (O) のパラメータ値を示す。表 3.9 と表 3.3, 3.4 と比べると、2 クラスベンチマークデータセットにおいて選ばれたパラメータ値がすべて異なっており、前節までの SSLS-SVM, SSLS-SVM (O) は汎化能力が最大となるためのモデル選択が行えていなかったと推測できる。

表 3.9 パラメータ値 (独自にモデル選択を行った SSLS-SVM)

Data	SSLS-SVM			SSLS-SVM (O)		
	κ	γ	C	κ	γ	C
Banana	99.9	15	50	99	15	1
B. cancer	99.9	0.5	10	99.9	0.5	5
Diabetes	95	3	10^3	99.9	3	50
German	80	3	10^4	85	10	500
Heart	80	1	1	80	3	0.5
Image	99	15	10^4	99.9	15	10^3
Ringnorm	99.9	1	10^4	80	0.1	1
F. solar	80	5	10^4	95	5	5
Splice	99	1	10^4	99	5	500
Thyroid	85	15	500	90	10	5
Titanic	80	5	0.1	95	5	5
Twonorm	80	0.1	50	80	0.5	0.1
Waveform	80	3	10^3	85	5	50
Iris	80	50	1	80	50	1
Numeral	99.9	10	10^3	99	50	5
Blood cell	99.9	20	500	99	500	5
Hiragana-50	99.9	15	100	95	100	50
Hiragana-105	99.9	10	100	99	50	10
Hiragana-13	99.9	30	10^4	99.9	500	10

表 3.10 にテストデータセットでの平均認識率と標準偏差を示す。各データセットに対して、最大となる平均認識率を太字で示す。また、各データセットに対して、最も高い平均認識率を太字で示し、SSLS-SVM, SSLS-SVM (O) において KSM (1) かつ KSM (E) より高い認識率の前に (*) を付ける。

2 クラス問題

2 クラス問題において、SSLS-SVM は 2 つの KSM と比べて 7 個のデータセットにおいて高い平均認識率を示している。ただし、KSM (1) のみと比較すると 8 個、KSM (E) のみと比較すると 11 個のデータセットにおいて高い平均認識率を示している。特に Ringnorm データセット、F. solar データセットにおいては汎化能力が極めて向上している。平均認識率が劣るデータセットにおいてもほぼ同程度であるといえる。また、SSLS-SVM (O) は 2 つの KSM と比べて 4 個のデータセットにおいて高い平均認識率を示している。ただし、KSM (1) のみと比較すると 6 個、KSM (E) のみと比較すると 8 個のデータセットにおいて高い平均認識率を示している。特に Ringnorm データセットにおいては汎化能力が極めて向上している。平均認識率が劣るデータセットにおいても Splice データセットを除いてはほぼ同程度であると確認できる。

表 3.6, 3.8 の結果と比べると大半の 2 クラス問題において SSLS-SVM, SSLS-SVM

(O) とともに平均認識率が向上している。これらのことから、通常の KSM に比べて SSL-SVM, SSL-SVM (O) の汎化能力が高いことがわかり、重みの最適化の有効性を示すことが確認できる。

多クラス問題

多クラス問題において、SSL-SVM はすべてのデータセットにおいて認識率が他の手法と比べ高くなっており、多クラス問題においても有効性を示している。しかしながら、SSL-SVM (O) では、Iris データセットを除いてはいずれかの KSM よりも認識率が低くなっている。この原因として、他のクラスの重みを考慮していないためであると考えられる。これにより、SSL-SVM において、高い汎化能力を要求する際は同時定式化方式を適用する方がよいといえる。

表 3.10 テストデータセットにおける平均認識率 (%) と標準偏差 (独自にモデル選択を行った SSL-SVM)

Data	KSM (1)	KSM (E)	SSL-SVM	SSL-SVM (O)
Banana	88.6±0.6	87.8±0.6	*88.9±0.6	88.7±0.6
B. cancer	75.0±4.3	75.1±4.4	73.2±4.6	72.2±4.6
Diabetes	73.5±1.8	71.7±2.3	72.0±2.3	71.7±2.4
German	75.1±2.2	73.6±2.1	73.8±2.2	71.5±2.2
Heart	80.5±3.3	82.4±3.6	*82.6±3.7	*82.6±3.5
Image	96.3±0.6	88.1±1.0	95.0±0.6	95.1±0.6
Ringnorm	76.6±0.6	64.1±2.4	*93.1±0.8	*91.8±0.8
F. solar	65.1±1.8	63.5±4.0	*66.9±1.6	*65.5±1.9
Splice	87.6±0.8	71.8±1.5	86.2±1.0	77.7±3.3
Thyroid	95.6±2.1	95.1±2.4	*95.9±2.1	95.5±2.2
Titanic	76.6±1.2	77.3±0.3	76.9±1.0	77.3±0.2
Twonorm	97.6±0.1	97.0±0.5	*97.7±0.1	*97.7±0.1
Waveform	88.5±0.6	88.0±1.1	*89.1±0.8	87.1±1.0
Iris	93.3	92.0	*96.0	*94.7
Numeral	99.0	99.0	*99.4	98.7
Blood cell	91.3	90.1	*92.9	90.9
Hiragana-50	95.3	95.1	*95.5	95.1
Hiragana-105	99.7	99.7	*99.9	99.7
Hiragana-13	98.4	98.3	*98.9	97.8

学習時間の比較

表 3.11 に各 KSM, 各 SSL-SVM における部分空間生成と重みの最適化 (各 SSL-SVM のみ) に要する平均学習時間を示す。

2 クラス問題において、大半のデータセットにおいて SSL-SVM が SSL-SVM (O) と比べて学習が同等または高速であることが明らかである。この原因として用いたデー

タセットの教師データ数が少ないことが考えられる。また、大半のデータセットにおいて学習コストが各KSMと比べ同程度である。しかしながら、ImageデータセットやSpliceデータセットなどの大規模データセットにおいてもSSL-SVMがSSL-SVM(O)に比べ高速になっている。これは各クラス部分空間の次元が非常に小さくなることが原因である。これらの結果より2クラス問題においては、一対他方式を用いることによる高速化があまりなされていないといえる。

多クラス問題において、Irisデータセットを除いてはSSL-SVM(O)がSSL-SVMと比べて学習時間が短い。特に、Hiragana-105データセット、Hiragana-13データセットなどのようにデータ数とクラス数が多いデータセットにおいて1000(s)以上の学習時間の短縮が行われている。これより、データ数とクラス数が多いデータセットにおいて学習コストの面ではSSL-SVM(O)が非常に有効であるといえる。また、これらのデータセットにおけるSSL-SVMまたはSSL-SVM(O)の学習時間の大半は部分空間生成に要していることが各KSMの学習時間と比較することにより明らかである。そこで、部分空間生成に要する学習コストの削減が重要になる。

表 3.11 SSL-SVMにおける学習時間 (s)

Data	KSM (1)	KSM (E)	SSL-SVM	SSL-SVM (O)
Banana	0.62	0.59	0.91	0.71
B.cancer	0.10	0.09	0.10	0.12
Diabetes	1.12	1.10	1.12	1.31
German	3.13	3.61	4.11	5.01
Heart	0.07	0.06	0.07	0.10
Image	21.9	22.0	22.5	31.6
Ringnorm	0.72	0.62	1.06	0.84
F.solar	1.31	1.33	1.35	1.34
Splice	10.3	10.9	13.1	18.2
Thyroid	0.03	0.03	0.04	0.05
Titanic	0.01	0.02	0.02	0.04
Twonorm	0.71	0.72	0.70	0.82
Waveform	0.74	0.74	0.75	1.05
Iris	0.01	0.01	0.01	0.01
Numeral	43.8	42.9	44.7	44.2
Blood cell	262	273	319	268
Hiragana-50	921	891	1027	930
Hiragana-105	2562	2460	4676	2583
Hiragana-13	1985	1962	3440	2214

3.6.5 SSLP-SVM の評価 (KSM との比較)

パラメータ設定

前節と同様に SSLP-SVM と KSM を比較するため, KSM の最適なカーネルパラメータ γ と累積寄与率の閾値 κ を設定し, それらのパラメータ値を用いて, SSLP-SVM の最適なマージンパラメータ C を五分割交差検定法により決定する. これにより, SSLP-SVM による重みの最適化処理の有効性を確認できる. また各パラメータ値の選択候補は前節までの計算機実験で同じにする. 表 3.12 に 2 クラスベンチマークデータセットにおける SSLP-SVM, SSLP-SVM (O) の選択されたパラメータ値を示す.

表 3.12 パラメータ値 (SSLP-SVM)

Data	SSLP-SVM		SSLP-SVM (O)	
	C_1	C_2	C_1	C_2
Banana	5	5	5	5
B.cancer	1	1	5	5
Diabetes	10	5	10	1
German	5	50	5	0.1
Heart	5	0.5	1	1
Image	100	50	100	100
Ringnorm	10^3	1	5	100
F. solar	5	5	1	0.05
Splice	10	1	0.1	0.1
Thyroid	5	5	1	1
Titanic	5	0.1	10^{-3}	1
Twonorm	10	5	0.05	0.1
Waveform	50	1	1	100

KSM (1) との比較 (バリデーションセット)

表 3.13 に 2 クラスベンチマークデータセットにおける KSM (1) と SSLP-SVM を用いたときのバリデーションデータセットの平均認識率 (%) と標準偏差を示す. 各データセットに対して, 最も高い平均認識率を太字で示す. また, それぞれ KSM (1) における累積寄与率の閾値 κ , カーネルパラメータ γ を用いていることを示すためにそれぞれの手法を SSLP-SVM (1), SSLP-SVM (O,1) と表中に示す. 各データセットに対して, 最も高い平均認識率を太字で示し, SSLP-SVM (1), SSLP-SVM (O,1) において KSM (1) より高い認識率の前に (*) を付ける.

表 3.13 より, SSLP-SVM (1) は 6 個のデータセットにおいて KSM (1) と比べて高い平均認識率を示している. SSLP-SVM (O,1) は KSM (1) と比べて B. cancer デー

タセットと Ringnorm データセットのみにおいて高い平均認識率を示している。しかしながら、大半のデータセットにおいて KSM (1) と比べて極めて低い平均認識率を示している。この原因として SSLP-SVM (O,1) は KSM (1) におけるパラメータ値では最適な解が導けないと考えられる。3.6.6 節に独自にパラメータ値を決定した計算機実験により確認する。これらより、バリデーションセットにおいて SSLP-SVM (1), SSLP-SVM (O,1) は KSM (1) に比べて汎化能力が低くなると推測される。

表 3.13 バリデーションセットを用いたときの平均認識率 (%) と標準偏差 (SSLP-SVM と KSM (1) の比較)

Data	KSM (1)	SSLP-SVM (1)	SSLP-SVM (O,1)
Banana	89.8 ± 3.1	89.0 ± 3.2	87.5±3.4
B. cancer	72.6 ± 4.6	*74.1 ± 5.0	* 74.8 ±2.9
Diabetes	73.5 ± 2.6	71.6 ± 3.7	71.4±4.0
German	73.9 ± 3.0	70.2 ± 5.9	64.5± 10.2
Heart	82.1 ± 3.9	* 82.2 ± 4.0	78.7±5.5
Image	95.7 ± 0.9	83.2 ± 14.4	90.2 ± 6.1
Ringnorm	52.8 ± 1.8	* 98.4 ± 1.2	*97.9±1.4
F. solar	65.2 ± 2.8	63.0 ± 6.2	54.4±10.1
Splice	87.4 ± 2.5	86.2 ± 1.4	69.2 ± 8.1
Thyroid	95.5 ± 2.5	* 95.7 ± 2.5	76.5±10.3
Titanic	79.6 ± 7.8	78.7 ± 7.8	69.3±1.2
Twonorm	97.1 ± 1.5	* 97.5 ± 1.3	51.7±3.5
Waveform	89.6 ± 3.0	* 90.0 ± 2.6	66.7±0.6

KSM (1) との比較 (テストデータ)

表 3.14 に 2 クラスベンチマークデータセットにおけるテストデータセットの平均認識率と標準偏差を示す。各データセットに対して、最も高い平均認識率を太字で示し、SSLP-SVM (1), SSLP-SVM (O,1) において KSM (1) より高い認識率の前に (*) を付ける。

表 3.14 より、SSLP-SVM (1) は KSM (1) と比べて 4 個のデータセットにおいて高い平均認識率を示している。SSLP-SVM (O,1) は Ringnorm データセットにおいてのみ KSM (1) と比べて高い平均認識率を示している。これらのように、Ringnorm データセットにおいて極めて高い汎化能力を示しているが、大半のデータセットにおいては KSM (1) と比べて低い平均認識率を示している。これより汎化能力の向上ができていないといえる。この原因として、KSM (1) によって与えられたパラメータ値が SSLP-SVM では最適なパラメータ値とは大きく違うためであると考えられる。

表 3.13, 3.14 を比較すると、大半のデータセットにおいて、バリデーションセットに

において最もよい結果を示しているときテストデータでも最もよい結果を示し、バリデーションセットにおいて悪ければ、テストデータでも悪い結果を示している。このように、SSLP-SVM においても五分割交差検定によりある程度の汎化能力の推測が行えているといえる。

表 3.14 テストデータセットを用いたときの平均認識率 (%) と標準偏差 (SSLP-SVM と KSM (1) の比較)

Data	KSM (1)	SSLP-SVM (1)	SSLP-SVM (O,1)
Banana	88.6 ± 0.6	88.6 ± 0.6	87.3±0.9
B.cancer	75.0 ± 4.2	73.3 ± 4.6	72.8±3.2
Diabetes	73.4 ± 1.7	71.3 ± 4.9	70.8±2.5
German	75.1 ± 2.2	71.2 ± 8.5	70.1±3.2
Heart	80.4 ± 3.3	* 82.9 ± 3.7	77.2±5.7
Image	96.3 ± 0.6	87.3 ± 13.5	91.5±3.9
Ringnorm	76.5 ± 11.1	* 98.2 ± 0.2	* 98.2 ±0.3
F. solar	65.1 ± 1.8	63.4 ± 4.7	53.6±9.3
Splice	87.6 ± 0.8	86.7 ± 0.8	70.2±10.8
Thyroid	95.6 ± 2.0	95.3 ± 3.8	75.4±13.2
Titanic	76.6 ± 1.2	* 77.0 ± 1.7	67.9±0.5
Twonorm	97.6 ± 0.1	97.5 ± 0.2	51.1 ±0.3
Waveform	88.5 ± 5.6	* 89.9 ± 0.7	77.0±2.4

KSM (E) との比較 (バリデーションセット)

表 3.15 に KSM (E) と SSLP-SVM と SSLP-SVM (O) を用いたときの 2 クラスベンチマークデータセットにおけるバリデーションデータセットの平均認識率 (%) と標準偏差を示す。各データセットに対して、最も高い平均認識率を太字で示す。また、それぞれ KSM (E) における累積寄与率の閾値 κ 、カーネルパラメータ γ を用いていることを示すためにそれぞれの手法を SSLP-SVM (E), SSLP-SVM (O,E) と表中に示す。各データセットに対して、最も高い平均認識率を太字で示す。SSLP-SVM (E), SSLP-SVM (O,E) において KSM (E) より高い認識率の前に (*) を付ける。

表 3.15 より、SSLP-SVM (E) は 7 個のデータセットにおいて KSM (E) と比べて高い平均認識率を示している。SSLP-SVM (O,E) は Image データセット、Ringnorm データセットのみにおいて KSM (E) と比べて高い平均認識率を示しており、大半のデータセットにおいて KSM (E) と比べて極めて低い平均認識率を示している。この原因として SSLP-SVM (O,E) は KSM (E) におけるパラメータ値では最適な解が導けないと考えられる。3.6.6 節に独自にパラメータ値を決定した計算機実験により確認する。これらより、バリデーションセットにおいて SSLP-SVM (E), SSLP-SVM (O,

E) は KSM (E) に比べて汎化能力が低くなると推測される。

表 3.15 バリデーションセットを用いたときの平均認識率 (%) と標準偏差 (SSLP-SVM と KSM (E) の比較)

Data	KSM (E)	SSLP-SVM (E)	SSLP-SVM (O,E)
Banana	88.3 ± 2.9	*89.2 ± 3.3	87.5 ± 3.4
B. cancer	75.7 ± 3.5	68.1 ± 6.4	70.9 ± 2.2
Diabetes	73.8 ± 3.0	70.9 ± 3.0	59.4 ± 8.6
German	71.6 ± 2.8	71.5 ± 2.5	71.1 ± 6.1
Heart	82.9 ± 3.9	* 83.5 ± 5.4	74.4 ± 6.8
Image	88.1 ± 1.4	* 95.2 ± 1.0	*93.8 ± 2.6
Ringnorm	62.5 ± 4.0	*63.9 ± 5.7	* 76.8 ± 20.8
F. solar	64.4 ± 3.7	* 65.1 ± 3.2	61.2 ± 4.1
Splice	72.0 ± 3.0	62.4 ± 9.2	60.3 ± 10.9
Thyroid	96.0 ± 2.7	* 96.1 ± 2.4	76.0 ± 11.6
Titanic	79.4 ± 7.0	73.5 ± 9.0	79.1 ± 6.1
Twonorm	97.2 ± 1.4	97.1 ± 1.4	92.8 ± 8.0
Waveform	89.0 ± 2.8	* 89.1 ± 2.5	74.0 ± 5.9

KSM (E) との比較 (テストデータ)

表 3.16 に 2 クラスベンチマークデータセットにおけるテストデータセットの平均認識率と標準偏差を示す。各データセットに対して、KSM (E), SSLP-SVM (E), SSLP-SVM (O,E) の結果の最も高い平均認識率を太字で示す。SSLP-SVM (E), SSLP-SVM (O,E) において KSM (E) より高い認識率の前に (*) を付ける。

表 3.16 より、SSLP-SVM (E) は 7 個のデータセットにおいて KSM (E) と比べて高い平均認識率を示している。SSLP-SVM (O,E) は Ringnorm データセット、Image データセットのみ KSM (E) と比べ高い平均認識率を示している。これより同時定式化手法を用いれば汎化能力がある程度向上するが、一対他方式では逆に汎化能力が非常に劣化するといえる。前述にあるように、この原因としてパラメータ値の設定方法に問題があるといえる。これらを確認するため 3.6.6 節にてパラメータ設定方法を変更した計算機実験により再び評価を行う。

また、バリデーションセットを用いたときの平均認識率とほぼ近い精度が得られたことから五分割交差検定法によるモデル選択により汎化能力の推測ができていくといえる。これらにより五分割交差検定法はモデル選択に有効であると検証できたため、次節以降バリデーションセットによる評価実験を省略する。

表 3.16 テストデータセットを用いたときの平均認識率 (%) と標準偏差 (SSLS-SVM と KSM (E) の比較)

Data	KSM (E)	SSLP-SVM (E)	SSLP-SVM (O, E)
Banana	87.8±0.6	*88.6±0.6	87.3±0.9
B. cancer	75.1±4.4	67.4±5.0	70.8±4.2
Diabetes	71.7±2.3	70.1±2.8	70.0±10.2
German	73.6±2.1	71.1±2.5	70.2±6.3
Heart	82.4±3.6	*83.1±3.8	78.2±4.4
Image	88.1±1.0	*95.8±0.8	*93.1±5.1
Ringnorm	64.1±2.4	64.1±2.4	78.9 ±20.2
F. solar	63.5±4.0	*65.0±1.7	59.1±1.0
Splice	71.8±1.5	51.4±4.5	61.3±10.2
Thyroid	95.1±2.4	*96.1±2.1	75.4±13.2
Titanic	77.3±0.3	76.1±8.9	77.2±1.2
Twonorm	97.0±0.5	*97.3±0.3	92.2±13.3
Waveform	88.0±1.1	*88.2±1.2	71.8±4.6

3.6.6 独自でモデル選択を行った SSLP-SVM の評価

前述のように, SSLP-SVM, SSLP-SVM (O) における累積寄与率の閾値 κ , カーネルパラメータ γ は KSM (1), KSM (E) のそれぞれで決定された値を選択していたが, SSLP-SVM, SSLP-SVM (O) において最適な選択とはなっていない. そこで, すべてのパラメータ値を独自に五分割交差検定法により決定した際, それぞれの KSM と比べて汎化能力が向上しているかどうかを評価する. ただし, SSLP-SVM では最適化の際, 同時に特徴選択することができる. そこで, どれほど削減でき, 必要な軸を削減していないか確認するために累積寄与率は $\kappa = 99.9(\%)$ とおく. また本実験では多クラス問題についても評価するが, 計算機のパフォーマンス上の問題で大半の多クラスデータセットにおける計算機実験は不可能だったため, Iris データセットと Numeral データセットのみを用いる.

