改定IP-OLDFによるSVMのアルゴリズム研究

新村　秀一

1970年代から，数理計画法（MP）による判別モデル（またはクラスター分析）の研究が数多く行われてきた．しかし，Stamが指摘するとおり，それらのモデルは統計分野で利用されていない．その一番大きな理由は，汎化能力（統計でいうEC）の検討が弱い点である．SVMは，この閉塞感を打開するものとして期待されている．筆者は，統計的な立場から，SVMはペナルティの客観的な決定法と汎化能力の実証の2点を統計ユーザーに明らかにすべきであると考えている．本稿は，そのうちのペナルティに関して，筆者の研究テーマである改定IP-OLDFと比較して問題点を提起したい．

キーワード：判別分析，誤分類最小化基準，マージン最大化基準，数理計画法，SVM

1. はじめに

幸運の女神は，ある日突然現れる．1976年ごろ，最小誤分類数（Minimum Misclassification Number, MMN）を判別基準とするヒューリスティックな最適線形判別関数（Optimal Linear Discriminant Function, OLDF）を提案した[3]．しかし，多くの統計家にとって過剰推定（オーバーエスティメイト）が考えられ受け入れがたいものであった．SVMでいう汎化能力が悪いと考えられた．しかし，判別する2群がFisherの線形判別関数（LDF）の理論的前提を満たせば，得られた誤分類数はMMNと等しくなる．そこで，何かおもしろい成果が得られることを期待し研究を続け壁に突き当たった．

1998年にこの問題が，計算時間がかかるということでMPの研究者が嫌っている整数計画法（IP）で定式化できることに気づき，IP-OLDFと命名し研究を再開した[6]．とにかく，計算の爆発というブラックホールに分け入り，大変であったが，これまでの判別分析の理論で説明できないことが分かり自己満足していた．

そして，Irisデータとスイス銀行紙幣データという統計で著名なテストデータと，筆者が長年種々の研究に用いてきた医療データの3つの実データを用い，LDF, 2次判別関数，ロジスティック分析，決定木分析等と比較しInternal Check（IC）で目覚ましい成果を得た．また，2変数の正規乱数データで115組の内部標本（教師データの事）と外部標本（評価データの事）を作成して，既存の判別手法と比較しICとExternal Check（EC）でIP-OLDFが他の判別手法に比べて良い結果を得た（[10～14]）．

統計モデルではECでの評価（汎化能力）が重要視される．しかし，最適凸体のどの内点を最終的な判別関数にすべきかが未検討であったため，大規模な外部標本を用いたECをこれまで行ってこなかった．しかし，パターン認識のマージン概念を取り入れた改定IP-OLDFを考えたことで，これが可能になった．

本研究では，これに加えてSVMのペナルティの意味を，改定IP-OLDFと改定LP-OLDFを用いて解明した．

しんむら　しゅういち
成蹊大学 経済学部
〒180-8633 武蔵野市吉祥寺北町3-3-1

2. 取り上げるモデル

2.1 IP-OLDFの定式化

クラス1を表す変数y_iの値を1とし，クラス2を表す変数y_iの値を-1とすれば，IP-OLDFは(1)のように定式化される．すなわち，誤分類されるケースに対して，0/1の整数変数のe_iを1にすることで判別境界を0から-99999として制約条件を緩め，正しく判別されるケースに対して0とすることで判別境界を0に固定すればよい．そして，目的関数で誤分類数($\sum e_i$)を最小化する．

MIN $\sum e_i$
ST
$$y_i*(\mathbf{x}_i'\,\mathbf{b}+1)>=-c*e_i \qquad (1)$$
END
$\mathbf{x}_i=(x_{i1}, x_{i2}, \cdots, x_{ip})$
$y_i=1$ for $x_i(i=1, \cdots, s)\in$クラス1,