パラメータ値の選択候補は前節までの候補と同じにしており, 表 3.17 に選択された SSLS-SVM, SSLS-SVM (O) のパラメータ値を示す.

表 3.18 にテストデータセットでの平均認識率と標準偏差を示す. 各データセットに対して, 最大となる平均認識率を太字で示す. また, 各データセットに対して, 最も高い平均認識率を太字で示し, SSLP-SVM, SSLP-SVM (O) において KSM (1) かつ KSM (E) より高い認識率の前に (*) を付ける.

表 3.17 パラメータ値 (独自でモデル選択を行った SSLP-SVM)

Data	SSLP-SVM		SSLP-SVM (O)	
	γ	C	γ	C
Banana	1	10	0.5	100
B. cancer	0.05	50	0.1	0.1
Diabetes	0.1	100	0.1	100
German	0.1	100	1	500
Heart	0.1	10	0.1	1
Image	0.05	10^3	0.5	10^3
Ringnorm	0.1	10^3	0.01	500
F. solar	0.5	500	0.1	10^3
Splice	0.05	100	0.05	100
Thyroid	1	1	1	1
Titanic	1.5	5	0.5	0.1
Twonorm	0.05	10	0.5	1
Waveform	1	10	1	1
Iris	1.5	100	0.01	10^4
Numeral	1	100	1	10^3

2 クラス問題

2 クラス問題では, SSLP-SVM は 2 つの KSM と比較して, 5 個のデータセットにおいて高い平均認識率を示している. ただし, KSM (1) のみと比較すると 7 個, KSM (E) のみと比較すると 10 個のデータセットにおいて高い平均認識率を示している. 特に Diabetes データセット, Ringnorm データセットにおいては汎化能力が極めて向上している. 平均認識率が劣るデータセットに対してもほぼ同程度といえる. SSLP-SVM (O) は 2 つの KSM と比較して, 4 個のデータセットにおいて高い平均認識率を示している. ただし, KSM (1) のみと比較すると 6 個, KSM (E) のみと比較すると 10 個のデータセットにおいて高い平均認識率を示している. 特に Ringnorm データセットにおいては汎化能力が極めて向上している. また平均認識率が劣るデータセットにおいても German データセットを除いてはほぼ同程度といえる.

また, 表 3.14, 3.16 の結果と比べると大半のデータセットにおいて SSLP-SVM, SSLP-SVM (O) とともに平均認識率が極めて向上している. これらの結果より, 前節までの SSLP-SVM, SSLP-SVM (O) の平均認識率が低い原因は最適なパラメータ値を選択できていないためであるといえる. また, 特徴選択を累積寄与率を用いずに行った結果がよいことから最適化と同時に行う特徴選択は有効である. SSLP-SVM と SSLP-SVM (O) はほぼ同程度の性能を示しており, 若干同時定式化方式を用いた場合の方が一対他方式を適用した場合より汎化能力が高い.

多クラス問題

多クラス問題では, SSLP-SVM は各 KSM と比べてほぼ同等の認識率を示している. SSLP-SVM (O) においては各 KSM と比べてすべてのデータセットにおいて認識率が向上している. これらより, 多クラス問題においては一対他方式が同時定式化方式に比べて優れていると推測できる.

表 3.18 テストデータに対する平均認識率 (%) 及び標準偏差

Data	KSM (1)	KSM (E)	SSLP-SVM	SSLP-SVM (O)
Banana	88.6 ± 0.6	87.8 ± 0.7	*88.7 ± 0.5	*89.0 ± 0.5
B. cancer	75.0 ± 4.2	75.1 ± 4.3	73.6 ± 4.3	70.5 ± 5.1
Diabetes	73.4 ± 1.7	71.7 ± 2.2	* 75.6 ± 1.6	73.3 ± 1.9
German	75.1 ± 2.2	73.7 ± 2.1	* 75.6 ± 2.4	71.3 ± 2.6
Heart	80.4 ± 3.3	82.4 ± 3.6	81.8 ± 3.3	*83.3 ± 3.4
Image	96.3 ± 0.6	88.0 ± 0.9	89.0 ± 15.6	92.9 ± 7.5
Ringnorm	76.6 ± 11.1	64.1 ± 2.4	*98.1 ± 0.1	*98.4 ± 0.1
F. solar	65.1 ± 1.8	63.5 ± 3.9	63.7 ± 1.9	64.1 ± 2.0
Splice	87.6 ± 0.8	71.9 ± 1.5	87.0 ± 0.9	85.0 ± 1.2
Thyroid	95.6 ± 2.0	95.1 ± 2.4	* 96.4 ± 2.0	*95.9 ± 2.0
Titanic	76.6 ± 1.2	77.3 ± 0.6	76.8 ± 1.1	76.7 ± 1.0
Twonorm	97.6 ± 0.1	97.0 ± 0.5	97.5 ± 0.1	97.4 ± 0.1
Waveform	88.5 ± 5.6	88.0 ± 1.1	88.2 ± 3.0	*89.3 ± 0.7
Iris	93.3	92.0	90.7	93.3
Numeral	99.0	99.0	98.9	99.3

SSLP-SVM による軸の削減割合

本節では SSLP-SVM でどの程度, 特徴を削減することができるかを 2 クラスベンチマークデータセットを用いて確認する. ここで用いる SSLP-SVM は同時定式化方式を用いた手法のみとする. 表 3.19 に SSLP-SVM, SSLP-SVM (1), SSLP-SVM (E) によって減少した固有ベクトルの数を示している. 各 SSLP-SVM の “Class 1”, “Class 2” は, KPCA により導かれた各クラスの部分空間の次元数を表し, “Rem 1”, “Rem 2” は各 SSLP-SVM の学習によって特徴選択された後の各クラス部分空間の次元数を表す. Heart と Twonorm に関しては, 部分空間の次元数が 1 となっており, SSLP-SVM による特徴選択が上手く行われているということがわかる.

表 3.19 特徴選択による特徴量の削減数 (SSLP-SVM)

Data	SSLP-SVM				SSLP-SVM (1)				SSLP-SVM (E)			
	Class1	Rem	Class2	Rem	Class1	Rem	Class2	Rem	Class1	Rem	Class2	Rem
Banana	67.7	11.5	73.2	10.7	123.1	18.2	128.8	20.5	123.1	18.2	128.8	20.5
B. cancer	75.6	0.8	53.5	0.9	75.6	0.8	53.5	0.9	20.2	5.5	22.7	9.4
Diabetes	99.7	11.0	103.3	11.2	229.4	26.9	159.0	42.3	19.0	15.1	54.7	36.4
German	494.6	33.2	212.8	129.1	113.0	19.8	83.0	54.0	398.8	260.0	192.3	100.6
Heart	94.4	1.7	75.6	1.4	21.2	6.9	28.1	8.7	37.3	1.2	42.0	1.3
Image	65.7	23.6	70.7	20.2	201.4	43.0	135.6	26.5	485.9	67.9	572.9	81.9
Ringnorm	21.0	2.2	85.6	0.7	21.0	2.2	85.6	0.7	225.4	39.3	199.0	84.8
F. solar	17.2	8.4	26.6	6.6	2.0	1.8	3.0	2.6	13.4	8.7	27.8	15.2
Splice	108.3	7.5	63.5	12.4	314.7	10.0	188.8	15.0	517.0	26.2	459.9	142.7
Thyroid	15.3	2.0	33.5	1.9	11.8	5.8	29.7	27.1	8.8	6.0	27.1	26.1
Titanic	7.6	1.4	9.5	0.6	1.0	1.0	1.0	0.9	5.3	1.1	7.7	1.0
Twonorm	187.2	1.0	190.9	1.0	1.0	1.0	1.0	1.0	120.0	31.1	121.8	116.2
Waveform	265.3	10.3	132.3	11.2	3.0	3.0	2.0	2.0	187.1	3.0	92.0	37.2

学習時間の比較

表 3.20 に各 KSM, 各 SSLP-SVM における部分空間生成と重みの最適化 (各 SSLP-SVM のみ) に要するそれぞれの学習時間 (s) を示す。

2 クラス問題では, 全データセットにおいて SSLP-SVM (O) が SSLP-SVM と比べて学習が高速化されている。これは解くべき変数が削減されているためである。ただし, 各 KSM と比べて Image データセットや Splice データセットなどの大規模データセットにおいて学習時間が極めて長くなっている。これは LP-SVM の解くべき変数がデータ数が膨大になると同じく膨大となってしまうためである。これにより大規模データセットにおいて SSLP-SVM は学習が困難であるといえる。

多クラス問題においても, SSLP-SVM (O) が SSLP-SVM と比べて高速な学習が行えている。Numeral データセットのように多クラス問題においてそれほどデータ量の多くないデータセットに対しても各 KSM と比べて学習時間が非常に長くなっている。これより SSLP-SVM は多クラス問題に対しても適用することが困難である。特に同時定式化方式を用いると極端に計算量が増加している。この原因として解くべき変数がクラス数が膨大になると同じく膨大となってしまうことが考えられる。

表 3.20 SSLP-SVM における学習時間 (s)

Data	KSM (1)	KSM (E)	SSLP-SVM	SSLP-SVM (O)
Banana	0.62	0.59	3.41	2.15
B.cancer	0.10	0.09	0.10	0.11
Diabetes	1.12	1.10	9.64	7.98
German	3.13	3.61	44.1	35.9
Heart	0.07	0.06	0.07	0.09
Image	21.9	22.0	80.8	67.5
Ringnorm	0.72	0.62	4.95	3.65
F.solar	1.31	1.33	18.9	7.95
Splice	10.3	10.9	69.5	58.1
Thyroid	0.03	0.03	0.04	0.03
Titanic	0.01	0.02	0.02	0.02
Twonorm	0.71	0.72	2.32	2.01
Waveform	0.74	0.74	5.21	3.56
Iris	0.01	0.01	0.02	0.01
Numeral	43.8	42.9	251	118

3.6.7 SSL2-SVM の評価 (KSM との比較)

前節までの計算機実験により KSM によって決定されたカーネルパラメータ γ と累積寄与率の閾値 κ は SS-SVM において最適な値ではない可能性が非常に高いことが確認できたため、本節では SSL2-SVM により独立してすべてのパラメータ値を五分割交差検定法により決定する。パラメータ値の選択候補は前節までと同じにする。また、アクティブ集合を用いるため初期のワーキングセットのサイズを $p = 50$ とする。表 3.21 に SSL2-SVM, SSL2-SVM (O) における選択されたパラメータ値を示す。ただし、計算機のパフォーマンス上、多クラス問題において 3 個のデータセットのみの結果しか取れなかった。

表 3.22 にテストデータセットでの平均認識率と標準偏差を示す。各データセットに対して、最大となる平均認識率を太字で示す。各データセットに対して、最も高い平均認識率を太字で示す。また、SSL2-SVM, SSL2-SVM (O) において KSM (1) かつ KSM (E) より高い認識率の前に (*) を付ける。

表 3.21 パラメータ値 (SSL2-SVM)

Data	SSL2-SVM			SSL2-SVM (O)		
	κ	γ	C	κ	γ	C
Banana	99.9	10	100	99	30	10
B. cancer	99	20	0.1	99.9	1.5	100
Diabetes	99	30	0.1	99.9	3	10^3
German	99.9	0.5	50	99.9	3	100
Heart	99.9	0.5	1	80	20	0.1
Image	99.9	200	500	99.9	20	10^3
Ringnorm	99.9	15	10	99.9	1	1
F. solar	99.9	5	50	90	3	10
Splice	99.9	10	500	99.9	1	100
Thyroid	99.9	15	500	99.9	5	10^3
Titanic	90	10	0.1	99.9	10	10^3
Twonorm	99.9	1.5	0.01	80	0.5	0.1
Waveform	85	50	1	80	3	100
Iris	80	50	10	80	30	10
Numeral	99.9	15	500	95	100	5
Blood cell	99.9	30	10^3	99	200	10

2 クラス問題

2 クラス問題では, SSL2-SVM は 2 つの KSM と比べ 9 個のデータセットにおいて高い平均認識率を示している. ただし, KSM (1) のみと比較すると 10 個, KSM (E) のみと比較すると 11 個のデータセットにおいて高い平均認識率を示している. 特に Ringnorm データセット, Diabetes データセット, F. solar データセット, Waveform データセットにおいては非常に高い平均認識率をしめしている. SSL2-SVM (O) は 2 つの KSM と比較して 5 個のデータセットにおいて高い平均認識率を示している. ただし, KSM (1) のみと比較すると 6 個, KSM (E) のみと比較すると 8 個のデータセットにおいて高い平均認識率を示している. 特に Ringnorm データセットに対して非常に高い平均認識率を示している. これらより, SSL2-SVM は通常の KSM と比べて汎化能力が非常に向上しているといえる. ただし SSL2-SVM (O) は Splice データセットのように非常に低い平均認識率を示していることから安定して汎化能力の向上ができていないといえる. これは重み付けの際に他のクラスの重みを考慮していないためである. すなわち, SSL2-SVM においても同時定式化方式が一対他方式よりも有効であるといえる.

多クラス問題

多クラス問題では, SSL2-SVM は全データセットにおいてすべての手法と比べて高い認識率を示している. SSL2-SVM (O) は各 KSM と比べて2個のデータセットにおいて高い認識率を示している. 他のデータセットにおいても同じ認識率を示している. これにより SSL2-SVM は多クラス問題においても各 KSM と比べて汎化能力の面で優れているといえる.

また, SSL2-SVM と SSL2-SVM (O) の学習コストはほぼ変わらなかったため, これらの学習コストの比較は省略する.

表 3.22 テストデータセットにおける平均認識率 (%) と標準偏差 (SSL2-SVM と KSM の比較)

Data	KSM (1)	KSM (E)	SSL2-SVM	SSL2-SVM (O)
Banana	88.6±0.6	87.8±0.6	* 89.4 ±0.4	*89.0±0.6
B. cancer	75.0±4.3	75.1 ±4.4	73.4±4.3	73.6±4.8
Diabetes	73.5±1.8	71.7±2.3	* 75.9 ±1.9	71.1±2.4
German	75.1±2.2	73.6±2.1	* 76.0 ±2.3	73.4±2.6
Heart	80.5±3.3	82.4±3.6	* 84.3 ±3.3	*83.1±3.9
Image	96.3 ±0.6	88.1±1.0	96.1±0.9	94.9±0.6
Ringnorm	76.6±0.6	64.1±2.4	*95.8±0.7	* 98.1 ±0.1
F. solar	65.1±1.8	63.5±4.0	* 66.5 ±1.6	*66.2±1.6
Splice	87.6 ±0.8	71.8±1.5	*89.0±0.6	77.2±2.3
Thyroid	95.6±2.1	95.1±2.4	95.4±2.5	* 95.9 ±2.1
Titanic	76.6±1.2	77.3 ±0.3	77.1±1.1	77.1±0.9
Twonorm	97.6±0.1	97.0±0.5	* 97.7 ±0.1	97.6±0.2
Waveform	88.5±0.6	88.0±1.1	* 90.2 ±1.2	86.3±1.6
Iris	93.3	92.0	* 97.3	* 97.3
Numeral	99.0	99.0	* 99.3	99.0
Blood cell	91.3	90.1	* 92.6	*91.5

SSL2-SVM と SSLS-SVM との学習時間の比較

SSL2-SVM ではアクティブ集合による学習法を用いている. この学習法によって, SSLS-SVM と比べてどれほどの学習コストの削減をされているかを確認する. 表 3.23 に部分空間生成部と重みの最適化に要する SSLS-SVM と SSL2-SVM の学習時間を示す.

2 クラス問題では大半のデータセットにおいて, SSLS-SVM に比べ SSL2-SVM の学習時間が短くなっていることがわかる. しかしながら, 小規模なデータにおいては, それほど学習時間に差がなく, 有効性を示しているとはいえないが, Image データセット

と Splice データセットの大規模データセットにおいて特に学習時間の短縮が行えている。これは全データ数に比べ、SV 数が非常に少ないためであるといえる。これより、2 クラス大規模データセットにおいて、SSL2-SVM は SSLS-SVM に比べて学習コストが非常に小さくなるといえる。

多クラス問題では、全データセットにおいて、SSL2-SVM に比べ SSLS-SVM の学習時間が短くなっていることがわかる。これは、SSLS-SVM の計算コストは最大オーダー (M^3) となるが、SSL2-SVM は最大オーダー ($(nM)^3$) となり、クラス数が多いと学習コストも膨大となる恐れがあるためである。よって SSL2-SVM は多クラス問題においては不向きである。

表 3.23 SSL2-SVM における学習時間 (s)

Data	SSLS-SVM	SSL2-SVM
Banana	0.91	0.13
B. cancer	0.10	0.02
Diabetes	1.12	0.69
German	4.11	3.72
Heart	0.07	0.08
Image	22.5	9.8
Ringnorm	1.06	2.11
F. solar	1.35	3.56
Splice	13.1	6.51
Thyroid	0.04	0.05
Titanic	0.02	0.01
Twonorm	0.70	0.71
Waveform	0.75	1.11
Iris	0.02	0.01
Numeral	44.7	101
Blood cell	319	372

3.6.8 SS-SVM における各モデルの性能比較

SS-SVM におけるすべてのモデルに対しての汎化能力の比較を行う。

表 3.24 に SVM, KSM, SSLS-SVM, SSLP-SVM, SSL2-SVM のテストデータにおける平均認識率と標準偏差を示す。ここで、表中の“KSM”は KSM (1), KSM (E) におけるバリデーションセットの結果のよいモデルを示す。また、表中の“SSLS-SVM”, “SSLP-SVM”, “SSL2-SVM”は同時定式化方式化、または一対他方式を適用した各モデルの中でバリデーションセットの結果のよいモデルを示す。汎化能力の高さを確認するために通常の SVM と平均認識率と標準偏差も記す。KSM と SS-SVM の各モデルに

おける平均認識率のうち、各データセットにおいて最大値を太字で示す。また SS-SVM の各モデルにおいて KSM と比較して高い平均認識率に対して (*) を付ける。

2 クラス問題

2 クラス問題では、SSLS-SVM は 3 個のデータセットにおいて最も高い平均認識率を示し、KSM と比べて 9 個のデータセットにおいて高い平均認識率を示している。特に、他の SS-SVM とは異なり、B. cancer データセットにおいて KSM と同じ平均認識率を示している。これらの結果より全モデルの中で KSM と比べて安定して高い汎化能力を示すモデルであるといえる。SSLP-SVM は 3 個のデータセットにおいて最も高い平均認識率を示し、KSM と比べて 8 個のデータセットにおいて高い平均認識率を示している。しかしながら、B. cancer データセットや German データセットや F. solar データセットにおいて非常に低い平均認識率を示している。これらの結果より SSLP-SVM はあまり安定した汎化能力を出すことのできないモデルであるといえる。SSL2-SVM は 8 個のデータセットにおいて最も高い平均認識率を示し、KSM と比べて 11 個のデータセットにおいて高い平均認識率を示している。また、B. cancer データセットを除いてはすべてにおいて安定して高い汎化能力を維持しているといえる。これらより、SS-SVM の全モデルは KSM と比べて汎化能力が高いことが確認できる。また、SSL2-SVM が全モデルの中で 2 クラス問題において最も汎化能力の高い識別手法であるといえる。

これらの手法を非常に高い汎化能力をもつ識別器のひとつである SVM と比べると SSLS-SVM では 4 個、SSLP-SVM では 2 個、SSL2-SVM では 6 個のデータセットにおいて高い平均認識率を示している。これにより SSLS-SVM と SSLP-SVM は SVM と比べると汎化能力は低く、SSL2-SVM においても若干劣る性能を示している。

多クラス問題

多クラス問題では、SSLS-SVM は 5 個のデータセットにおいて最も高い認識率を示し、すべて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。SSLP-SVM は多クラス問題において計算機のパフォーマンス上 2 個のデータセットのみしか結果を取れなく、いずれのデータセットにおいても最も高い認識率を示したものはない。しかしながら、KSM と比べると Numeral データセットにおいて高い認識率を示し、Iris データセットにおいても同じ認識率を示していることから汎化能力が向上しているといえる。SSL2-SVM は多クラス問題において計算機のパフォーマンス上、3 個のデータセットのみしか結果を取れていないが、Iris データセットにおいて最も高い認識率を示している。また KSM と比べて全データセットにおいて高い認識率を示している。

これらの結果より, SSSL-SVM が多クラス問題において全モデルの中で最も汎化能力が優れており, かつ大規模データセットに対しても結果が取れていることから実用的なモデルであるといえる.