$y_i=-1$ for $x_i(i=(s+1),\cdots,(s+t))\in$クラス2

b：p次元判別係数ベクトル

e_i：各x_iに対応した0/1決定変数

c：99999の定数（Big M定数）

このモデルはp次元の判別係数の空間で考えると，MMNの集合は（最適）凸体になり，その頂点を求めることに対応している．頂点に対応した判別関数は，データ空間で考えると判別超平面上にp個（データが一般位置[7]にある場合）あるいは（p+1）個以上（データが一般位置にない場合）のケースがある．これらの判別スコアは0になる（判別超平面上にあるこれらのケースをどう扱うかは，統計でもはっきりしていない）．そこで，次のステップで内点（全てのケースは，どちらかの群に必ず判別されることがIP-OLDFで分った）に移る必要があるが，この点がIP-OLDFでは未検討であった．

2.2 LP-OLDFの定式化

LP-OLDFは，数多くの研究成果のある線形計画法（LP）による判別モデルの一種であり，(1)のe_iを0/1の整数変数から非負の決定変数に変えただけである．LP-OLDFは，誤分類されるケースの判別超平面からの距離の和を最小化している．統計の立場からいえば，この基準が何に役立つのか，あるいは汎化性が良いか否かの検討がLP判別の研究に乏しい事である．

2.3 改定IP-OLDFと改定LP-OLDFの提案

改定IP-OLDFは，パターン認識で古くから考えられているマージン概念を取り入れて次のように定式化する．これによって，最適凸体の内点が1ステップで求まり，これを用いてECが行える．

MIN Σe_i

ST

$$y_i*(\mathbf{x}_i'\ \mathbf{b}+b_0)>=1-c*e_i \qquad (2)$$

END

2.4 SVMの定式化

SVMは，線形分離可能な場合，2つのサポートベクタでマージンをとって完全に分離できる．線形分離できない場合，幾つかのケースがサポートベクタの反対側にある事を認め，その距離e_iの和を最小化することとマージンを最大化することを考えている．(3)の目的関数の2番目の項にあるΣe_iはサポートベクタの反対側にあるケースの距離の和であり，cはペナルティを表す．e_iが正になるものを認めることによって，全てのケースが見かけ上線形分離可能になる．また，最初の項でマージン最大化を行っている．本稿のテーマは，折角最適化手法を用いているのに，恣意的なペナルティのcを用いることの是非についてである．

MIN $\|b\|^2/2+c\Sigma e_i$

ST

$$y_i*(\mathbf{x}_i'\ \mathbf{b}+b_0)>=1-e_i \qquad (3)$$

END

3. MPによる判別モデルの歴史

3.1 MP判別モデルはなぜ利用されないのか

1970年代から，MP判別モデルの研究が数多く行われてきた．しかし，Stam[5]が指摘するとおり，それらのモデルは統計分野で利用されていない．この理由として，筆者は次の点を指摘したい．

従来のMPモデルの多くが，MPで初めて解決できるものが多い．これに対し，MP判別モデルは，すでに統計で確立された判別分析を研究テーマにしているという認識がこの分野の研究者に希薄なことである[17]．統計理論に対し何が優れているのか，判別成績（特に汎化能力）の検討，計算速度や操作性が統計手法に比べ劣らない事，などを検討しなければいけない．これまで，SVM研究と筆者の研究のみが汎化性をとりあげているが，多くのMP判別研究はこの検討を行っていないことが問題である．

3.2 判別理論に貢献しないLP-OLDF

MP判別モデルの主流は，LPを用いたモデルが多い．Glover[4]は，LPモデルの総轄を行っているが，わずか数例の玩具のデータ（tiny data）による説明で，汎化能力の検討まで至っていない．LP-OLDFは，この種のモデルの一種であり，判別分析の理論に何も貢献していないことが分かっている．

3.3 SVMのこれまでのMP判別モデルとの違い

この閉塞感を打開するものとして，SVMが期待されている．SVMは，マージン最大化基準で汎化能力が優れている事を主張している．また，カーネルトリックという美しい理論で多くの研究者をひきつけている．もしこれらが事実であれば，これまでの「判別モデルは説明変数の少ないシンプルなモデルが汎化能力に優れている[1]」とする統計の知識を覆す新しい考え方を提起していることになる．筆者は，統計的な立場から，SVMはこれらの点を統計ユーザーに分かりやすく説明すべきであると考えている．

3.4 IP判別モデル

IPは，最近筆者の使用するWhat'sBest！（やLINDO API）でも100万整数変数のモデルが解ける

ようになるまで，多くの研究者が利用を避けてきた．

IP-OLDFと同じタイプのIP判別モデルには，Liitschwager & Wang[2]がある．しかし，このモデルは判別超平面上に（p+1）個以上の数多くのケースを呼び寄せることが分かっている[16]．しかし，この分野の研究者はこのモデルの欠点を理解していないようだ．

IP-OLDFの特徴は，判別関数の定数項を1に固定している点である．これによって，統計アプローチと異なる新しい事実が発見された．例えば，MMNの集合が判別係数の空間で凸体になること（最適凸体と呼ぶ）．判別モデルに1個説明変数を追加したモデルのMMNは必ず元のモデルのそれより単調減少すること．判別関数の定数項の役割と判別係数の関係，などである[6][8][9]．

3.5 汎化能力の検討

一方問題点として，IP-OLDFは第1ステップで最適凸体の頂点を求める．しかし，次のステップでどの内点を選ぶか明確な指針がなかった．改定IP-OLDFはマージン概念を取り入れる事で，最適凸体の内点が1ステップで求まる．これを用い，スイス銀行1,000フラン紙幣の真札と偽札データを教師データとしてICを，それと同じデータ構造を持つ2万件の乱数データを評価データとしてECを行うことで，以下の著しい成果を得た[18]．

変数選択法は教師データ（スイス銀行データ）で5変数モデルを選ぶ．しかし，乱数データを評価データとして，教師用データで得られたIP-OLDFの判別関数で判別すると，3変数モデルの誤分類数が一番少なかった．すなわち，統計の代表的手法である逐次変数選択法に問題があることや改定IP-OLDFの汎化能力も証明できた．

4. 改定LP-OLDFとSVMの比較

本研究では，SVMのペナルティcの意味を改定LP-OLDFやIP-OLDFと比較することで解明する．表1のデータは，現在早稲田大学教授の高森寛氏が作成した「学生の生活実態調査データ」で，筆者のHPからダウンロードできる（Goggleに新村秀一と入力し検索可能）．あるいは著書[15]に統計ソフトのJMPと一緒にCD-ROMに格納されている．

「座標」は（勉強時間，支出）の値である．「合格」の数字は60点以上の学生番号，「不合格」はそれ未満の学生番号である．「SV」は以下で紹介する判別超平面とサポートベクタに選ばれたことを示す．「$e_i>0$」はサポートベクタの反対側にある誤分類ケースを示す．

表1 データ

座標	合格	不合格	SV	$e_i>0$
1,6		3	—	④
1,8		35	④	
2,4	6		④	①②③④
2,5		18,38	①②③	④
2,6		1	—	④
3,2		8	①	①②③④
3,3	10,23	40	③	①②③④
3,5	34	33	④	①②③④
3,6		28	①	②③④
3,7		17,29	③	④
3,10		9	④	—
4,4	26		—	①②③④
4,5	30		①②③	①②③④
4,6		13	④	①②③④
5,2	24	15	④	①②③④
5,3	39		③	④
5,4	7,37		①	④
5,5		14	—	①②③④
6,3	21,25,27		④	
6,5	5		①②③	④
7,3	2,11,12,16		—	
7,4	32		④	
xが	4,19,20		—	
8以上	22,31,36			
合計	25	15		

表2 改定LP-OLDFとSVMの比較

Cの値	Margin	判別関数 勉強時間	支出	定数項
①0.3以上	2.82	0.500	-0.500	0.500
②0.24	2.88	0.500	-0.480	0.400
③0.1	3.58	0.500	-0.250	-0.750
④0.01	7.07	0.200	-0.200	0.400
⑤0.0074	7.1	0.198	-0.200	0.410
⑥0固定	7.1	0.198	-0.200	0.410
改定LP-OLDF	2.82	0.500	-0.500	0.500
改定IP-OLDF	0	2.000	-33332	166653