これらの手法を SVM と比べると, SS-SVM の全モデルは全データセットにおいて劣っているか, または同じ認識率を示している. 最も汎化能力の高いモデルである SSSL-SVM においても, 6 個中 2 個のデータセットにおいて同じ認識率を示しており, その他のデータセットにおいては低い認識率を示している. Iris データセットにおいては 1.3% の劣化が見られるがテストデータ数が 75 個しかないため, SVM と比べて誤認識したデータが 1 個多くなったのみである. そのため,それほど問題とはならないが, Hiragana-50 データセットにおいて 3.4% の劣化が見られた. このデータセットでは KSM においても 3.6% の劣化が見られることから, 各クラス部分空間を構成する軸の選択が識別の観点から最適ではないことが原因として考えられる. これらの結果より, SS-SVM は多クラス問題においても KSM と比べ高い汎化能力を示すが, SVM と比べてはまだ汎化能力が低いといえる.

表 3.24 テストデータセットを用いたときの平均認識率 (%) と標準偏差 (各 SS-SVM の比較)

Data	SVM	KSM	SSLS-SVM	SSLP-SVM	SSL2-SVM
Banana	89.3 ± 0.5	88.6 ± 0.6	*88.9 ± 0.6	*89.0 ± 0.6	* 89.4 ± 0.4
B. cancer	72.4 ± 4.6	75.1 ± 4.4	75.1 ± 4.4	73.3 ± 4.6	73.6 ± 4.8
Diabetes	76.3 ± 1.8	71.7 ± 2.3	*72.0 ± 2.3	* 73.5 ± 2.0	* 75.9 ± 1.9
German	76.2 ± 2.2	75.1 ± 2.2	73.8 ± 2.2	70.9 ± 6.6	* 76.0 ± 2.3
Heart	83.7 ± 3.4	82.4 ± 3.3	*82.6 ± 3.7	*83.1 ± 3.8	* 84.3 ± 3.3
Image	97.3 ± 0.4	96.3 ± 0.6	95.4 ± 0.6	95.1 ± 1.0	96.1 ± 0.9
Ringnorm	97.8 ± 0.3	64.1 ± 2.4	*97.6 ± 0.3	* 98.2 ± 0.2	*98.1 ± 0.1
F. solar	67.6 ± 1.7	65.1 ± 1.8	* 66.9 ± 1.6	64.7 ± 1.9	*66.5 ± 1.6
Splice	89.2 ± 0.7	87.6 ± 0.8	86.2 ± 1.0	88.2 ± 0.6	* 89.0 ± 0.6
Thyroid	96.1 ± 2.0	95.1 ± 2.4	* 95.4 ± 2.1	* 96.3 ± 2.2	*95.9 ± 2.1
Titanic	77.2 ± 1.1	76.6 ± 1.2	* 76.9 ± 1.0	*76.8 ± 1.1	* 77.1 ± 1.1
Twonorm	97.6 ± 0.1	97.0 ± 0.5	* 97.7 ± 0.1	* 97.6 ± 0.2	* 97.7 ± 0.1
Waveform	90.0 ± 0.4	88.5 ± 0.6	* 89.1 ± 0.8	* 89.3 ± 0.7	* 90.2 ± 1.2
Iris	97.3	93.3	*96.0	93.3	* 97.3
Numeral	99.6	99.0	* 99.4	*99.3	*99.3
Blood cell	92.9	91.3	* 92.9	—	*92.6
Hiragana-50	98.9	95.3	* 95.5	—	—
Hiragana-105	99.9	99.7	* 99.9	—	—
Hiragana-13	99.1	98.4	* 98.9	—	—

3.6.9 識別部の解析

部分空間法においての大きな利点のひとつである識別部の解析を行う。図 3.3 に Thyroid データセットにおける各軸に対する類似度を示す。ここで、 \circ はクラス 1 のデータを示しており、青色の \times はクラス 2 のデータを示している。また、図示しやすくするため、識別に影響を与えやすい軸における解析、すなわち軸における重みの値が特に大きいものを 2 本選び、2 次元座標上にプロットした。横軸は重みが最大となる軸における類似度、縦軸は 2 番目に重みが大きくなるような軸を示している。また本実験で用いた SS-SVM のモデルは SSSL-SVM である。

ここで、各軸に対する重みは横軸が約 4.2 であり、縦軸は約 1.1 である。図 3.3 を見てわかるように、重みが最大となる軸においてクラス 1 のデータセットの類似度がクラス 2 の類似度より大きくなりやすい傾向にある。そのような軸においての重みは大きくなりやすくなっている。一部のデータはこの軸においての類似度が高くなっており、このようなデータが誤認識される。このように、重みが最大となる軸が識別に大きく影響を受けていることが確認できる。

このように、SS-SVM では各軸に関する類似度の解析を行うことにより、どのような軸において識別を行えやすくなっているかどうかを確認できる。これは識別部が解析可能であることを意味しており、SVM より非常に有効な点である。

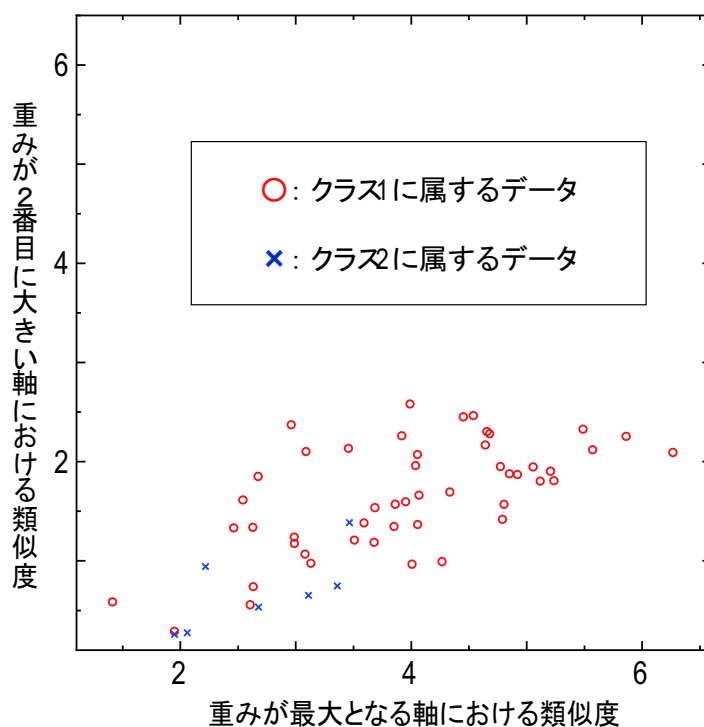


図 3.3 各軸に対する類似度

3.7 結言

本章では、2章の結言で述べた KSM の汎化能力が低いという問題点を解消するために SVM のマージン最大化の概念を KSM に取り入れた SS-SVM について述べた。SS-SVM では、クラスごとに生成された部分空間を構成する軸の重みに対してクラス間のマージンを最大にするように重みの最適化処理を行う。ここで SVM のモデルとして 2章に述べた LS-SVM, LP-SVM, L2-SVM をそれぞれ用いた手法における最適化処理のアルゴリズムを説明した。これらの手法により、部分空間を構成する軸に対して識別の観点から最適な重み付けをできる。また、それぞれのモデルに対して同時定式化方式を基本的に使用するが計算コストを削減するために一対他方式をそれぞれ適用する SS-SVM (O) についても述べた。

計算機実験では SS-SVM の各モデルに対して KSM と同じパラメータ値を与え、汎化能力が向上するか確認した。大半の 2 クラスベンチマークデータセットにおいて向上できていないことが確認できた。そこで、各モデルにおいて独立して五分割交差検定法によりモデル選択を行ったところ、KSM と同じパラメータ値を与えるより汎化能力が非常に向上した。さらに大半のデータセットにおいて各モデルとも KSM より高い平均認識率を示した。また、同時定式化方式と一対他方式の学習コストの比較を行うためにそれぞれにおける学習時間の比較を行った。SSLS-SVM では、2 クラス問題において同時定式化方式を適用した場合の方が一対他方式よりも高速である。しかしながら、多クラス問題においては一対他方式が同時定式化方式よりも高速となっていることから多クラス問題においては学習コストの面で一対他方式が有効であるといえる。また SSLP-SVM では一対他方式を適用した場合の方が大規模データセットに対して非常に高速になることが確認できた。しかしながら、通常の KSM と比べると学習時間が非常に長くなり、多クラス大規模データセットにおいては学習が困難となることからそれほど実用的ではないといえる。また SSL2-SVM は 2 クラス問題において SSLS-SVM より高速であり、さらに汎化能力が高いことから 2 クラス問題においては非常に有効な手法であるといえる。ただし、多クラス問題においては計算コストが非常に膨大となり、学習が困難となることから多クラス問題においては適用しがたいといえる。SS-SVM の全モデルの汎化能力を SVM と比較すると、すべてのモデルにおいて劣っていたことから、さらなる汎化能力の向上が必要となる。そのための手法を 4 章に説明する。また、SS-SVM の学習時間の比較により、SSLP-SVM と多クラス問題における SSL2-SVM を除いて学習コストの大半は重みの最適化問題ではなく、部分空間生成におけるものではないかと考えることができ、部分空間生成における学習コストの削減が必要となる。これを解決する手法を 5 章で述べる。

第4章

重みに基づく特徴選択法

4.1 緒言

通常のカネル部分空間法 (KSM: Kernel Subspace Method)^{13, 14)}, 前章で述べた部分空間法に基づくサポートベクトルマシン (SS-SVM: Subspace-Based Support Vector Machine)^{23, 24)} において, 部分空間を構成する軸の選択には部分空間法に基づく線形計画サポートベクトルマシン (SSLP-SVM: Subspace-Based Linear Programming Support Vector Machine) を除いては, 共分散行列の固有値分解によって, 導かれた固有値の大きいものから順に選択する手法を用いる. しかしながら, この手法では, 他のクラスにおける部分空間を考慮していないため, 識別に必要な固有ベクトルが部分空間を構成する軸として選ばれていない可能性がある. また, 識別において冗長な軸が選択されうることからサポートベクトルマシン (SVM: Support Vector Machine)^{5, 6, 7)} と比べまだ汎化能力が劣っているのではないかと考えられる.

そこで本章では, SS-SVM による部分空間を構成する軸への最適な重み付けを利用し, それらの重みを用いて, 特徴選択を行う手法^{25, 26)} を提案する. まず, あらかじめ累積寄与率の閾値 κ を非常に高い値 (99.9%) に設定し, カネル主成分分析 (KPCA: Kernel Principal Component Analysis)^{10, 11)} による特徴抽出を行う. そして, すべての固有ベクトルを用いて最適化を行い, SS-SVM により決定された重みを識別における軸の重要度とする. その重みの中である条件を満たす軸のみを用いて, 再び重みの最適化を行い, 選択された軸の重みが条件を満たすまで繰り返す. このとき, 2つの条件を設ける. 1つ目は, 非負条件である. これは, 通常のカネル部分空間法に基づく最小自乗サポートベクトルマシン (SSLS-SVM: Subspace-Based Least Squares Support Vector Machine) と部分空間法に基づく L2 サポートベクトルマシン (SSL2-SVM: Subspace-Based L2 Support Vector Machine) では負の重みが存在しうるためである. しかしながら, SSL2-SVM では重みの学習コストが次元数にほとんど影響を受けないため, サポートベクトル数が減らない限りは繰り返しごとの学習コストが同じである. このとき, 繰り返し回数が増加すれば重みの最適化における学習コストが膨大となる. このため SSLS-SVM にのみ適用する. 2つ目の条件は, 重みの絶対値があらかじめ設定した閾値よりも大きくなることである. 重みの小さいものは識別において影響をあまり与えないことから, 識別において冗長な可能性が高い. こ

のため、それらの軸を削除することにより、識別の観点から有用な軸のみを選択でき、汎化能力の向上が期待できる。また、これらの手法においても SSLS-SVM を適用し、SSL2-SVM への導入を行わないのは前述にある学習コストの増加につながるためである。さらに、この2つの条件を設定したそれぞれの SSLS-SVM を比較することにより、負の重みを持つ軸の必要性を確認することができる。

計算機実験により、SSLS-SVM と SSLP-SVM を用いて負の重みをもつ軸の識別における必要性を確認する。さらに、SSLS-SVM における重みの絶対値に基づく特徴選択手法を通常の KSM, SS-SVM における寄与率による特徴選択手法と比較することにより、有効性を示す。また、汎化能力の非常に高い識別器である SVM と比較することにより、重みに基づく特徴選択手法を用いた SS-SVM の汎化能力の高さを示す。

4.2 重みに基づいた軸の選択

本節では、クラス部分空間を構成する軸から重みが正となる軸のみを選択し、または絶対値の大きい重みに対応する軸のみを選択することでクラス部分空間を形成する軸の最適な選択処理を行う SSLS-SVM について述べる。これらのように前章までの固有値を基準とした手法とは異なり、識別の観点から決定された重みを基に特徴選択することができる。

4.2.1 非負制約を設けた SSLS-SVM

KSM では各クラス部分空間の各軸の重みがすべて正の値をとるが、SSLS-SVM では負の値をとりうる。図 4.1 のような、2 クラス問題において考える。まずクラス 1 において 2 次元の部分空間が生成されており、クラス 2 においては 1 次元の部分空間が生成されたとする。このとき φ_{ik} はクラス i 部分空間における k 番目の軸を意味する。 φ_{11} と φ_{21} はそれぞれの対応するクラスのデータを最も表現できるような軸になっている。しかしながら、 φ_{12} は属するクラス 1 のデータよりもクラス 2 のデータをよく表現できるような軸となっている。このようなとき、軸の重みは負となる。

しかしながら、重みが負となる軸は他のクラスに属するデータを近似してしまう軸であるといえるが、負の重みによって識別に対してよい影響を及ぼす可能性がある。そのため、識別の観点から不必要な軸とは必ずしもいえない。SSLS-SVM に非負制約を設け、負の重みをもつ軸が識別に対してどのような影響を与えるかを 4.3 節の計算機実験で確認する。ただし、SSLP-SVM とは異なり、最適化問題に直接的に非負制約を組み込むことが困難である。そこで、全データセットにおいて寄与率をすべて 99.9 (%) と設定し、SSLS-SVM により導かれた重みが負となる軸を削除し、重みが正となる軸のみを用いて再学習し、全重みが正となるまでこの処理を繰り返す。これによりデー

タセットごとに寄与率を設定することなく、軸の選択処理を行うことができる。

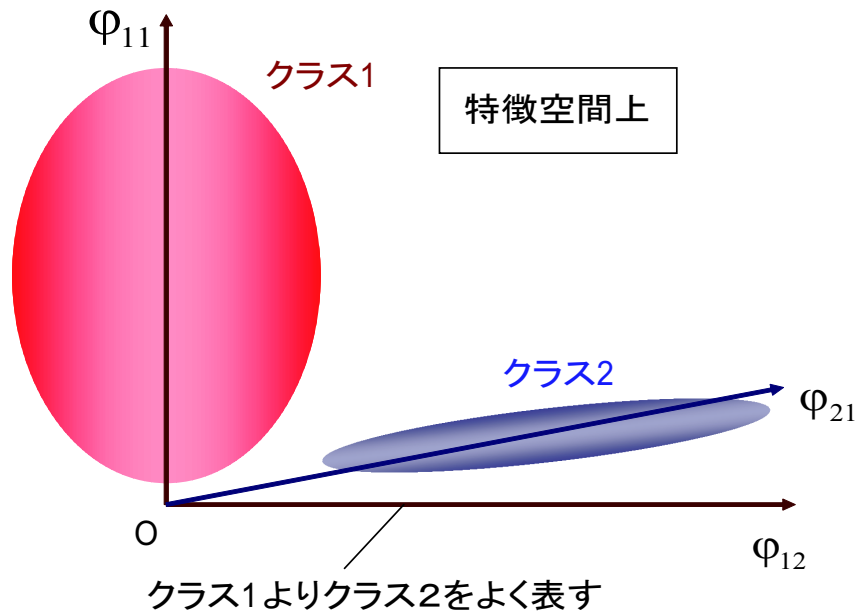


図 4.1 負の重みと軸の関係

以下にこの選択処理を用いた SLS-SVM のアルゴリズムを示す。

アルゴリズム 8

- Step 1** 五分割交差検定を用いて、カーネルパラメータ γ とマージンパラメータ C の値をそれぞれ決定する。寄与率 $\kappa = 99.9\%$ とおく。
- Step 2** Step 1 で決定したパラメータの値を用いてコレスキー分解によって教師データから一次独立なデータを選択する。
- Step 3** Step 2 で求めた一次独立なデータで構成した標本特徴空間への写像関数を生成する。また、式 (2.14) よりクラス i ($i = 1, \dots, n$) の固有ベクトル φ_{ik} と固有値 λ_{ik} を導く。
- Step 4** Step 3 で求めた固有ベクトルを各クラス i 部分空間を形成する軸とし、次元数 r_i ($i = 1, \dots, n$) を決定する。
- Step 5** 各クラス部分空間の軸を用いて式 (2.22) より $f_i(\mathbf{x}_j)$ ($i = 1, \dots, n, j = 1, \dots, M$) を計算する。
- Step 6** 式 (3.12) を用いて重み w_i ($i = 1, \dots, n$) を導く。
- Step 7** Step 6 によって導かれた w_{ij} がすべて正ならば学習を終了する。そうでなければ、負となる重みに対応する軸を削除し、Step 5 に戻る。

以下、重みに非負制約を設けた SSLP-SVM を SSLP-SVM (P) とよぶ。

4.2.2 SSLP-SVM における負の重みを許す最適化問題

SSLP-SVM では学習過程で同時に特徴選択を行うため、重みに基づく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いる。負の重みをもつ軸の重要性を確認するため、SSLP-SVM の非負制約を省いたモデルを考える必要がある。そこで負の重みを許すために式 (3.24)-(3.26) で与えられる最適化問題から式 (3.26) で与えられる制約条件を省く。このとき、SSLP-SVM の最適化問題は以下のように表せる。

$$\text{minimize } Q(\mathbf{w}, \mathbf{b}, \boldsymbol{\xi}, \mathbf{u}) = \sum_{i=1}^n \sum_{k=1}^{r_k} w_{ik} + \sum_{j=1}^M \sum_{\substack{i \neq y_j, \\ i=1}}^n \frac{C M}{n |X_{y_j}|} \xi_{ji} \quad (4.1)$$

$$\text{subject to } \mathbf{w}_{y_j}^T \mathbf{f}_{y_j}(\mathbf{x}_j) + (b_{y_j}^+ - b_{y_j}^-) - \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}_j) - (b_i^+ - b_i^-) = 1 - \xi_{ji} + u_{ij} \\ \text{for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (4.2)$$

$$\xi_{ji}, u_{ji} \geq 0 \text{ for } i \neq y_j, i = 1, \dots, n, j = 1, \dots, M \quad (4.3)$$

学習アルゴリズムは通常の SSLP-SVM のアルゴリズム 4, 5 とほぼ同じであるため、省略する。以下、非負制約を用いた SSLP-SVM, 負の重みを許す SSLP-SVM をそれぞれ SSLP-SVM (P), SSLP-SVM (N) とよぶ。