4.1 SVM結果

表2と図1は分析結果である．図の＋は合格者群，□は不合格者群，○は後で述べるサポートベクタの回転の中心になる点である．SVMの分析結果は，Cの値を次の6つに分けて示した．

①Cが0.3以上の場合（表2の①の行）

Cの値を0.3以上に設定すると，SVMによる判別関数として$f(x)=0.5*$勉強時間$-0.5*$支出$+0.5$が得られた．マージンは2.82になる．図1の3本の

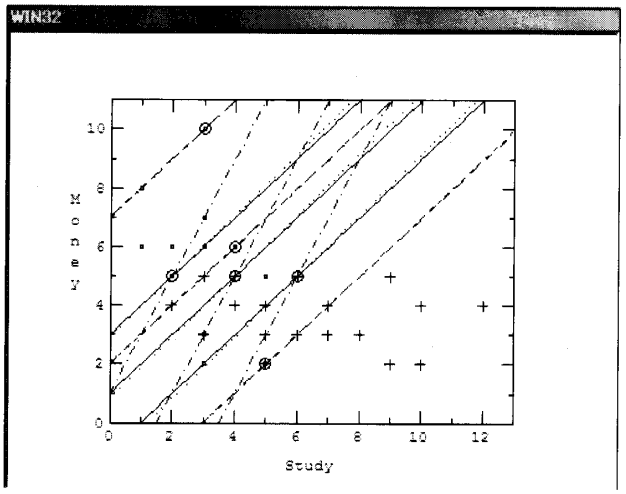

図1 ペナルティの違いによる判別超平面とサポートベクタ

実線の真ん中にある判別超平面がこれに対応し，支出＝勉強時間+1になる．表1のSVで①のついた点(4,5)を含んでいる．

図1の下側の実線のf(x)=1は合格群のサポートベクタであり，点(6,5)と点(5,4)を含む．合格群でこのサポートベクタの反対側にくるものは，表1の「$e_i>0$」で①のついた(2,4)，(3,3)，(3,5)，(4,4)，(4,5)，(5,2)である．このうち判別超平面で不合格群と誤分類された学生は(2,4)と(3,5)である．

①の上側の実線のf(x)=−1は不合格群のサポートベクタであり，点(2,5)と(3,6)を含む．

②Cが0.24の場合（表2の②の行）

Cを0.24に設定すると，判別超平面は支出=1.042＊勉強時間+0.833になる．マージンは2.88と①より大きくなる．図1の実線の少し重なって上にある点線が，判別超平面とサポートベクタを表す．

図1と表1から，合格群のサポートベクタは(6,5)であり，不合格群のサポートベクタは(2,5)である．

合格群の点(5,4)と不合格群の点(3,6)が①のサポートベクタから外れることで，判別超平面は点(3,5)，合格群のサポートベクタは点(6,5)，不合格群のサポートベクタは点(2,5)の各1点で支えられている．これらは支出=5の値をもつ3つの○で表してある．すなわち，cを0.3から次の0.1まで動かしても，その間に他のケースが関係してこないので，判別結果（誤分類されるケース）は変わらないことになる．

③Cが0.1の場合（表2の③の行）

Cの値を0.1に設定すると，判別超平面は支出=2＊勉強時間−3になる．図1の①と②と同じ○で示した3点(3,5)，(6,5)，(2,5)を回転の中心とした一点鎖線が，判別超平面とサポートベクタを表す．また，マージンは3.58と②より大きくなる．

合格群のサポートベクタ(6,5)に新たに(5,3)が加わり，不合格群のサポートベクタ(2,5)に(3,7)が加わり，2点で固定される．

①と③は，複数の学生にサポートされ，固定されている．一方，②のサポートベクタは学生1人でサポートされている．すなわち，Cの値を0.3から0.1まで連続的に減少させると，両方のサポートベクタは，(6,5)と(2,5)の各1点を回転の中心として，①の実線から③の一点鎖線まで連続的に反時計回りに回転する．それに伴い，マージンは2.82から3.58へ連続的に増加する．

④Cが0.01の場合（表2の④の行）

Cの値を0.01に設定すると，判別超平面は支出=勉強時間+2になる．傾き45度の3本の破線が，判別超平面とサポートベクタを表す．マージンは7.07と①②③より大きくなる．

合格群のサポートベクタは(5,2)と(6,3)と(7,4)の3点で，不合格群は(1,8)と(3,10)の2点で固定されている．

⑤Cが0.0074の場合（表2の⑤の行）

次に，Cを例えば0.0074にすると，判別超平面は支出=0.99＊勉強時間+2.05になる．マージンは7.10と①②③④より大きくなる．合格群のサポートベクタは(5,2)の1点で，不合格群は(3,10)の1点でのみ拘束されている．④とほぼ重なるので，図にはこれらを示していない．

⑥Cを0に固定した場合（表2の⑥の行）

C=0にすると目的関数は$\|\mathbf{b}\|^2/2$の最小化になる．サポートベクタの制約が無ければ，判別係数は$\mathbf{b}=(0,0)$になる．このため，表2の⑤で選ばれた(5,2)と(3,10)をサポートベクタに固定してモデルを解くと，⑤と同じ結果になる．すなわち，⑥はサポートベクタを考えた場合，目的関数を最小化する仮りの限界になる．しかし，図1から分かるとおり，マージンが最大のSVMは合格群では(12,4)，不合格群では(3,10)で，マージンは10.817になることは明らかである．判別関数としてf(x)=0.154＊勉強時間−0.103＊支出−0.436，最適な超平面は支出=1.5＊勉強時間−4.25になる．Cを場当たり的に小さな値(0.0074)にして⑥を計算しても，これを求めることは難しかった．

4.2 改定LP-OLDFと改定IP-OLDFの結果

表2には，改定LP-OLDFと改定IP-OLDFの結果が示してある．改定LP-OLDFの結果は，SVMのCが0.3以上と同じである．SVMはいってみれば$\|b\|^2$と$\sum e_i$の2目的最適化である．Cを大きくしていくことは結局$\|b\|^2$の影響を無視し改定LP-OLDFと等しくなる．

改定IP-OLDFの結果を見ると，マージンは限りなく0に近く，SVMでいうところの汎化性すなわちオーバーエスティメイトが疑われる．誤分類数は5個と改定LP-OLDFより1件少ない．しかし，先に述べたように2万件乱数データでは，汎化性は悪くない．

4.3 Cの問題点

多目的最適化の欠点は，恣意的なCの値によって，結果が異なってくることである．一体，①から⑥のいずれを選ぶのが正しいか判断できない．

さらにSVMのCは，今回6つの異なった状態で説明できたが，データ数が多くなるほど，この状態は増えてくる問題もある．

4.4 Cの役割をシミュレートする

ここでcの役割をシミュレーションする．改定LP-OLDFと同じ①は，支出=5の3つの○の3点を回転の中心とし，③まで回転する．

しかし，③から④への道筋はまだ明らかでないが，③の後は点 (5,3)，(3,7) を回転の中心に選び，その後合格群と不合格群の一方だけがあるケースに固定され，もう一方はケースに固定されないで④まで変化するように思われる．今後，この解明を行う必要がある．

5. まとめ

5.1 ペナルティCの検証

SVMでは，ペナルティCの値によって，得られるサポートベクタがデータ空間でダイナミックに異なってきて，どれを選べばよいかの問題があった．これは，多目的最適化に共通する問題点である．

あるいは，数多くの代替案の中からCをチューニングして汎化性の良いものに決定するというのであれば，他の手法に比べて優位に立つのは当然である．

今後SVM研究では，Cの客観的な決定法と外部標本による汎化性の検証が望まれる．

一つの提案として，今回分かったように，改定LP-OLDFはCが無限大からかなり小さな実数までの範囲で変更したソフトマージン最大化のSVMと一致することが分かった．また，Cを正の小さな値に減少させていき，サポートベクタがなくなる寸前のC値の分析結果を求める．そして，ここで選ばれたサポートベクタに関係するケースを固定して，C=0にしてこの変更モデルを再計算する．このとき，モデルに解があれば，マージンがほぼ最大になる限界を示すと考えられる．