4.2.3 重みの絶対値による選択処理

図 4.2 のような 2 クラス問題において、クラス 1, 2 それぞれの部分空間を構成する軸が 2 本ずつ得られたとする。ここで φ_{ik} はクラス i 部分空間における k 番目の軸を意味する。そして、 φ_{11} と φ_{21} はそれぞれのクラスに属するデータを最もよく表現できる軸であり、それぞれの軸に対応する固有値は $\lambda_{11} \gg \lambda_{12}, \lambda_{21} \gg \lambda_{22}$ であるとする。通常の KSM では、固有値の大きい順から部分空間を選択するため、 $\lambda_{12}, \lambda_{12}$ のような値の小くなる軸は削減される。しかしながら、図 4.2 より、 φ_{11} と φ_{21} は非常に似た軸である。これらの軸のみが部分空間として残された場合、明らかに識別が困難となる。すべての軸を SVM によって最適化すると φ_{11} と φ_{21} は識別において冗長であるため、重みの絶対値がそれぞれ正負に関わらず小さくなる。一方、 φ_{12} と φ_{22} は識別に有用であることから重みの絶対値がそれぞれ正負に関わらず大きくなるのがわかる。そこで識別の観点から最適な軸の選択処理を行うため、正負に関わらず重みの絶対値が非常に小さい軸を識別の観点から不必要な軸とみなし削除する手法を考える。すべての重みの絶対値の総和と比較して重みの絶対値が非常に小さい軸は識別に関連がない軸、すなわち冗長な軸である。これらの軸を削除することにより識別に有用な

軸のみを選択することができ、識別の観点から最適な軸の選択といえる。ここで、特徴選択の基準としてある重みの絶対値の自乗がすべての重みの絶対値の自乗の総和の閾値 T 倍よりも小さくなれば、その軸を削除する。また、本論文では閾値 $T = 10^{-5}, 10^{-6}$ から五分割交差検定法により T を決定する。

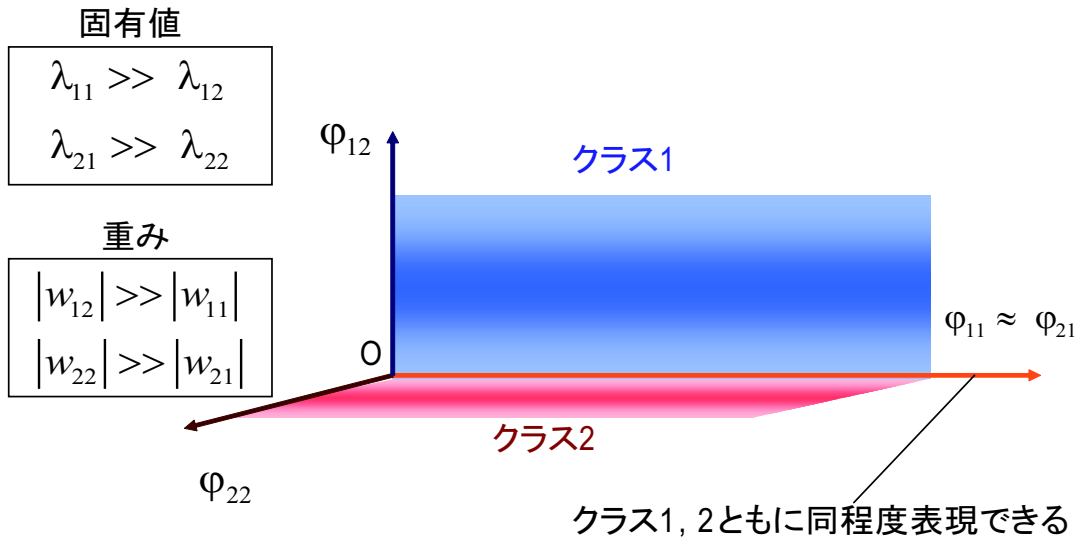


図 4.2 重みの絶対値と軸の関係

以下に SSSL-SVM (Ab) のアルゴリズムを示す。

アルゴリズム 9

- Step 1** 五分割交差検定を用いて, γ, C, T の値をそれぞれ決定する。寄与率 $\kappa = 99.9\%$ 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って教師データから一次独立なデータを選択する。
- Step 3** Step 2 で求めた一次独立なデータで構成した標本特徴空間への写像関数を生成する。また、式 (2.14) よりクラス i ($i = 1, \dots, n$) の固有ベクトル φ_{ik} と固有値 λ_{ik} を導く。
- Step 4** Step 3 で求めた固有ベクトルを各クラス i 部分空間を形成する軸とし、次元数 r_i ($i = 1, \dots, n$) を決定する。 $d = 1$ とおく。
- Step 5** $f_i(x_j)$ ($i = 1, \dots, n, j = 1, \dots, M$) を計算する。
- Step 6** 式 (3.12) を用いて重み w_i ($i = 1, \dots, n$) を導く。
- Step 7** すべての w_{ij} の絶対値の自乗がすべての重みの絶対値の自乗の総和の T 倍より大きければ学習を終了する。そうでなければ、 T より絶対値が小さくなる重

みに対応する軸を削除し, Step 5 から繰り返す.

以下, 重みの絶対値に基づく特徴選択手法を用いた SSLS-SVM を SSLS-SVM (Ab) とよぶ.

4.3 計算機実験

本節では, ベンチマークデータセット^{29, 30)}を用いた計算機実験により負の重みをもつ軸の重要性の評価を行うために, 非負制約を設けた SS-SVM と非負制約を設けない SS-SVM を比較する. また, SSLS-SVM (Ab) と他の SS-SVM とを比較することにより, 重みの絶対値の小さい軸は識別において冗長であるか確認すると共に通常の SVM とも比較することによりこれらの特徴選択手法を用いた SS-SVM の汎化能力の評価を行う. ただし, 多クラスデータセットにおいて各 SSLP-SVM の結果は用いている計算機のパフォーマンス上, 大半のデータセットの計算不可能なため, 比較的小さなデータセットである Iris データセットと Numeral データセットのみの結果を示す. さらに, SSLS-SVM (P) と SSLS-SVM (Ab) においても計算時間が膨大となり Hiragana-105 データセットと Hiragana-13 データセットの結果が取れなかったため, これらの結果を割愛する. 以下, 固有値に基づく特徴選択手法を用いた従来の SSLS-SVM を SSLS-SVM (A) とよぶ.

4.3.1 識別における負の重みをもつ軸の重要性の評価

従来の固有値に基づく特徴選択手法を用いた SSLS-SVM (A), 非負制約条件を設けた SSLP-SVM (P) と提案手法である非負制約を設けた SSLS-SVM (P), 重みの絶対値に基づく特徴選択手法を用いた SSLS-SVM (Ab), 非負制約条件を省いた SSLP-SVM (N) を比較・評価する. 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 を五分割交差検定を用いて決定する. ここで3章で用いていたパラメータ値を選択候補として用いる. また, T は $T = \{10^{-5}, 10^{-6}\}$ から選択する. 表 4.1 に SSLS-SVM (P), SSLP-SVM (N) のパラメータ値を示す.

表 4.2 にテストデータセットにおける SSLS-SVM (A), SSLS-SVM (P), SSLP-SVM (P), SSLP-SVM (N) の平均認識率と標準偏差を示す. ただし, 各多クラスデータセットはそれぞれ組が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, SSLS-SVM (A) または SSLP-SVM (P) と比べ, 高い認識率を示した各データセットにおける SSLS-SVM (P), SSLP-SVM (N) の結果の前にアスタリスク (*) を付ける.

2クラス問題

2クラス問題では, SSLS-SVM (P) は8個のデータセットにおいて最も高い平均認識率を示しており, SSLS-SVM (A) と比べて8個のデータセットにおいて高い平均認識率を示している. また, Welch の t 検定 (有意水準: 5%) により SSLS-SVM (A) と比べて7個のデータセットにおいて有意性を示し, 3個のデータセットにおいて有意差なしと判定されている. これにより, SSLS-SVM (P) は SSLS-SVM (A) よりも汎化能力が優れているといえる. ただし, SSLS-SVM (A) では固有値を基準に特徴選択をしており, 一方 SSLS-SVM (P) では, 重みを基準に特徴選択をしている. この結果から固有値ではなく重みを基準にした特徴選択の方が優れているといえる. 4.3.2 節にて SSLS-SVM (P) と SSLS-SVM (Ab) を比較することで SSLS-SVM において負の重みをもつ軸の重要性を確認する.

また, SSLP-SVM (N) は4個のデータセットにおいて最も高い平均認識率を示しており, SSLP-SVM (P) と比べても9個のデータセットにおいて高い平均認識率を示している. また SSLP-SVM (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有意差なしと判定されている. これにより非負制約を設けた SSLS-SVM (P) よりも非負制約を省いた SSLP-SVM (N) が識別において優れているといえ, SSLP-SVM において重みが負となる軸は識別の観点から必要となる.

多クラス問題

多クラス問題では SSLS-SVM (P) は SSLS-SVM (A) と比べて1個のデータセットにおいて高い認識率, 1個のデータセットにおいて同じ認識率, 2個のデータセットにおいて低い認識率を示している. ただし, Numeral データセットでは低い認識率を示したが, 誤差が 0.1 (%) とほぼ同等であるといえる. Hiragana-50 データセットにおいては SSLS-SVM (A) と比べて 1.2 (%) と汎化能力が大幅に向上している. このため, 多クラス問題においてはほぼ同等の汎化能力をもつといえる.

SSLP-SVM (N) は SSLS-SVM (P) と比べて, すべてのデータセットにおいて認識率を示している. これより, 多クラス問題においても SSLP-SVM (N) は SSLP-SVM (P) と比べて同等以上の汎化能力をもち, 重みが負となる軸が識別の観点から必要であるといえる.

表 4.1 SSLS-SVM (P) と SSLP-SVM (N) におけるパラメータ値

Data	SSLS-SVM (P)		SSLP-SVM (N)	
	γ	C	γ	C
Banana	1	100	0.5	10
B. cancer	1.5	5×10^{-4}	0.1	0.1
Diabetes	0.01	10	0.01	1
German	3	0.01	0.5	1
Heart	0.01	0.01	0.5	10
Image	15	100	0.5	100
Ringnorm	0.5	1	0.1	0.01
F. solar	0.1	10	0.5	1
Splice	1	1	0.5	0.5
Thyroid	1.5	0.1	0.5	1
Titanic	3	10^{-4}	0.01	1
Twonorm	0.1	0.1	0.1	0.1
Waveform	1	0.1	1	0.1
Iris	15	100	0.01	10^4
Numeral	20	10	1	100
Blood cell	200	5	–	–
Hiragana-50	50	1	–	–

表 4.2 テストデータにおける認識率 (%) と標準偏差の比較 (負の重みの重要性の評価)

Data	SSLS-SVM (A)	SSLS-SVM (P)	SSLP-SVM (P)	SSLP-SVM (N)
Banana	88.9±0.6	*89.3±0.4	89.0±0.6	*89.3±0.5
B. cancer	75.1±4.4	75.1±4.2	73.3±4.6	68.6±4.7
Diabetes	72.0±2.3	*75.2±2.0	73.5±2.0	*74.7±2.0
German	73.8±2.2	*75.2±2.3	70.9±6.6	*71.1±2.3
Heart	82.6±3.7	*83.9±3.3	83.1±3.8	*84.0±3.4
Image	95.4±0.6	*96.3±0.7	95.1±1.0	*95.7±0.9
Ringnorm	97.6±0.3	96.6±0.4	98.2±0.2	*98.5±0.1
F. solar	66.9±1.6	65.8±1.8	64.7±1.9	*66.7±1.8
Splice	86.2±1.0	*88.9±0.6	88.2±0.6	*88.9±0.5
Thyroid	95.4±2.1	95.3±2.3	96.3±2.2	95.4±2.7
Titanic	76.9±1.0	*77.0±1.5	76.8±1.1	*76.9±1.0
Twonorm	97.7±0.1	97.6±0.1	97.6±0.2	97.6±0.1
Waveform	89.1±0.8	*89.8±0.5	89.3±0.7	89.5±0.6
Iris	96.0	96.0	90.7	90.7
Numeral	99.4	99.3	99.3	99.3
Blood cell	92.9	91.5	–	–
Hiragana-50	95.5	96.7	–	–

4.3.2 重みの絶対値に基づいた選択手法の評価

本節では, SSLS-SVM (Ab) を他の SSLS-SVM と比較することにより重みの絶対値に基づいた特徴選択の評価を行う. また, 同時に SSLS-SVM における負の重みをもつ軸の重要性も確認する.

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 を五分割交差検定を用いて決定する. ここで各パラメータ値の選択候補は 4.3.1 節で用いていた候補と同じにする. 表 4.3 にそれぞれのパラメータ値を示す.

表 4.3 SSLS-SVM (Ab) におけるパラメータ値

Data	SSLS-SVM (Ab)		
	γ	C	T
Banana	1	100	10^{-6}
B. cancer	1.5	5×10^{-4}	10^{-5}
Diabetes	0.05	0.1	10^{-5}
German	3	0.1	10^{-6}
Heart	0.01	0.01	10^{-5}
Image	1.5	100	10^{-6}
Ringnorm	0.5	1	10^{-5}
F. solar	0.1	1	10^{-5}
Splice	1	1	10^{-6}
Thyroid	0.5	50	10^{-5}
Titanic	1.5	10^{-4}	10^{-6}
Twonorm	1	0.01	10^{-5}
Waveform	1	0.1	10^{-5}
Iris	10	100	10^{-6}
Numeral	10	10^3	10^{-6}
Blood cell	20	5000	10^{-5}
Hiragana-50	50	5	10^{-5}

表 4.4 にテストデータセットにおける SSLS-SVM (A), SSLS-SVM (P), SSLS-SVM (Ab) 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. 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す. また SSLS-SVM (A) または SSLS-SVM (P) と比べて高い認識率を示す各データセットにおける SSLS-SVM (Ab) の結果の前にアスタリスク (*) を付ける.

2クラス問題

2クラス問題では, SSLS-SVM (Ab) は9個のデータセットにおいて最も高い平均認識率を示している. SSLS-SVM (A), SSLS-SVM (P) と比べてそれぞれ10個, 9個のデータセットにおいて高い平均認識率を示している. また SSLS-SVM (A) と比べて9個のデータセットにおいて Welch の t 検定 (有意水準: 5%) により有意性を示し, 4個のデータセットにおいて有意差なしと判定されている. これにより, 従来の各クラスの各軸における固有値を基にする特徴選択手法と比べて重みに基づく特徴選択手法が識別において優れており, 特に SSLS-SVM (P) と比べても優れているといえ, 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.

多クラス問題

多クラス問題では, SSLS-SVM (Ab) は3個のデータセットにおいて最も高い認識率を示しており, Hiragana-50 データセットにおいても SSLS-SVM (A) と比べて高い認識率を示している. これより, 多クラス問題においても重みの絶対値に基づく特徴選択手法は固有値に基づく特徴選択手法に比べて安定して高い汎化能力をもつことがわかる. また SSLS-SVM (P) と比べても大半のデータセットにおいて高い認識率を示していることから, 多クラス問題においても負の重みに対応する軸が識別の観点から重要となりうるといえる.

表 4.4 テストデータにおける認識率 (%) と標準偏差の比較 (SSLS-SVM (P) と SSLS-SVM (Ab) の比較)

Data	SSLS-SVM (A)	SSLS-SVM (P)	SSLS-SVM (Ab)
Banana	88.9±0.6	89.3±0.4	*89.5±0.4
B. cancer	75.1±4.4	75.1±4.2	75.0±4.1
Diabetes	72.0±2.3	75.2±2.0	75.1±2.2
German	73.8±2.2	75.2±2.3	74.4±2.1
Heart	82.6±3.7	83.9±3.3	83.9±3.2
Image	95.4±0.6	96.3±0.7	*97.0±0.6
Ringnorm	97.6±0.3	96.6±0.4	*98.2±0.2
F. solar	66.9±1.6	65.8±1.8	66.5±1.6
Splice	86.2±1.0	88.9±0.6	*89.0±0.5
Thyroid	95.4±2.1	95.3±2.3	*96.2±1.9
Titanic	76.9±1.0	77.0±1.5	*77.3±0.3
Twonorm	97.7±0.1	97.6±0.1	*97.7±0.1
Waveform	89.1±0.8	89.8±0.5	*90.0±0.5
Iris	96.0	96.0	*97.3
Numeral	99.4	99.3	*99.5
Blood cell	92.9	91.5	*93.0
Hiragana-50	95.5	*96.7	*95.8

4.3.3 部分空間を構成する各軸に対応する固有値と重みの関係

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、図 4.3 に Thyroid データセットにおけるそれぞれの固有値と SLS-SVM によって最適化された重みを表す。図 4.3 (a) に各軸に対する固有値を示し、図 4.3 (b) に SLS-SVM によって最適化された重みをそれぞれ示している。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中で、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な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といえる。

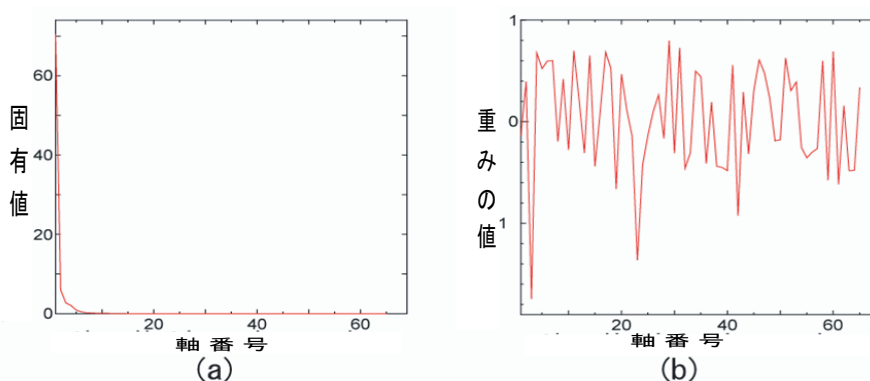


図 4.3 Thyroid データにおける各軸に対応する固有値と重みの関係 (a) 各軸に対する固有値 (b) 各軸に対する重み

4.3.4 特徴選択による軸の削減割合

本実験において、SLS-SVM (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。本節では、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 4.5 に 2 クラスベンチマークデータにおける SLS-SVM (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 4.5 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 個となっている。これよ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といえる。特に Ringnorm データセットでは表より SSLS-SVM (A) では $\kappa = 99.9(\%)$ では軸の削減が行われていないが、SSLS-SVM (Ab) により 74.7 (%) も軸が削減されている。さらに表 4.4 より特徴選択によって識別性能が向上していることがわかる。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 の値が $T = \{10^{-5}, 10^{-6}\}$ 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といえる。

表 4.5 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 (%)

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

4.3.5 正規直交系の標本特徴空間上での識別との評価

SS-SVM では 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。^{32, 33)}

$$\mathbf{h}(\mathbf{x}) = \Lambda^{-1/2} P^T (H(\mathbf{x}_1, \mathbf{x}), \dots, H(\mathbf{x}_M, \mathbf{x})) \quad (4.4)$$

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 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。^{32, 33)} このため、標本特徴空間を構成する軸が直交していないことによって汎化能力の影響はないかを式 (2.14)

に式 (4.4) 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. 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-SVM (P) と SSLS-SVM (Ab) を用い, 2 クラスベンチマークデータセットを用いて評価する. 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-SVM (P, O), SSLS-SVM (Ab, O) とよぶ. また, マージンパラメータ C 以外に対応するそれぞれの手法のパラメータ値と等しい値を設定し, C のみ五分割交差検定で決定する.

表 4.6 より, SSLS-SVM (P), SSLS-SVM (Ab) の 2 つの手法ともそれぞれ SSLS-SVM (P, O), SSLS-SVM (Ab, O) と比べて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いる. これより, 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. しかしながら, いくつかのデータセットにおいて若干の差が見られる. これは本実験において SSLS-SVM (P) と SSLS-SVM (Ab) では一次独立なデータの選定にコレスキー分解³²⁾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 としており, 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ためと考えられる. しかしながら, その差の小ささから問題にする必要はない. この実験より, 標本特徴空間の基底ベクトルを直交化せずとも同等の汎化能力を示すことが確認できる.

表 4.6 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

Data	SSLS-SVM (P)	SSLS-SVM (P, O)	SSLS-SVM (Ab)	SSLS-SVM (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

4.3.6 KSM と SVM との比較・評価

表 4.7 において通常の SVM と KSM とそれぞれの特徴選択手法を用いた SSLS-SVM と SSLP-SVM の平均認識率 (%) 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 よりも高い平均認識率の前に (*) を付

ける。また、“vs KSM”, “vs SVM” はそれぞれ KSM, SVM と比較して平均認識率が“(高いデータセット数) : (等しいデータセット数) : (低いデータセット数)”を示す。

2 クラス問題

2 クラス問題では, SSLS-SVM (P) は KSM と比べて 8 個のデータセットにおいて高い平均認識率を示し, 3 個のデータセットにおいて同じ平均認識率を示している。その他のデータセットにおいてもほぼ同等な認識率を示しており, Welch の t 検定 (有意水準: 5%) によって有意差なしと判定されている。これより, SSLS-SVM (P) は KSM と比べて 2 クラス問題において汎化能力が高いといえる。SVM と比べると 2 個のデータセットにおいて高い平均認識率を示し, 2 個のデータセットにおいて同じ平均認識率を示している。しかしながら, その他のデータセットにおいて低い平均認識率を示している。

SSLS-SVM (Ab) は KSM と比べて 10 個のデータセットにおいて高い平均認識率を示し, 1 個のデータセットにおいて同じ平均認識率を示している。その他のデータセットにおいても有意差なしと判定されたことから, SSLS-SVM (Ab) も KSM と比べて極めて汎化能力の高い手法であるといえる。また, SVM と比べても 7 個のデータセットにおいて高い平均認識率を示しており, 1 個のデータセットにおいて同じ平均認識率を示している。また SVM と比べて 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 2 クラスベンチマークデータセットにおいて SSLS-SVM (Ab) は SVM と比べ, 同等以上の高い汎化能力をもつ識別器であるといえる。

SSLP-SVM (N) は KSM と比べて 7 個のデータセットにおいて高い平均認識率を示し, 1 個のデータセットにおいて同じ平均認識率を示している。有意性の観点から B. cancer データセットと German データセットにおいて劣っていると判定されたが, それ以外では有意差なしと判定されている。これらより, SSLP-SVM (N) は KSM と比べて汎化能力が優れた識別器といえる。SVM と比べると 2 個のデータセットにおいて高い平均認識率を示し, 2 個のデータセットにおいて同じ平均認識率を示している。しかしながら, その他のデータセットにおいて低い平均認識率を示している。

多クラス問題

多クラス問題では, SSLS-SVM (P) は KSM と比べて, すべてのデータセットにおいて高い認識率を示している。しかしながら, SVM と比べてすべてのデータセットにおいて低い認識率を示している。このことから SVM と比べて汎化能力の低い識別手法であるといえる。一方, SSLS-SVM (Ab) は KSM と比べて, すべてのデータセットにおいて高い認識率を示している。SVM と比べると Blood cell データセットにおいて高い

認識率, Iris データセットにおいて同じ認識率を示している. Numeral データセットにおいては低い認識率を示しているが, 誤差が0.1 (%) であり, ほぼ同等といえる. また, SSLP-SVM (N) は KSM と比べて Numeral データセットにおいてのみ高い認識率を示している. このことから多クラス問題においても SSLP-SVM の汎化能力はそれほど高くないといえる.

これらの結果より, これらの SS-SVM は KSM と比べて非常に高い汎化能力をもち, SSLS-SVM (Ab) に関しては SVM と比べても同等以上の汎化能力をもつといえる. 本論文の目的である高い汎化能力をもち, 冗長な特徴量を削減でき, さらに識別部が解析可能である3点を満たす非常に優れた識別手法であるといえる.

表 4.7 テストデータセットにおける平均認識率 (%) と標準偏差 (KSM と SVM との比較)

Data	SVM	KSM	SSLS-SVM			SSLP-SVM	
			(A)	(P)	(Ab)	(P)	(N)
Banana	89.3±0.5	88.6±0.6	88.9±0.6	89.3±0.4	*89.5±0.4	89.0±0.6	89.3±0.5
B. cancer	72.4±4.6	*75.1±4.4	*75.1±4.4	*75.1±4.2	*75.0±4.1	*73.3±4.6	68.6±4.7
Diabetes	76.3±1.8	73.5±1.8	72.0±2.3	75.2±2.0	75.1±2.2	73.5±2.0	74.7±2.0
German	76.2±2.2	75.1±2.2	73.8±2.2	75.2±2.3	74.4±2.1	70.9±6.6	71.1±2.3
Heart	83.7±3.4	82.4±3.6	82.6±3.7	*83.9±3.3	*83.9±3.2	83.1±3.8	*84.0±3.4
Image	97.3±0.4	96.3±0.6	95.4±0.6	96.3±0.7	97.0±0.6	95.1±1.0	95.7±0.9
Ringnorm	97.8±0.3	76.6±11.2	97.6±0.3	96.6±0.3	*98.2±0.2	*98.2±0.2	*98.5±0.1
F. solar	67.6±1.7	65.1±1.8	66.9±1.6	65.8±1.8	66.5±1.6	64.7±1.9	66.7±1.8
Splice	89.2±0.7	87.4±2.5	86.2±1.0	88.9±0.6	89.0±0.5	88.2±0.6	88.9±0.5
Thyroid	96.1±2.0	95.6±2.1	95.4±2.1	95.3±2.3	*96.2±1.9	*96.3±2.2	95.4±2.7
Titanic	77.2±1.1	77.3±0.6	76.9±1.0	77.0±1.5	*77.3±0.3	76.8±1.1	76.9±1.0
Twonorm	97.6±0.1	97.6±0.1	*97.7±0.1	97.6±0.1	*97.7±0.1	97.6±0.2	97.6±0.1
Waveform	90.0±0.4	88.5±0.6	89.1±0.8	89.8±0.5	90.0±0.5	89.3±0.7	89.5±0.6
Iris	97.3	93.3	96.0	96.0	97.3	90.7	90.7
Numeral	99.6	99.0	99.4	99.3	99.5	99.3	99.3
Blood cell	92.9	91.3	92.9	91.5	*93.0	-	-
Hiragana-50	98.9	95.3	95.5	96.7	95.8	-	-
vs KSM	14:1:2	-	10:1:6	12:3:2	14:1:2	7:2:6	8:1:6
vs SVM	-	2:1:14	2:1:14	2:2:13	8:2:7	3:1:11	3:1:11

4.4 結言

本章では SS-SVM における負の重みをもつ軸の重要性の確認と重みの絶対値に基づいた特徴選択手法を提案した. この特徴選択手法では, SS-SVM によって決定された各軸における重みを識別における重要度とし, それらの値を基に各クラス部分空間を構成する軸を選択する. これにより, 他のクラスを考慮した特徴選択ができる. このとき, 軸の選択における基準として2つの条件を設けた. 1つ目は重みの非負制約条件である. 2つ目は重みの絶対値が閾値より大きくなければならないという条件である. ただし, SSLP-SVM は重みの最適化と同時に特徴選択を行うため, これらの手法を適用しなかった. また, これらの提案手法は SSLS-SVM, SSL2-SVM とともに繰り返し学習が必要となり, SSL2-SVM では計算コストが膨大になると推測できる. そのため SSLS-SVM にのみこれらの特徴選択手法を適用した.

計算機実験により、重みに基づく特徴選択手法を適用することにより SSLS-SVM の汎化能力が大幅に向上し、汎化能力の非常に高い識別器である SVM と同程度であることが確認できた。さらに、非負制約条件を設けた SSLS-SVM では、3 章の SSLS-SVM に比べ汎化能力が向上した。しかしながら、重みの絶対値に基づく特徴選択手法を用いた SSLS-SVM と比べると汎化能力は劣っていた。これにより、負の重みは識別において重要となりうることを確認できた。また、これは SSLS-SVM において非負制約を設けた場合と設けなかった場合を比較することによっても確認できた。さらに、この特徴選択手法により本当に軸の削減がなされているのかを確認するため SSLS-SVM における軸の削減割合を確認した。これにより次元削減が行われ、それにより汎化能力の向上に繋がっていることを確認できた。また、KPCA を適用することに求められる固有値と SS-SVM により最適化された重みの値の関係を調査し、固有値と重みの値はほぼ無関係であるということが確認できた。これにより重みの最適化の重要性を再確認できた。これらより、重みの絶対値に基づく特徴選択を SS-SVM に適用することで汎化能力を SVM と同等以上まで向上させることができ、これらの手法の有効性を検証できた。

第5章

学習の高速化

5.1 緒言

カーネル部分空間法 (KSM: Kernel Subspace Method)^{13, 14)} や部分空間法に基づくサポートベクトルマシン (SS-SVM: Subspace-Based Support Vector Machine)^{23, 24)} では、各クラスの部分空間生成の前処理として全教師データから一次独立なデータを選択し、それらの教師データを用いて標本特徴空間^{32, 33)} を生成し、全教師データをあらかじめ標本特徴空間に写像する。これにより、部分空間生成の際に解くべき固有値問題の計算コストを削減していた。しかしながら、一次独立なデータの選択にはコレスキー分解を用いており、大規模データセットにおいて、この選択処理における計算コストが膨大となる。またクラスごとにカーネル主成分分析 (KPCA: Kernel Principal Component Analysis)^{10, 11)} を用いるため、共分散行列の固有値問題を解く必要がある。このときクラスごとにデータ数の膨大なデータセット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。したがって、部分空間生成の際における学習コストの削減が重要となる。

一方、固有値分解を解かずにクラスごとに部分空間を生成する部分空間法として錐制約部分空間法 (CRSM: Cone Restricted Subspace Method)¹⁸⁾ が提案されている。CRSM では、各クラスに属する正規化された一次独立な教師ベクトルを凸錐の部分空間を構成する軸とする。そして、未知の入力ベクトルに対して、それらのクラスごとの部分空間との角度を類似度として識別処理を行う。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¹⁾ 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、クラス部分空間同士で重なりができる。その重複した領域は未分類領域となる。

そこで本章では、これらの問題を解消するための、SS-SVM における2つの高速学習手法^{25, 26)} を提案する。1つ目は、前述のKSM, SS-SVM における前処理である標本特徴空間の生成の際にも部分空間生成と同様にクラスごとに標本特徴空間を生成することにより、計算コストを削減する手法²⁵⁾ である。各クラスの部分空間は元々そのクラスに属する教師データのみにより生成されるため、全データではなく各クラスのデータのみを用いた生成した標本特徴空間上で部分空間を生成してもそれぞれ同じク

ラス部分空間を生成することができる。2つ目は、標本部分空間法に基づくサポートベクトルマシン (ES-SVM: Empirical Subspace-Based Support Vector Machine)²⁶⁾ である。まず、CRSM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、CRSM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。これにより、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。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 (ESM: Empirical Subspace Method) とよぶ。ES-SVMでは、SS-SVMと同様にESMによって生成された各クラス部分空間同士の分離の最大化と誤識別の最小化を行うことにより識別の観点から最適な重み付けを行う。この手法により、SS-SVMと比べて部分空間の生成における学習コストが削減できる。さらにCRSMと比べて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。

5.2 クラス標本特徴空間の生成による学習の高速化

通常、SS-SVMでは部分空間生成の前処理として、全教師データから一次独立なデータを選択し、それらのデータから標本特徴空間を生成する。これにより、部分空間生成における固有値問題のサイズを縮小できる。しかしながら、教師データ数が膨大な場合、一次独立なデータも膨大になる恐れがあり、さらにそれらの一次独立なデータの選択も困難となる。一次独立なデータが膨大になれば、解くべき固有値問題のサイズも非常に大きくなり、膨大な学習コストが必要となる。

そこで、SS-SVMの標本特徴空間生成と部分空間生成における学習コスト削減のためにクラスごとに標本特徴空間を生成する高速学習方式を提案する。SS-SVMではクラスごとのデータを用いてそのクラスを表すことのできる必要最低限の特徴量を抽出・選択することにより部分空間を生成する。そのため、他のクラスに属するデータを表現できるような特徴量は必要ない。そこで、クラスごとに標本特徴空間を生成し、そのクラスの部分空間を生成するとしても部分空間に全教師データを用いた標本特徴空間上で生成した部分空間とほとんど相違ない。これにより、部分空間生成の際の固有値問題のサイズが非常に小さくなるため学習コストの削減が行える。

写像関数 $\mathbf{h}(\mathbf{x})$ の代わりにクラスごとの写像関数 $\mathbf{h}_i(\mathbf{x})$ とおき、 $\mathbf{h}_i(\mathbf{x})$ を

$$\mathbf{h}_i(\mathbf{x}) = \left(H(\mathbf{x}_{ik_1}, \mathbf{x}), \dots, H(\mathbf{x}_{ik_{N_i}}, \mathbf{x}) \right)^T \quad (5.1)$$

と表す。ここで、 N_i はクラス i の特徴空間における一次独立なデータ数を表し、 \mathbf{x}_{ik_j} ($ik_j \in X_i, j = 1, \dots, N_i$) はクラス i の一次独立なデータを表す。式 (2.14) を解く代わりに

$\varphi_{ik} (i = 1, \dots, n, k = 1, \dots, r_i)$ を以下のように導く.

$$\frac{1}{|X_i|} \sum_{j \in X_i} \mathbf{h}_i(\mathbf{x}_j) \mathbf{h}_i^T(\mathbf{x}_j) \varphi_{ik} = \lambda_{ik} \varphi_{ik}^T \quad (5.2)$$

このように、解くべき固有値問題の行列のサイズは $N_i \times N_i$ 、学習コストも非常に削減することができる.

しかしながら、単純に式 (2.22) における $\mathbf{h}(\mathbf{x})$ を $\mathbf{h}_i(\mathbf{x})$ に変換するだけでは、正しい類似度を求めることができない. 図 5.1 を用いて説明する. 図ではクラス i における部分空間はひとつの φ_i のみで構成されている. ここで、 \mathbf{x} はクラス i 以外のクラスに属するある入力とし、 $\mathbf{g}(\mathbf{x})$, $\mathbf{h}_i(\mathbf{x})$ の φ_i との角度をそれぞれ α , α' とする. このとき、 $\mathbf{g}(\mathbf{x})$ のクラス i 部分空間における類似度を得たいが、単純に式 (2.22) における $\mathbf{h}(\mathbf{x})$ を $\mathbf{h}_i(\mathbf{x})$ に変換では明らかに $\alpha \neq \alpha'$ となる. このとき、 $\mathbf{h}_i(\mathbf{x})$ のクラス i 部分空間における射影長は $\mathbf{g}(\mathbf{x})$ のクラス i 部分空間における射影長と等しくなるが類似度は異なるため正しい類似度が求めることができていないことがわかる.

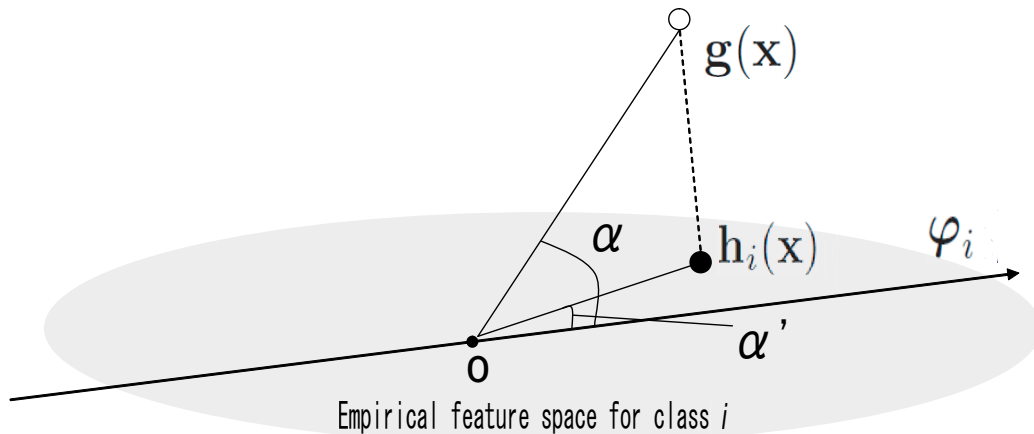


図 5.1 標本特徴空間の縮小

そこで $\mathbf{g}(\mathbf{x})$ のクラス i 部分空間における類似度を以下のように表す.

$$\mathbf{f}_i(\mathbf{x}) = \frac{(\varphi_i^T \mathbf{g}(\mathbf{x}))^2}{\|\varphi_i\|^2 \|\mathbf{g}(\mathbf{x})\|^2} = \frac{(\varphi_i^T \mathbf{g}(\mathbf{x}))^2}{\|\varphi_i\|^2 H(\mathbf{x}, \mathbf{x})} \quad (5.3)$$

ここで φ'_i は φ_i の高次元特徴空間への拡張した固有ベクトルを表す. そして $\mathbf{h}_i(\mathbf{x})$ のクラス i 部分空間における類似度は

$$\mathbf{f}'_i(\mathbf{x}) = \frac{(\varphi_i^T \mathbf{h}_i(\mathbf{x}))^2}{\|\varphi_i\|^2 \|\mathbf{h}_i(\mathbf{x})\|^2} \quad (5.4)$$

となる. $\mathbf{g}(\mathbf{x})$ のクラス i 部分空間における射影長は $\mathbf{h}_i(\mathbf{x})$ となり、

$$\varphi_i^T \mathbf{g}(\mathbf{x}) = \varphi_i^T \mathbf{h}_i(\mathbf{x}) \quad (5.5)$$

である。したがって、スケールファクターは $\|\mathbf{h}_i(\mathbf{x})\|^2/H(\mathbf{x}, \mathbf{x})$ となり、類似度は

$$\begin{aligned} \mathbf{f}_i(\mathbf{x}) &= \frac{\|\mathbf{h}_i(\mathbf{x})\|^2}{H(\mathbf{x}, \mathbf{x})} \left(\frac{(\varphi_{i1}^T \mathbf{h}_i(\mathbf{x}))^2}{\|\varphi_{i1}\|^2 \|\mathbf{h}_i(\mathbf{x})\|^2}, \dots, \frac{(\varphi_{ir_i}^T \mathbf{h}_i(\mathbf{x}))^2}{\|\varphi_{ir_i}\|^2 \|\mathbf{h}_i(\mathbf{x})\|^2} \right)^T \\ &= \left(\frac{(\varphi_{i1}^T \mathbf{h}_i(\mathbf{x}))^2}{\|\varphi_{i1}\|^2 H(\mathbf{x}, \mathbf{x})}, \dots, \frac{(\varphi_{ir_i}^T \mathbf{h}_i(\mathbf{x}))^2}{\|\varphi_{ir_i}\|^2 H(\mathbf{x}, \mathbf{x})} \right)^T \end{aligned} \quad (5.6)$$

となる。

以下に学習アルゴリズムを示す。

アルゴリズム 10

Step 1 κ, γ を五分割交差検定法により決定する。 $i = 1$ を設定する。

Step 2 Step 1 で決定した γ を用いて、コレスキー分解によりクラス i における一次独立なデータを選択する。

Step 3 Step 2 により得られた一次独立なデータを用いてクラス i における標本特徴空間を生成する。式 (5.2) によりクラス i の固有ベクトル φ_{ik} と固有値 λ_{ik} を求める。

Step 4 クラス i の次元数 r_i を Step 1 により決定した κ の値を用いて決定する。

Step 5 式 (5.9) により $\mathbf{f}_i(\mathbf{x}_j)$ ($j = 1, \dots, M$) を決定する。もし $i = n$ ならば、アルゴリズムを終了する。もし $i < n$ ならば、 $i = i + 1$ とおき Step 2 へ戻る。

しかしながら、この手法においても部分空間の生成に KPCA を用いるため、固有値問題を解く必要がある。このため、各クラスのデータ数がそれぞれ膨大となれば、学習コストも膨大となりうる。そこで、次節に部分空間の生成に KPCA を用いる必要のない錐制約部分空間法について述べる。

5.3 錐制約部分空間法

通常の KSM や SS-SVM 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。前節ではこの固有値問題で解くべき共分散行列のサイズを小さくすることで学習の高速化を図った。しかしながら、それでもなおクラスごとの一次独立なデータ数 N_i ($i = 1, \dots, n$) 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_i の値が大きければ、学習コストが膨大となる。

一方、CRSM では特徴ベクトルの要素が全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く。そして、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。図 5.2 に 2 次元 2 クラス問題について例を示して説明する。

はそれぞれクラス1, クラス2に属する教師データを示しており, まずクラスごとに一次独立なデータを選択する. そして, それぞれのクラスにおいて一次独立なデータによって覆われた凸錐の空間を部分空間とする. また, 前述までの部分空間法とは異なり, 射影長ではなく各クラスにおける凸錐との角度を類似度とする. その角度が小さければ小さいほどそのクラスに近いとみなし, 最も角度の小さいクラスに未知のデータを分類する. ここで, x は未知のデータを意味しており, 図 5.2 の例ではクラス2に分類される. またこのような2次元の部分空間の場合, 分離超平面は凸錐間の2等分線となることが明らかである.