すなわち，SVMにおけるCの決定は，改定LP-OLDFとC=0でサポートベクタのある2つのモデルの上限と下限を明らかにし，その範囲の中でCをどう選ぶかの議論が必要であろう．

5.2 今後の課題

統計の知恵に「けちの原理」がある．予測モデルは，できるだけ少ない説明変数を用い，多くの内部標本でモデル構築しなければ，汎化能力に劣る予測モデルが導かれるという結論である[1]．本研究では，統計研究家の多くが抱いているSVMへの不信の一つを検証した．

ただし，マージンを最大化すれば，カーネルトリックのような高次元で判別を行っても汎化能力が優れているというSVMの最大の売りは，統計のこれまでの知識と背反するが，今後の検討課題としたい．

SVMはマージン概念を最大化ということで今日大きな研究勢力になっているが，今後マージン概念を取り入れた改定IP-OLDFとSVMをスイス銀行紙幣データのような実データを内部標本とし，それと同じ分散共分散行列をもつ乱数を外部標本として比較研究を行いたい．すなわち，改定IP-OLDFのMMN基準とSVMのマージン最大化基準の比較が今後の課題になる．

参考文献

[1] Miyake, A., Shinmura, S.: Error rate of linear discriminant function. F. T. de Dombal & F. Gremy, editors, North-Holland Publishing Company, 435-445 (1976).

[2] Liitschwager, J. M., Wang, C.: Integer programming solution of a classification problem. Management Science, 24/14, 1515-1525 (1978).

[3] 三宅章彦，新村秀一：最適線形判別関数のアルゴリズムとその応用．医用電子と生体工学，18/1, 15-20 (1980).

[4] Glover, F.: Improve linear programming models for discriminant analysis. Decision Sciences, 2, 771-785 (1990).

[5] Stam, A.: Nontraditional approaches to statistical classification: Some perspectives on Lp-norm

methods. Annals of Operations Research, 74, 1-36 (1997).
[6] 新村秀一：数理計画法を用いた最適線形判別関数. 計算機統計学. 11/2, 89-101 (1998).
[7] 石田健一郎, 上田修功, 前田英作, 村瀬洋：パターン認識. オーム社 (1998).
[8] 新村秀一, 垂水共之：2変量正規乱数データによるIP-OLDFの評価. 計算機統計学, 12/2, 107-123 (1999).
[9] Shinmura, S.: A new algorithm of the linear discriminant function using integer programming. New Trends in Probability and Statistics 5, 133-142 (2000).
[10] 新村秀一：数理計画法を用いた最適線形判別関数(1), オペレーションズ・リサーチ, 47/1, 38-45 (2002).
[11] 新村秀一：数理計画法を用いた最適線形判別関数(2), オペレーションズ・リサーチ, 47/2, 109-113 (2002).
[12] 新村秀一：数理計画法を用いた最適線形判別関数(3), オペレーションズ・リサーチ, 47/3, 172-185 (2002).
[13] 新村秀一, 数理計画法を用いた最適線形判別関数(4), オペレーションズ・リサーチ, 47/4, 244-250 (2002).
[14] 新村秀一：数理計画法を用いた最適線形判別関数(5), オペレーションズ・リサーチ, 47/5, 315-321 (2002).
[15] 新村秀一：JMP活用統計学とっておき勉強法. 講談社 (2004).
[16] 新村秀一：整数計画法による判別分析の新世紀—1000スイスフラン偽札紙幣の分析, 日本オペレーションズ・リサーチ学会2005年春季研究発表会, 118-119 (2005).
[17] 新村秀一：従来の整数計画法による判別モデル研究に対する批判, 日本オペレーションズ・リサーチ学会2005年春季研究発表会, 120-121 (2005).
[18] 新村秀一：IP-OLDFによる線形判別関数の新しいモデル選択法の提案, 日本計算機統計学会, (2005). 投稿中.