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. このため, 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全て削減することができる. 通常の CRSM では以下のように入力 x とクラス i 部分空間への正射影ベクトルとの角度 $\beta_i(x)$ を決定関数 $D_i(x)$ とする.

$$D_i(x) = \beta_i(x) = \arcsin \left(\sqrt{\frac{\min_{\alpha_{ij} \geq 0} \|\mathbf{x} - \sum_{j \in X_i} \alpha_{ij} \mathbf{x}_j\|^2}{\|\mathbf{x}\|^2}} \right) \quad (5.7)$$

ここで, α_{ij} は教師ベクトル x_j における係数を表す.

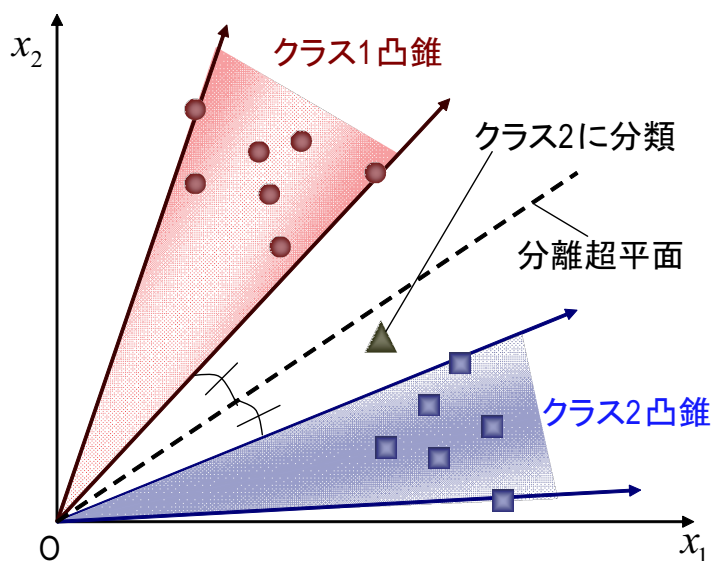


図 5.2 CRSM による識別

しかしながら, 通常のパターン認識問題では線形分離可能な問題はほとんど現実にはなく, 非線形問題である. 各クラス凸錐部分空間において, 内部に位置するデータは決定関数 $D_i(x) = 0$ となる. このとき, 図 5.3 のように, 各クラス凸錐部分空間同士で重なる領域は未分類領域となる. この問題点を解消するため, カーネル法を用いることによりこのような部分空間同士の重なりを最小限に抑えることができる. しかしな

がら、過学習となる可能性もあり、汎化能力が低下する恐れがある。

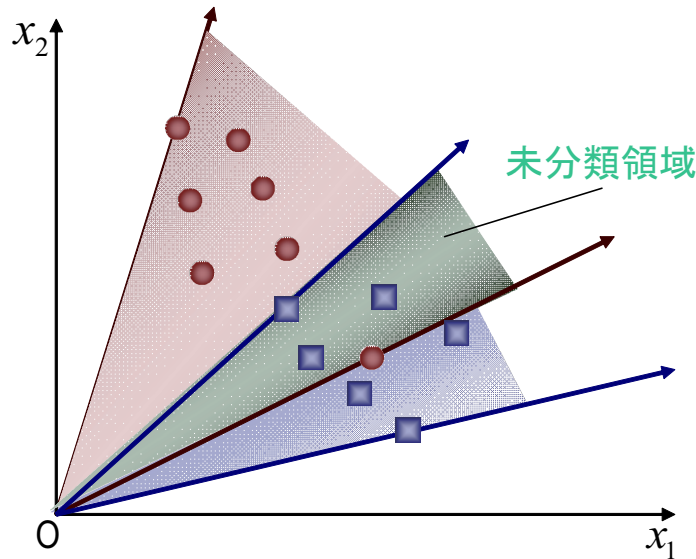


図 5.3 CRSM の問題点

また、この決定関数 $D_i(\mathbf{x})$ は未知の入力がある度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。

5.4 標本部分空間法に基づくサポートベクトルマシン

重複する領域を小さくするためカーネル法を用いる。さらに CRSM 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 ESM を考える。そして、ESM に SS-SVM と同様に SVM のマージン最大化と誤認識最小化の概念を導入する ES-SVM を提案する。

はじめに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。

$$D_i(\mathbf{x}) = \sum_{j \in X_i} \frac{w_{ij} \mathbf{g}^T(\mathbf{x}) \mathbf{g}(\mathbf{x}_{i_j})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{i_j})\|} = \sum_{j=1}^{N_i} \frac{w_{ij} H(\mathbf{x}, \mathbf{x}_{i_j})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{i_j})\|} \quad (5.8)$$

ここで、高次元特徴空間上のベクトル \mathbf{x} のノルムはカーネル $\|\mathbf{g}(\mathbf{x})\| = \sqrt{H(\mathbf{x}, \mathbf{x})}$ により容易に求めることができる。また類似度 $f_i(\mathbf{x})$ は

$$\mathbf{f}_i(\mathbf{x}) = \left(\frac{H(\mathbf{x}, \mathbf{x}_{i_1})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{i_1})\|}, \dots, \frac{H(\mathbf{x}, \mathbf{x}_{i_{|X_i|}})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{i_{|X_i|}})\|} \right)^T \quad (5.9)$$

と定義でき、決定関数を式 (2.23) と同様に表せる。ここで重み w_{ij} に SVM を用いて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い。このとき、内積のみを求めればよいから、CRSM に比べて識別コストの大幅な削減が期待できる。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、よりよ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 (5.8) 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。

$$D_i(\mathbf{x}) = \sum_{j=1}^{N_i} \frac{w_{ij} H(\mathbf{x}, \mathbf{x}_{ik_j})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_j})\|} \quad (5.10)$$

このとき、類似度を

$$\mathbf{f}_i(\mathbf{x}) = \left(\frac{H(\mathbf{x}, \mathbf{x}_{ik_1})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_1})\|}, \dots, \frac{H(\mathbf{x}, \mathbf{x}_{ik_{N_i}})}{\|\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x}_{ik_{N_i}})\|} \right)^T \quad (5.11)$$

と表すことができる。このとき、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、サポートベクトルマシン (SVM: Support Vector Machine)^{5, 6, 7)} を用いてマージン最大化と誤認識最小化の概念の基で SS-SVM と同様に最適化を行うことができる。SS-SVM と同様に用いる SVM のモデルとして最小自乗サポートベクトルマシン (LS-SVM: Least Squares Support Vector Machine) と線形計画サポートベクトルマシン (LP-SVM: Linear Programming Support Vector Machine) を用いる。以下、それぞれの手法を標本部分空間法に基づく LS-SVM (ESLS-SVM: Empirical Subspace-Based Support Vector Machine)、標本部分空間法に基づく LP-SVM (ESLP-SVM: Empirical Subspace-Based Support Vector Machine) とよぶ。また、ESLP-SVM の最適化問題では4章と同様に非負制約を省く。ここで、L2 サポートベクトルマシン (L2-SVM: L2 Support Vector Machine) を適用しない理由は、ES-SVM では固有値問題を解かないことから固有値に基づく特徴選択手法が適用できず、4章で述べた重みに基づく特徴選択手法を用いるか SSLP-SVM のように LP-SVM で重みの最適化と同時に特徴選択を行う方法しかないためである。以下、非負制約を設けた ESLS-SVM を ESLS-SVM (P)、重みの絶対値に基づく特徴選択を用いた ESLS-SVM を ESLS-SVM (Ab) とよぶ。

5.5 計算機実験

ベンチマークデータセット^{29, 30)} を用いた計算機実験により、まずクラスごとに標本特徴空間を生成することによる高速学習手法を従来の SS-SVM と学習時間の比較・評価を行う。また、ES-SVM における学習時間を SS-SVM における学習時間と比較す

ることにより高速化されているかを評価する。さらに、教師データそのものを部分空間を構成する軸として扱うことにより汎化能力が劣化しないかを SS-SVM と比較・評価することにより確認する。

5.5.1 クラス標本特徴空間による高速化の評価

表 5.1 に全てのデータ、またはクラスごとのデータを用いて標本特徴空間を生成したときのその標本特徴空間生成にかかる学習時間とその後 KPCA により前クラス部分空間を生成するためにかかる学習時間を足した時間を示す。表中の “All” は標本特徴空間の生成の際、全データを用いたときの学習コストを表し、“Class” は各クラスのデータを用いたときの学習コストを表す。

2 クラス問題

2 クラス問題では、すべてのデータセットにおいて学習時間が短くなっている。特に Image データセットや Splice データセットなどの比較的大規模なデータセットにおいては大幅な削減が確認できることから、データ数が多ければ多いほど有効性がみられるといえる。

多クラス問題

多クラス問題でも、すべてのデータセットにおいて学習時間が短くなっている。また、Numeral データセットのようにデータ数があまり多くないデータセットに対しても大幅な学習コストが削減されている。これより、クラス数の多いデータセットに対して、この手法はより有効であるといえる。この手法により多クラス問題に対して、部分空間生成における学習コストが非常に低コストで済み、非常に有効性の高い手法であるといえる。

表 5.1 学習時間 (s) の比較

Data	All	Class
Banana	0.60	0.13
B. cancer	0.08	0.01
Diabetes	0.91	0.12
German	2.80	0.58
Heart	0.05	0.01
Image	20.4	1.02
Ringnorm	0.67	0.09
F. solar	1.10	0.51
Splice	9.12	0.89
Thyroid	0.01	0.01
Titanic	0.01	0.01
Twonorm	0.56	0.19
Waveform	0.71	0.14
Iris	0.01	0.01
Numeral	39.8	0.19
Blood cell	259	2.81
Hiragana-50	913	6.03
Hiragana-105	2513	45.1
Hiragana-13	1937	20.7

5.5.2 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

ESLS-SVM (P), ESLS-SVM (Ab), ESLP-SVM に用いるパラメータを4章の計算機実験におけるパラメータの選択候補から五分割交差検定法より決定する。表 5.2 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, Iris データセット, Numeral データセットのみの結果を示す。さらに, ESLS-SVM (P) と ESLS-SVM (Ab) においても計算時間が膨大となり Hiragana-105 データセットと Hiragana-13 データセットの結果が取れなかったため, これらの結果を割愛する。

表 5.3 に ESLS-SVM (P), ESLS-SVM (Ab), ESLP-SVM 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。ここで, 各データセットにおいて最大平均認識率を太字で示し, 2 クラス問題においてのみ4章の SSLS-SVM (P), SSLS-SVM (Ab), SSLP-SVM (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-SVM (P), ESLS-SVM (Ab), ESLP-SVM の結果の前にアスタリスク (*), 有意差なしと判定された結果の前に三角印 (Δ) を付ける。

表 5.2 ES-SVM におけるパラメータ

Data	ESLS-SVM (P)		ESLS-SVM (Ab)			ESLP-SVM	
	γ	C	γ	C	T	γ	C
Banana	3	10	1.5	500	10^{-5}	0.5	100
B. cancer	1	10^{-3}	0.05	10^{-3}	10^{-5}	0.1	1
Diabetes	0.01	500	0.01	500	10^{-6}	0.01	10
German	5	1	5	10	10^{-6}	0.05	1
Heart	0.1	1	0.01	100	10^{-5}	0.1	1
Image	5	5×10^3	5	500	10^{-6}	0.5	1000
Ringnorm	1	10	1	10	10^{-6}	0.1	1
F. solar	0.1	10	0.1	10	10^{-6}	0.5	1
Splice	1.5	100	1	10	10^{-6}	0.5	50
Thyroid	1	10	1	10	10^{-6}	1	10
Titanic	1.5	10^{-3}	1.5	10^{-3}	10^{-6}	1	1
Twonorm	1.5	0.1	0.5	1	10^{-5}	0.05	1
Waveform	1.5	1	1.5	1	10^{-6}	1.5	1
Iris	15	1	3	100	10^{-5}	0.5	500
Numeral	50	1	15	1000	10^{-6}	1	10^3
Blood cell	500	10^3	50	10^4	10^{-5}	-	-
Hiragana-50	50	10^3	100	10^4	10^{-6}	-	-

2 クラス問題

2 クラス問題において ESLS-SVM (P) は SSLS-SVM (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有意差なしと判定されている。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もつといえる。また ESLS-SVM (Ab) は SSLS-SVM (Ab) と比べて 10 個のデータセットにおいて有意差なしと判定されている。有意性を示している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ほぼ同等であるといえる。また, ESLS-SVM (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よいといえる。

また, ESLP-SVM は SSLP-SVM (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有意差なしと判定されている。これより, ESLP-SVM は SSLP-SVM (N) と比べ, 識別能力が若干劣るといえる。

ESLS-SVM (P) と ESLS-SVM (Ab), ESLP-SVM は大半のデータセットにおいて SSLS-SVM (P), SSLS-SVM (Ab), SSLP-SVM (N) と比べてそれぞれほぼ同等の汎化能力を示している。

多クラス問題

多クラス問題では ESLS-SVM (P) は SSLS-SVM (P) と比べて2個のデータセットにおいて高い認識率を示し, その他のデータセットにおいて同じ認識率を示している. これより, 多クラス問題において ESLS-SVM (P) は SSLS-SVM (P) と比べて汎化能力が高いといえる. また, ESLS-SVM (Ab) は SSLS-SVM (Ab) と比べて1個のデータセットにおいてのみ高い認識率を示している. しかしながら, その他のデータセットにおいても認識率は SSLS-SVM (Ab) と比べてほぼ同等である. これより, 多クラス問題においても ESLS-SVM (Ab) は SSLS-SVM (Ab) と比べ, 同等の識別性能といえる. また, ESLS-SVM は SSLP-SVM (N) のそれぞれと比べ, すべてのデータセットにおいて高い認識率を示している.

表 5.3 ES-SVM のテストデータにおける平均認識率 (%) と標準偏差

Data	ESLS-SVM (P)	ESLS-SVM (Ab)	ESLP-SVM
Banana	* 89.5 ±0.4	△ 89.5 ±0.4	89.1±0.5
B. cancer	△73.9±4.7	△ 74.1 ±4.5	△69.0±4.5
Diabetes	△ 75.0 ±2.0	△ 75.0 ±2.0	△ 75.0 ±2.1
German	74.3±2.1	△ 74.7 ±2.5	△71.6±2.5
Heart	△84.0±3.3	△84.0±3.2	△ 84.5 ±3.1
Image	*96.8±0.5	△ 97.0 ±0.6	△95.8±0.6
Ringnorm	*96.8±0.4	96.7±0.4	98.0 ±0.3
F. solar	*66.5±1.5	△66.5±1.6	* 67.2 ±1.7
Splice	△ 89.1 ±0.7	△ 89.1 ±0.6	87.7±0.9
Thyroid	*95.9±1.9	△ 96.0 ±2.0	△94.9±2.3
Titanic	* 77.3 ±0.4	△ 77.3 ±0.3	76.3±2.1
Twonorm	△ 97.6 ±0.1	97.6 ±0.1	97.3±0.2
Waveform	89.7±0.5	89.8 ±0.5	88.8±0.7
Iris	96.0	96.0	*93.3
Numeral	99.3	99.3	* 99.8
Blood cell	* 92.6	92.6	—
Hiragana-50	*96.8	* 98.0	—

5.5.3 ES-SVM の学習時間の比較・評価

表 5.4 に SSLS-SVM (P), SSLS-SVM (Ab), ESLS-SVM (P), ESLS-SVM (Ab), SSLP-SVM (N), ESLP-SVM の各データセットにおける1ファイルあたりの部分空間の生成と重みの最適化における平均学習時間 (s) 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かったため省略する. ここで各 SS-SVM と各 ES-SVM を記載する際, サイズの都合からそれぞれの手法名から “SVM” を省略する. 例えば, SSLS-SVM (Ab) ならば SSLS (Ab) と示す.

2クラス問題

2クラス問題では, ESLS-SVM (P), ESLS-SVM (Ab) は SLS-SVM (P), SLS-SVM (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. ESLS-SVM においても SLS-SVM (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 SLS-SVM 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.

多クラス問題

多クラス問題では, すべての ES-SVM がそれぞれの S-SVM に比べて高速化されている. ただし, 2クラス問題と比べると学習時間の短縮割合は小さい. これは, 用いた多クラスデータセットは各クラスのデータ数が少ないため固有値問題における学習コストが小さいことが原因である. すなわち, 学習時間の大半は重みの最適化部における学習時間であると考えられる.

表 5.4 SS-SVM と ES-SVM の学習時間 (s) の比較

Data	SLS (P)	SLS (Ab)	SLS (N)	ESLS (P)	ESLS (Ab)	ESLS
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	0.09
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	1.13	0.56	162	1.10	0.55	160
Blood cell	109	62.3	–	79.3	60.8	–
Hiragana-50	104	121	–	92	82	–

5.6 結言

本章では, SS-SVM における学習コストの高速化を目的とした手法を 2 つ述べた. 1 つ目は, 部分空間生成の前処理として生成する標本特徴空間を全データではなくクラスごとのデータのみを用いてクラスごとに生成することにより, 一次独立なデータの選択と部分空間生成における学習コストを大幅に削減する手法である. また, この手法による識別処理に対する影響はないことを説明した. 2 つ目は, さらに部分空間生成における学習コストを削減するために固有値分解を解く必要のない CRSM の概念を取り入れた ES-SVM である. この手法では KPCA により部分空間を生成するのではなく, 各クラスのデータは一部の領域に密集していることを利用して, それらの中で一次独立なデータそのものをそのクラスの部分空間を構成する軸とみなす. これにより, 固有値問題を解く必要がなくなり, 学習コストを大幅に削減することができる. さらに, それぞれの軸に対して SS-SVM と同様にマージン最大化の概念のもと重みを最適化することにより識別の観点から最適な重みを決定し, 汎化能力を高めることができる. ただし, 固有値問題を解かないため, 累積寄与率を基に特徴選択を行えない. そのため, LP-SVM による特徴選択と 4 章で提案した重みに基づく特徴選択を適用した.

計算機実験により, 標本特徴空間を縮小することによる高速化手法は前章までに用いていた手法と比べて学習コストを大幅に削減することができた. また ES-SVM では SS-SVM と比べて汎化能力を劣化させることなく学習コストが非常に削減できていることを確認でき, 非常に有効な手法であるといえる.

第6章

多重非線形部分空間法の導入

6.1 緒言

部分空間法に基づくサポートベクトルマシン (SS-SVM: Subspace-Based Support Vector Machine)^{23, 24)} では, 各クラスの部分空間を構成する軸に対して様々なモデルのサポートベクトルマシン (SVM: Support Vector Machine)^{5, 6, 7)} を適用して識別の観点から最適な重み付けを行う. これにより, カーネル部分空間法 (KSM: Kernel Subspace Method)^{13, 14)} の汎化能力を向上させた. しかしながら, KSM と SS-SVM はともに用いるカーネル¹²⁾ の種類とカーネルパラメータを五分割交差検定法により決定する必要がある. ただし, 現在までの実験では用いているカーネルはラジアル基底関数 (RBF: Radial Basis Function) カーネルのみである. そして, カーネル主成分分析 (KPCA: Kernel Principal Component Analysis)^{10, 11)} を用いて, 各クラス部分空間を生成し, 部分空間を構成する軸に対する類似度を設定する. このとき, 各軸に対して SVM により重み付けを行う. しかしながら, カーネルパラメータごとの部分空間はそれぞれ異なり, 五分割交差検定法により選択されなかったカーネルパラメータにおける部分空間は学習に全く考慮されない. また, 五分割交差検定法では前述のようにあらかじめ選択した選ばれるパラメータ値の候補をそれぞれ用いた学習を複数回する必要がある. 非負制約を設けた部分空間法に基づく最小自乗サポートベクトルマシン (SSLS-SVM: Subspace-Based Least Squares Support Vector Machine)^{25, 26)} を例とすると, 選択する必要のあるパラメータはマージンパラメータ C とカーネルパラメータ γ である. このとき, それぞれ 10 個の候補があるとする合計 $10 \times 10 = 100$ 通りの組み合わせの学習を 1 ファイルにつき 5 回行わなければならない.

そこで本章では, これらの問題点を解決するために複数のカーネルパラメータすべて考慮した多重非線形部分空間法に基づく SVM (MNS-SVM: Multiple Nonlinear Subspace-Based Support Vector Machine) を提案する. すなわち, あらかじめ設定しておいた複数のカーネルパラメータの候補を用いて, それぞれ KPCA によりクラスごとに部分空間を生成する. このとき, それらの軸に対しては重み付けを行わない. 重み付けを行わないとは, すべての軸の重みを等しく設定することである. これは前章までの計算機実験により固有値を重みとして設定する場合より高い汎化能力を示したためである. そして, それぞれのカーネルパラメータにおけるクラスごとの部分空間への

射影長を求める。これにより、ひとつの教師ベクトルに対してクラスごとにカーネルパラメータ数の射影長を得られる。これらの射影長をその部分空間の属するクラスにおける類似度を構成するための特徴量のひとつとして考え、それらの特徴量を用いて SS-SVM と同様に重みの最適化処理を行う。したがって、各クラス部分空間の次元数はカーネルパラメータ数となる。これにより、すべてのカーネルパラメータを用いたときの部分空間を識別に考慮することができる。さらにカーネルパラメータを候補から五分割交差検定法により決定する必要がなく、前述の例ではマージンパラメータ数と同じ 10 通りの組み合わせの学習を 1 ファイルにつき 5 回行えばよいのみである。これにより五分割交差検定法における学習コストが大幅に削減できる。ただし、識別に悪影響を及ぼす部分空間が存在する恐れがあり、各部分空間ごとの固有値を求めることができないため、累積寄与率を用いた特徴選択は行えない。この問題を回避するため、最小自乗サポートベクトルマシン (LS-SVM: Least Squares Support Vector Machine)^{7, 21)} を用いる多重非線形部分空間法に基づく LS-SVM (MNSLS-SVM: Multiple Nonlinear Subspace-Based Least Squares Support Vector Machine) では 4 章の重みに基づく特徴選択手法^{25, 26)} を用いる。

6.2 多重非線形部分空間法に基づくサポートベクトルマシン

KSM または SS-SVM では、カーネルパラメータをひとつ選択し、各クラスの部分空間はひとつずつ生成される。しかしながら、他のカーネルパラメータにおいても識別において重要な部分空間が生成される可能性がある。そこで複数のカーネルパラメータを用いて複数の部分空間をクラスごとにそれぞれ生成し、それらすべてを用いて識別処理を行う。この学習法を本論文では多重非線形部分空間法 (MNSM: Multiple Nonlinear Subspace Method) とよぶ。図 6.1 に、9 つの RBF カーネルパラメータ $\gamma = \{1, \dots, 9\}$ を用意し、それらのカーネルパラメータをすべて用いた MNS の学習法を示す。まずクラス 1 からクラス n においてそれぞれ 9 つの部分空間をそれぞれのカーネルパラメータを用いて KPCA により生成する。このとき、それぞれの部分空間を構成する軸の重みはすべて等しく設定する。図のようにある入力进行分类するためにまず、それぞれの部分空間に写像し射影長を求める。そして、その射影長に対してそれぞれ重みを付ける。図中の w_{ij} ($i = 1, \dots, n, j = 1, \dots, 9$) はクラス i における j 番目のカーネルパラメータを用いたときの射影長に対する重みを意味しており、 S_i はクラス i における類似度を示している。このとき、重み付けされた総和がそれぞれのクラスにおける類似度となる。これにより、各パラメータごとの部分空間をすべて考慮することができ、さらにあらかじめカーネルパラメータをひとつ選択する必要がなくなる。

それぞれの部分空間への射影長に対する重み付けにより汎化能力が大きく左右され

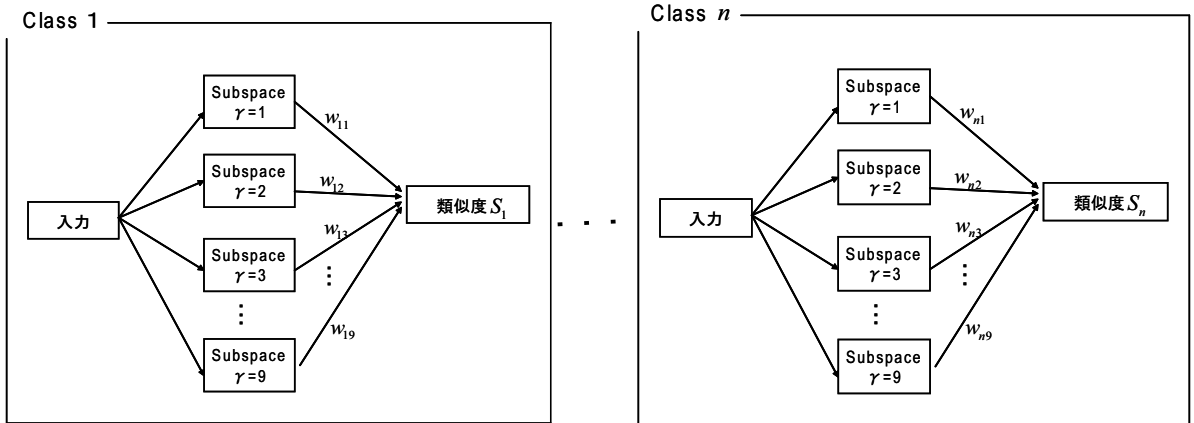


図 6.1 多重非線形部分空間法

るため、これらの重み付けは非常に重要となる。そこで前章までの SS-SVM と同様にクラス間のマージンを最大化するように重みの最適化を行う SVM を取り入れる。このとき、データセットに関わらず各クラスにおける特徴量の数はカーネルパラメータ数と等しくなる。

m 次元入力ベクトル \mathbf{x} の n クラス問題において、はじめに用いる K 個のカーネルパラメータを設定し、 k 番目のカーネルパラメータを用いたときの写像関数 $g_k(\mathbf{x})$ により l 次元特徴空間に \mathbf{x} を写像する。このとき、 M 個の教師ベクトル $\mathbf{x}_j (j = 1, \dots, M)$ を用いてクラス i 部分空間を構成する r_{ik} 本の軸 $\varphi_{ikz} (z = 1, \dots, r_{ik})$ は固有値問題の固有ベクトルとなる。しかしながら、高次元特徴空間は無限次元となりうるため計算が困難になる。そこで、計算速度を向上させるため、 k 番目のカーネルパラメータを用いたときのクラス i の N_{ik} 個の一次独立なデータ $\mathbf{x}_{ikj} (j = 1, \dots, N_{ik})$ によって構成できる N_i 次元標本特徴空間^{32, 33)} に写像する。カーネルを $H_k(\mathbf{x}, \mathbf{x}') = g_k^T(\mathbf{x})g_k(\mathbf{x}')$ とすると、標本特徴空間への写像関数 $\mathbf{h}_k(\mathbf{x})$ は次式のように定義できる。

$$\mathbf{h}_{ik}(\mathbf{x}) = \left(H_k(\mathbf{x}_{ik1}, \mathbf{x}), \dots, H_k(\mathbf{x}_{ikN_{ik}}, \mathbf{x}) \right)^T \quad (6.1)$$

このとき、一次独立なデータの選択にはコレスキー分解を用いる。解くべき固有値問題は以下ようになる。

$$\sum_{j \in X_i} \mathbf{h}_{ik}(\mathbf{x}_j)\mathbf{h}_{ik}(\mathbf{x}_j^T)\varphi_{ikz} = \lambda_{ikz}\varphi_{ikz} \quad (6.2)$$

ここで、 $\varphi_{ikz}, \lambda_{ikz}$ はそれぞれ k 番目のカーネルパラメータを用いたときのクラス i における z 番目の固有ベクトル、固有値を表し、 $\lambda_{ik1} > \dots > \lambda_{ikr_{ik}}$ となるように並べ替える。

また, k 番目のカーネルパラメータを用いたときのクラス i 部分空間の次元数 r_{ik} は寄与率の閾値 κ を設定し, 次式のような累積寄与率 $a(r_{ik})$ を用いることにより決定する.

$$a(r_{ik}) = \frac{\sum_{j=1}^{r_{ik}} \lambda_{ikj}}{\sum_{j=1}^{N_{ik}} \lambda_{ikj}} \quad (6.3)$$

ここで, $a(r_{ik} - 1) < \kappa \leq a(r_{ik})$ となるような r_{ik} 個の部分空間の軸を決定する. この処理をすべてのカーネルパラメータを用いて行う.

クラス i 類似度 $S_i(\mathbf{x})$ はクラス i 部分空間への射影長と写像前の空間 (ここでは, 高次元特徴空間) における入力ベクトルの長さの比に重み付けした値で定義でき, 以下のように表せる.

$$S_i(\mathbf{x}) = \sum_{k=1}^K w_{ik} \sum_{j=0}^{r_{ik}} \frac{(\boldsymbol{\varphi}_{ikj}^T \mathbf{h}_{ik}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{ikj}\|^2 \|\mathbf{g}_k(\mathbf{x})\|^2} \quad (6.4)$$

ここで, w_{ik} は k 番目のカーネルパラメータを用いたときのクラス i 部分空間への射影長に対する重みである. w_{ik} を SVM により最適化することにより識別性能を向上させる. SVM の最適化問題に適用するために次式のように $\mathbf{f}_i(\mathbf{x})$ を定義する.

$$\mathbf{f}_i(\mathbf{x}) = \left(\sum_{j=0}^{r_{i1}} \frac{(\boldsymbol{\varphi}_{i1j}^T \mathbf{h}_{i1}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{i1j}\|^2 \|\mathbf{g}_1(\mathbf{x})\|^2}, \dots, \sum_{j=0}^{r_{iK}} \frac{(\boldsymbol{\varphi}_{iKj}^T \mathbf{h}_{iK}(\mathbf{x}))^2}{\|\boldsymbol{\varphi}_{iKj}\|^2 \|\mathbf{g}_K(\mathbf{x})\|^2} \right)^T \quad (6.5)$$

これにより, 式 (6.4) は次式のように書き換えることができる.

$$S_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{f}_i(\mathbf{x}) \quad (6.6)$$

ここで, $\mathbf{w}_i = (w_{i1}, \dots, w_{iK})^T$ である. この形式は SVM における決定関数と非常に類似しており, SVM の最適化問題が容易に適用できる. ただし, すべてのカーネルパラメータにおける部分空間が識別において有用であるとは必ずしもいえなく, 中には冗長なカーネルパラメータも含まれている可能性がある. そこで, 前章までのさまざまな特徴選択の中で汎化能力が最も高くなる重みの絶対値に基づく特徴選択手法を MNS-SVM に適用する. このため, 適用する SVM のモデルは LS-SVM であり, MNSLS-SVM とよぶ.

以下に MNSLS-SVM における $\mathbf{f}_i(\mathbf{x})$ 決定までの学習アルゴリズムを説明する.

学習アルゴリズム 11

Step 1 用いる K 個のカーネルパラメータを $\gamma = \{\gamma_1, \dots, \gamma_K\}$ とする. $\kappa = 99(\%)$, $k = 1$ とおく.

Step 2 γ_k を用いてコレスキー分解を用いて全教師データから一次独立なデータを選択する.

- Step 3** Step 2 で求めた一次独立なデータを用いて式 (6.1) を計算し, クラス i ($i = 1, \dots, n$) における標本特徴空間への写像関数 $\mathbf{h}_{ik}(\mathbf{x})$ を生成する.
- Step 4** 式 (6.2) よりクラス i ($i = 1, \dots, n$) の $\varphi_{ik_z}, \lambda_{ik_z}$ ($z = 1, \dots, N_{ik}$) を求める.
- Step 5** κ の値を用いて, クラス i 部分空間の次元数 r_{ik} を決定する. $k = K$ ならば, Step 7 へ進む. そうでなければ, $k = k + 1$ として Step 2 に戻る.
- Step 6** それぞれのクラスにおける部分空間への全教師データ \mathbf{x}_j ($j = 1, \dots, M$) の射影長を用いて式 (6.5) を計算する.

6.3 計算機実験

計算機実験により, 2 クラスベンチマークデータセット^{29, 30)}を用いて MNSLS-SVM の汎化能力を KSM と比較し, 向上しているかを確認する. また, SS-SVM の中で最も汎化能力の高いモデル SSSL-SVM (Ab) と比較し, 汎化能力の評価を行う. また, 五分割交差検定法による学習時間を SSSL-SVM (Ab) と比較することにより, モデル選択における学習コストが削減できているかを確認する.

6.3.1 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 γ, C , 特徴選択における重みの閾値 T を 4 章の計算機実験に用いていたパラメータの選択候補から最適な値を五分割交差検定を用いて決定する. 表 6.1 に重みによる特徴選択を行う SSSL-SVM (Ab) と MNS-SVM のパラメータを示す.

6.3.2 汎化能力の評価

表 6.2 に通常の部分空間法 (KSM) と SSSL-SVM (Ab) と MNSLS-SVM の平均認識率と標準偏差を示す. ここで KSM はすべての軸における重みをすべて等しく設定している. 各データセットにおける最大平均認識率をそれぞれ太字で示す. また KSM の平均認識率に比べ高い MNSLS-SVM の平均認識率の前に (*) を付ける.

表 6.2 より, 5 個のデータセットにおいて KSM と比べて MNSLS-SVM の平均認識率が高く, 1 個のデータセットにおいて同じ平均認識率を示している. 特に Ringnorm データセットにおいては KSM と比べて非常に平均認識率が向上しており, SSSL-SVM (Ab) と比べても高い平均認識率を示している. しかしながら, 複数のデータセットにおいて平均認識率の劣化がみられる. これらのようなデータセットにおいて, 教師データにおける平均認識率がほぼ 100 (%) となっており, 過学習していると考えられる. これは, KSM における最適なカーネルパラメータは 15 以下と比較的小さい値を

表 6.1 MNSLS-SVM のパラメータ値

Data	SSLS-SVM (Ab)			MNSLS-SVM	
	γ	C	T	C	T
Banana	1	100	10^{-6}	10^{-4}	10^{-5}
B. cancer	1.5	10^{-4}	10^{-5}	10^{-4}	10^{-5}
Diabetes	0.05	0.1	10^{-5}	10^{-2}	10^{-5}
German	3	0.1	10^{-6}	10^{-3}	10^{-6}
Heart	0.01	0.01	10^{-5}	10^{-3}	10^{-5}
Image	1.5	100	10^{-6}	1000	10^{-6}
Ringnorm	0.5	1	10^{-5}	10^{-4}	10^{-5}
F. solar	0.1	1	10^{-5}	10^{-2}	10^{-5}
Splice	1	1	10^{-6}	5×10^{-4}	10^{-6}
Thyroid	0.5	50	10^{-5}	10^{-2}	10^{-5}
Titanic	1.5	10^{-4}	10^{-6}	10^{-2}	10^{-5}
Twonorm	1	0.01	10^{-5}	10^{-2}	10^{-6}
Waveform	1	0.1	10^{-5}	10^{-2}	10^{-5}

選択されているが、MNSLS-SVM ではカーネルパラメータの値が大きいほど、その値における部分空間に対する重みの値が大きくなるためであると考えられる。確認のため、Diabetes においてはカーネルパラメータの最大値を 15 としたところ、平均認識率が 75.4 (%) と大幅に向上していた。これにより、重みに基づく特徴選択法では冗長なパラメータの削除はあまりできないことが確認できる。

6.3.3 学習時間の評価

表 6.3 に SSLS-SVM (Ab) と MNSLS-SVM における各パラメータ選択のための五分割交差検定法に要する学習時間 (s) を示す。

表 6.3 より MNSLS-SVM における五分割交差検定法における学習時間がすべての SSLS-SVM (Ab) に比べて非常に短くなっている。特に Image データセットと Splice データセットのような大規模データセットにおいては大幅な学習時間の削減が行えている。この理由を以下に述べる。SSLS-SVM (Ab) と MNSLS-SVM のモデル選択において、部分空間生成のための固有値問題を解く回数は等しくなり、重みの最適化問題を解く回数が SSLS-SVM (Ab) と比べて、 $\frac{1}{5}$ となっている。このため、モデル選択における学習コストの差は重みの最適化問題における学習コストの差となる。よって、大規模データセットなどの重みの最適化問題における学習コストが小さいデータセットに対しては他のデータセットと比べて、モデル選択における学習コストの削減率は大きくなる。

本実験により、モデル選択の高速化が可能となることを確認できた。

表 6.2 テストデータにおける平均認識率 (%) と標準偏差の比較 (MNSLS-SVM)

Data	KSM	SSLS-SVM (Ab)	MNSLS-SVM
Banana	88.6 ± 0.6	89.5 ±0.4	*88.9±0.6
B. cancer	75.0 ± 4.3	75.0 ±4.1	65.0±5.0
Diabetes	73.5 ± 1.8	75.1 ±2.2	*73.7±2.1
German	75.1 ± 2.2	74.4±2.1	72.7±3.1
Heart	80.5 ± 3.3	83.9 ±3.2	*83.4±3.4
Image	96.3 ± 0.6	97.0 ±0.6	92.0±1.1
Ringnorm	76.6 ± 11.2	98.2±0.2	* 98.5 ±0.1
F. solar	65.1 ± 1.8	66.5 ±1.6	*65.2±1.7
Splice	87.6 ± 0.8	89.0 ±0.5	84.8±1.3
Thyroid	95.6 ± 2.1	96.2 ±1.9	95.0±2.4
Titanic	76.6 ± 1.2	77.3 ±0.3	76.0±2.1
Twonorm	97.6 ± 0.1	97.7 ±0.1	97.6±0.1
Waveform	88.5 ± 0.6	90.0 ±0.5	87.0±1.3

表 6.3 五分割交差検定法の学習時間 (s) の比較 (SSLS-SVM (Ab) と MNSLS-SVM)

Data	SSLS-SVM (Ab)	MNSLS-SVM
Banana	1730	654
B. cancer	380	176
Diabetes	4524	1314
German	28916	5531
Heart	298	136
Image	175742	19536
Ringnorm	2807	875
F. solar	2244	2161
Splice	93337	10188
Thyroid	146	67
Titanic	80	51
Twonorm	613	324
Waveform	3246	1137

6.4 結言

本章では、複数のカーネルパラメータを同時に考慮しながら学習を行う新しい SS-SVM のモデルである MNS-SVM を提案した。この手法ではあらかじめカーネルパラメータをひとつ選ぶのではなく、すべてのカーネルパラメータを選択する。これにより、全パラメータにおける部分空間を考慮でき、通常の KSM と比べ汎化能力が向上する。さらに五分割交差検定法によりカーネルパラメータを選択する必要がなくなることから前章までの SS-SVM と比べてモデル選択における学習の高速化が可能となる。

計算法実験により、MNS-SVM における汎化能力の評価・五分割交差検定法における学習コストの評価を行った。通常の KSM と比べて多くのデータセットにおいて汎化能力は向上していることが確認できた。しかしながら、いくつかのデータセットにおいて汎化能力が極めて劣化した。これは、それらのデータセットにおいて過学習しているためと考えられる。また、SS-SVM の中で優れたモデル SLS-SVM (Ab) と比べると汎化能力が劣っていた。これは、MNS-SVM ではそれぞれの部分空間全体にひとつの重みをつけるものであり、各部分空間を構成する各軸に対して重み付けをしていないことが原因であると考えられる。また、五分割交差検定法におけるモデル選択において学習コストが大幅に削減できており大規模データセットに対しては MNS-SVM は有効であることを確認できた。今後、MNS-SVM においてあらかじめ選択するカーネルパラメータの候補の決定法と各軸に対する重み付けをどのようにするかを考える必要がある。

第7章

結論

本論文では、カーネル部分空間法 (KSM: Kernel Subspace Method) の利点である解くべき空間が低次元かつ識別部が解析可能である点とサポートベクトルマシン (SVM: Support Vector Machine) の利点である汎化能力が高い点をすべて備えた識別手法の開発を目的とした部分空間法に基づく SVM (SS-SVM: Subspace-Based Support Vector Machine) に関する研究について述べ、現在までに提案した SS-SVM のモデルとそれらの有効性について議論した。

1章では、研究背景と研究目的と研究における実験環境について述べた。2章では、本研究において従来手法とする部分空間法と SVM の理論とその概略、またそれらに用いられるカーネル法や多クラス問題への拡張方式について述べた。

3章では、KSM の汎化能力を向上させるために部分空間を構成する軸における重みを SVM のマージン最大化の概念を取り入れて決定する SS-SVM を提案した。SS-SVM では KSM とは異なり他のクラスにおける部分空間との分離を考慮して識別ルールである類似度を決定できる。SS-SVM のモデルとして連立一次方程式を解くことにより最適解を求める部分空間法に基づく最小自乗サポートベクトルマシン (SSL-SVM: Subspace-Based Least Squares Support Vector Machine)、線形計画問題を解くことにより最適解を求めることのできる、さらに識別と同時に特徴選択を行える部分空間法に基づく線形計画サポートベクトルマシン (SSLP-SVM: Subspace-Based Linear Programming Support Vector Machine)、アクティブ集合を用いた高速学習法を容易に適用できる部分空間法に基づく L2 サポートベクトルマシン (SSL2-SVM: Subspace-Based L2 Support Vector Machine) の3つのモデルを用いた。計算機実験では、それぞれのモデルごとに汎化能力を KSM と等しいパラメータを設定し、比較・評価を行うことにより汎化能力が向上することを示したが、一部のデータセットにおいては、汎化能力が KSM と比べ劣っていた。そこで、それぞれの SS-SVM において、KSM のパラメータに関係なく独立して五分割交差検定法により、それぞれのパラメータを決定し、それらのパラメータを用いて KSM と比較したところ、KSM に比べ、大半のデータセットにおいて高い汎化能力を示した。このことから、KSM における最適なパラメータは必ずしも SS-SVM においても最適なパラメータではないことが確認できた。また、同時に五分割交差検定法におけるバリデーションセットに対しての識別性能とテストデータセットにおける識別性能を比較することにより、大半のデータセットにおいて

ほぼ同等な性能を示し、五分割交差検定法によって汎化能力の推測が可能であることを確認できた。また、それぞれの SS-SVM において、同時定式化方式と一対他方式との学習時間を比較・評価したが、SSLP-SVM を除いては大半のデータセットにおいてそれほど学習時間の差がなく、SS-SVM における学習コストの大半は部分空間生成における学習コストではないかと考えられる。また、SSLP-SVM における各クラスにおける部分空間の次元数を確認することにより、識別と同時に特徴選択が行えていることを確認できた。また、これらの SS-SVM のモデルと SVM の汎化能力の比較・評価を行い、2 クラス問題において SS-SVM の中で SSL2-SVM が最も汎化能力が高いが、SSLS-SVM もほぼ同等であることを示し、また通常の SVM との比較ではすべてのモデルにおいて汎化能力が少し劣ることが確認できた。また、最後に SVM に対してのメリットである識別部の解析を SSLS-SVM を用いて Thyroid データセットで行った。

4 章では、SS-SVM における負の重みをもつ軸の重要性の確認と重みの絶対値に基づく特徴選択手法を提案した。SS-SVM は通常の KSM とは異なり、負の重みをもつ可能性があり、その負の重みが識別にどのような影響が確認するために非負制約を設けた SS-SVM と比較・評価した。また、SS-SVM では識別の観点から最適な重み付けを決定することができ、それらの重みは識別における対応する軸の重要度を意味している。すなわち、重みの絶対値が小さい軸は識別に与える影響は小さく、識別の観点から冗長な軸であるといえ、これらの軸を削減することにより汎化能力を高めることができる。ただし、SSLP-SVM では識別と同時に特徴選択が行え、また SSL2-SVM では、繰り返し計算が必要となり、学習コストが次元数に依存せずにサポートベクトル数に影響するため、学習コストが膨大となりうることから、この特徴選択手法は SSLS-SVM にのみ適用した。このときの SSLS-SVM を SSLS-SVM (Ab) とよぶ。計算機実験により非負制約を設けた SS-SVM と負を許す SS-SVM を比較することにより、負を許す SS-SVM の汎化能力が優れており、負の重みをもつ軸は識別において重要となりうることを確認できた。また、SSLS-SVM (Ab) は、他の SSLS-SVM に比べて汎化能力が非常に高く、さらに通常の SVM と比べて同等以上の性能を示し、この特徴選択手法の有効性を示すことができた。また、この特徴選択手法によって特徴量が削減できていることも確認し、さらに各軸における重みと固有値の解析により、これらの値には関連性がなく、通常の KSM または 3 章までの SS-SVM における特徴選択手法により識別の観点から重要な軸が削減されていたことが確認できた。これらの結果により、SSLS-SVM (Ab) は本研究の目的であった情報量を大幅に削減でき、かつ識別部が解析可能で、さらに SVM と同等以上の汎化能力の非常に高い識別手法といえる。

5 章では、SS-SVM における学習コストを削減するため、標本特徴空間生成部の改良と標本部分空間法に基づくサポートベクトルマシン (ES-SVM: Empirical Subspace-Based Support Vector Machine) を提案した。標本特徴空間生成部の改良では、全教

師データを用いて標本特徴空間を生成するのではなく、クラスごとに生成することで次元数を削減し、解くべき固有値問題のサイズを小さくすることにより学習の高速化を行う。また、ES-SVMでは、クラスごとの教師データが密集する傾向があることを利用し、それらの教師データそのものを部分空間を構成する軸とし、部分空間生成における固有値問題を解く必要をなくすことにより、学習の高速化を行う。ただし、この手法では、各軸に対応する固有値が得られないことから、識別と同時に特徴選択を行うか、重みを基準に特徴選択する必要がある、このことから適用するSVMのモデルを最小自乗サポートベクトルマシン (LS-SVM: Least Squares Support Vector Machine) と線形計画サポートベクトルマシン (LP-SVM: Linear Programming Support Vector Machine) のみとした。計算機実験により、通常のSS-SVMと標本特徴空間生成部の改良を行ったSS-SVMとの学習時間の比較を行うことにより、学習の高速化を確認でき、有効性を示した。また、ES-SVMと4章で用いたSS-SVMの学習時間の比較では大幅に学習時間が削減されており、さらに汎化能力の比較により、この手法による汎化能力の劣化はほとんど見られないことからES-SVMの有効性を示した。このことから、標本部分空間法に基づくLS-SVM (ESLS-SVM: Empirical Subspace-Based Least Squares Support Vector Machine) はSSLS-SVM (Ab) と同様にSVMと同等以上の高い汎化能力備えており、非常に優れた識別器であるといえる。

6章では、複数のパラメータにおける部分空間を同時に考慮し、さらにモデル選択における学習コストを大幅に削減することのできる多重非線形部分空間法に基づくSVM (MNS-SVM: Multiple Nonlinear Subspace-Based Support Vector Machine) を提案した。この手法では、あるひとつのカーネルパラメータを部分空間生成のために用いるのではなく、複数用いることによりクラスごとに複数の部分空間を定義し、それらへの入力の射影長をそれぞれ、それらのクラスにおけるその入力のひとつの特徴量として扱い、SVMによりそれぞれの特徴量に対して識別の観点から重み付けすることにより、識別処理を行う。これによりカーネルパラメータを五分割交差検定法により決定する必要がないため学習コストを大幅に削減できる。計算機実験により、通常のKSMと比較し、KSMと比べて汎化能力が向上していることを確認したが、SS-SVMの最も汎化能力の高いモデルであるSSLS-SVM (Ab) と比較すると、大半のデータセットにおいてMNSLS-SVMの汎化能力が劣っていた。この原因として、各部分空間を構成する軸に対してすべて等しい値の重みを設定していることが考えられる。しかしながら、モデル選択における学習時間をSSLS-SVM (Ab) と比較したところ、大幅な学習コストの削減が確認でき、特に大規模データセットにおいては非常に有効であることを示した。

このように、本研究においてKSMの利点とSVMの利点をそれぞれ備えた識別手法の開発を行えた。今後の課題として、モデル選択における学習コストは膨大となりやす

く、特に最もよい性能を示した SSLS-SVM (Ab) においては、決定すべきパラメータが3つ存在することからこれらの求めるパラメータ数を削減するなどし、汎化能力を落とすことなく学習の高速化を行うことが今後の課題のひとつとして挙げられる。また今後、現在まで用いていたデータセットだけでなく実データに対しても有効であるかどうかの検証が課題となる。

謝 辞

本研究の機会を与えて頂き、また学部4年生から4年間に渡り、終始熱心かつ適切な御指導を賜りました神戸大学大学院工学研究科電気電子工学専攻 阿部重夫教授に謹んで感謝の意を表します。

本研究をまとめるにあたって、多くの貴重な御教示ならびに御助言を賜りました神戸大学大学院工学研究科電気電子工学専攻 増田澄男教授、八坂保能教授に深く感謝致します。

本研究を始めるきっかけと部分空間法について懇切な御指導および御教授を賜りました筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻 福井和広准教授に深く感謝致します。

また、本研究を進めるにあたり、適切な御指導と御助言を賜りました神戸大学大学院工学研究科電気電子工学専攻 小澤誠一准教授に深く感謝致します。日頃、様々な方面で御世話になりました神戸大学大学院工学研究科電気電子工学専攻知的学習論講座の諸氏に心より感謝致します。特に、竹内翔吾氏、田尻康之氏、藪脇諒佑氏、福田創氏には関連研究の勉強や計算機実験の補助や研究環境の整備など多大な御協力賜りましたことを感謝致します。また、博士課程後期課程に進学することを納得し、私生活において様々な面倒を見てくださった家族に深く感謝致します。

参 考 文 献

- 1) C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, 2006.
- 2) K. Edamatsu and Y. Nitta, “Automated Capsule Inspection Method,” *Pattern Recognition*, Vol. 14, Nos. 1–6, pp. 365–374, 1981.
- 3) S. Watanabe and N. Pakvasa, “Subspace Methods of Pattern Recognition,” *Proc. 1st International Joint Conference on Pattern Recognition*, pp. 283–328, 1973.
- 4) E. Oja, *Subspace Methods of Pattern Recognition*, Research Studies Press, Letchworth, UK, 1983.
- 5) V.N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, NY, 1998.
- 6) V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, 1995.
- 7) S. Abe, *Support Vector Machines for Pattern Classification (Advances in Pattern Recognition)*, Springer-Verlag, London, UK, 2010.
- 8) S. Haykin, *Neural Networks: A Comprehensive Foundation, second edition*, Prentice Hall, Upper Saddle River, NJ, 1999.
- 9) T.R. Reed and J.M.H. Du Buf, “A Review of Recent Texture Segmentation and Feature Extraction Techniques,” *CVGIP: Image Understanding*, Vol. 57, No. 3, pp. 359–372, 1993.
- 10) B. Schölkopf, S. Mika, C.J.C. Burges, P. Knirsch, K.R. Müller, C. Ratsch, K. Tsuda, and A.J. Smola, “Input Space Versus Feature Space in Kernel-Based Methods,” *IEEE Trans. Neural Networks*, Vol. 10, No. 5, pp. 1000–1016, 1999.
- 11) B. Schölkopf, A.J. Smola, and K.R. Müller, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem,” *Neural Computation*, Vol. 10, No. 3, pp. 1299–1319, 1998.
- 12) T. Gärtner, “A Survey of Kernels for Structured Data,” *ACM SIGKDD Explorations Newsletter*, Vol. 5, No. 1, pp. 49–58, 2003.
- 13) K. Tsuda, “Subspace Classifier in the Hilbert Space,” *Pattern Recognition Letters*, Vol. 20, No. 5, pp. 513–519, 1999.

- 14) E. Maeda and H. Murase, "Kernel-Based Nonlinear Subspace Method for Pattern Recognition," *Systems and Computers in Japan*, Vol. 33, No. 1, pp. 38–52, 2002.
- 15) H. Sakano, N. Mukawa, and T. Nakamura, "Kernel Mutual Subspace Method and Its Application for Object Recognition," *Electronics and Communication in Japan*, Part 2, Vol. 88, No. 6, pp. 45–53, 2005.
- 16) K. Fukui and O. Yamaguchi, "The Kernel Orthogonal Mutual Subspace Method and Its Application to 3D Object Recognition," *Proceedings of Asian Conference on Computer Vision*, pp. 467–476, 2007.
- 17) K. Fukui, B. Stenger, and O. Yamaguchi, "A Framework for 3D Object Recognition Using the Kernel Constrained Mutual Subspace Method," *Proceedings of Asian Conference on Computer Vision*, pp. 315–324, 2006.
- 18) 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.
- 19) S. Abe, "Is Primal Better than Dual," *Proceedings of International Conference on Artificial Neural Networks (ICANN2009)*, pp. 854–863, 2009.
- 20) R. Yabuwaki and S. Abe, "Convergence Improvement of Active Set Support Vector Training," *Proceedings of International Joint Conference of Neural Networks (IJCNN2010)*, pp. 1426–1430, 2010.
- 21) J.A.K. Suykens, "Least Squares Support Vector Machines for Classification and Nonlinear Modelling," *Neural Network World*, Vol. 10, Nos. 1–2, pp. 29–47, 2000.
- 22) L. Zhang and L. Jiao, "Linear Programming Support Vector Machines," *Pattern Recognition*, Vol. 35, No. 12, pp. 2927–2936, 2000.
- 23) T. Kitamura, S. Takeuchi, S. Abe, and K. Fukui, "Subspace-Based Support Vector Machines for Pattern Classification," *Neural Networks*, Vol. 22, Nos. 5–6, pp. 558–567, 2009.
- 24) T. Kitamura and S. Abe, "Subspace-Based L2 Support Vector Machines," *Australian Journal of Information Intelligent Processing Systems*, Vol. 12, No. 3, pp. 38–43, 2010.
- 25) T. Kitamura, S. Takeuchi, and S. Abe, "Feature Selection and Fast Training of Subspace Based Support Vector Machines," *Proceedings of International Joint Conference on Neural Networks (IJCNN2010)*, pp. 1967–1972, 2010.
- 26) 北村拓也, 竹内翔吾, 阿部重夫, "標本部分空間法に基づくサポートベクトルマシン

- ンの開発,” 情報処理学会論文誌 数理モデル化と応用, Vol. 3, No. 3, pp. 12–23, 2010.
- 27) L. Breiman, “Arcing Classifiers,” *The Annals of Statistics*, Vol. 26, No. 3, pp. 801–849, 1998.
- 28) I. Guyon, J. Weston, S. Barnhill, and V.N. Vapnik, “Gene Selection for Cancer Classification Using Support Vector Machines,” *Machine Learning*, Vol. 46, No. 1–3, pp. 389–422, 2002.
- 29) G. Rätsch, T. Onda, and K.R. Müller, “Soft Margins for AdaBoost,” *Machine Learning*, Vol. 42, No. 3, pp. 287–320, 2001.
- 30) <http://ida.first.fraunhofer.de/projects/bench/benchmarks.htm>.
- 31) P.A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, London, UK, 1982.
- 32) S. Abe, “Sparse Least Squares Support Vector Training in the Reduced Empirical Feature space,” *Pattern Analysis and Applications*, Vol. 10, No. 3, pp. 203–214, 2007.
- 33) H. Xiong, M.N.S. Swamy, and M.O. Ahmad, “Optimizing the Kernel in the Empirical Feature Space,” *IEEE Trans. Neural Networks*, Vol. 16, No. 2, pp. 460–474, 2005.

関連発表論文

学術論文

- 1) T. Kitamura, S. Takeuchi, S. Abe, and K. Fukui, “Subspace-Based Support Vector Machines for Pattern Classification,” *Neural Networks*, Vol. 22, No. 5–6, pp. 558–567, 2009.
- 2) 北村拓也, 竹内翔吾, 阿部重夫, “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,” *情報処理学会論文誌 数理モデル化と応用*, Vol. 3, No. 3, pp. 12–23, 2010.
- 3) T. Kitamura and S. Abe, “Subspace-Based L2 Support Vector Machines,” *Australian Journal of Information Intelligent Processing Systems*, Vol. 12, No. 3, pp. 38–43, 2010.

国際会議論文

- 1) T. Kitamura, S. Abe, and K. Fukui, “Subspace Based Least Squares Support Vector Machines for Pattern Classification,” *Proceedings of International Joint Conference on Neural Networks (IJCNN 2009)*, pp. 1640–1646, 2009.
- 2) S. Takeuchi, T. Kitamura, S. Abe, and K. Fukui, “Subspace Based Linear Programming Support Vector Machines for Pattern Classification,” *Proceedings of International Joint Conference on Neural Networks (IJCNN 2009)*, pp. 3067–3073, 2009.
- 3) T. Kitamura, S. Takeuchi, and S. Abe, “Feature Selection and Fast Training of Subspace Based Support Vector Machines,” *Proceedings of International Joint Conference on Neural Networks (IJCNN2010)*, pp. 1967–1972, 2010.
- 4) Y. Tajiri, R. Yabuwaki, T. Kitamura, and S. Abe, “Feature Extraction Using Support Vector Machines,” *Proceedings of 17th International Conference on Neural Information Processing (ICONIP2010)*, pp. 108–115, 2010.

その他論文

- 1) 北村拓也, 阿部重夫, 福井和広, “部分空間法に基づく最小自乗サポートベクトルマシン,” *平成 21 年電気学会電子・情報・システム部門大会講演論文集*, pp. 1164–1169, 2009.
- 2) 竹内翔吾, 北村拓也, 阿部重夫, 福井和広, “部分空間法に基づく線形計画サポートベクトルマシン,” *平成 21 年電気学会電子・情報・システム部門大会講演論文集*,

pp. 1158–1163, 2009.

- 3) 北村拓也, 竹内翔吾, 阿部重夫, “錘制約部分空間法に基づくサポートベクトルマシンの開発,” 第 77 回情報処理学会数理モデル化と問題解決研究会, Vol. 2010–MPS–77, No. 28, pp. 1–6, 2010.
- 4) 北村拓也, 阿部重夫, “部分空間法に基づく最小自乗サポートベクトルマシンの汎化能力の向上方式,” 計測制御学会 第 37 回知能システムシンポジウム, pp. 111–114, 2010.
- 5) 北村拓也, 阿部重夫, “部分空間法に基づく L2 サポートベクトルマシンの開発,” 平成 22 年電気学会電子・情報・システム部門大会講演論文集, pp. 1485–1489, 2010.
- 6) 田尻康之, 藪脇諒佑, 北村拓也, 阿部重夫, “サポートベクトルマシンによる特徴抽出,” 平成 22 年電気学会電子・情報・システム部門大会講演論文集, pp. 1305–1309, 2010.
- 7) 北村拓也, 藪脇諒佑, 田中佑典, 阿部重夫, “多重非線形部分空間法に基づくサポートベクトルマシンの開発,” 計測制御学会 第 38 回知能システムシンポジウム, 2010. (2011 年 3 月発表予定)
- 8) 藪脇諒佑, 北村拓也, 阿部重夫, “関数近似用 サポートベクトルマシンの高速学習方式,” 計測制御学会 第 38 回知能システムシンポジウム, 2010. (2011 年 3 月発表予定)
- 9) 田尻康之, 北村拓也, 藪脇諒佑, 阿部重夫, “サポートベクトルマシンを用いた特徴抽出に基づく識別器の構成法,” 計測制御学会 第 38 回知能システムシンポジウム, 2010. (2011 年 3 月発表予定)
- 10) ペイマイダ, 阿部重夫, 北村拓也, 藪脇諒佑, “L2 サポートベクトルマシンによるファジイ識別器のチューニング法 の開発,” 計測制御学会 第 38 回知能システムシンポジウム, 2010. (2011 年 3 月発表予定)

受賞

- 1) 神戸大学大学院工学研究科 竹水会優秀論文賞, 2010.
(受賞論文) “部分空間法に基づくサポートベクトルマシンの開発” (修士論文)
- 2) 第 7 回 IEEE 関西支部学生研究奨励賞, 2011.
(受賞論文) “Feature Selection and Fast Training of Subspace-Based Support Vector Machines” (IJCNN2010)