SVMの2次計画問題に関する解法の考察

戸田　健一（沼田　一道助教授）

1　はじめに

人間の情報処理を支援強化するための手段であるパターン認識の分野において，サポートベクターマシン (Support Vector Machine, SVM) と呼ばれる学習機械が提唱されている．SVMは2クラスのパターン判別手法であり，判別能力に優れている点が特に注目されている．しかしSVMは，パターン判別問題を2次計画問題として定式化して解くが，サンプル数が多くなるにつれて問題が巨大なものになってしまい，一般的な2次計画問題の解法では計算が困難になる．そこで，Platt[1]は，SVMに由来する2次計画問題の特殊性に注目し，Sequential Minimal Optimization(SMO) と呼ばれる解法を提案した．SMOは，2変数の部分問題を逐次最適化していく手法で，計算量の飛躍的な減少を見込める効率的な解法である．部分問題でとりあげる2変数はヒューリスティックな方法により決定されるが，選び方によって全体の計算量が大きく違ってくる．本研究では，SMOの変数の選び方に着目し，2回に分けてSMOを実行する2段階SMOを提案し，その性能を評価する．

2　SVM

文字認識などのパターン認識問題においては，既知のデータとして n 個のサンプルベクトル $\mathbf{x}_i$ とその所属クラス y_i を学習し，未知データ $\mathbf{x}$ が与えられたときに学習データに基づきその所属クラスを判別するという方法がとられる．

学習データは特徴量の次元が m で，各データは χ_1, χ_2 のどちらかに属するとする．

$$\mathbf{x}_i = (x_{i_1}, \ldots, x_{im}), \quad y_i = \begin{cases} 1 & \mathbf{x}_i \in \chi_1 \\ -1 & \mathbf{x}_i \in \chi_2 \end{cases} \qquad (i = 1, \ldots, n)$$

このように与えられた学習データに対して，判別境界付近に位置するデータの「最適分離」を目的として考え出されたのがSVMである．SVMは，判別関数を

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} - b \tag{1}$$

とおき，判別境界を $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} - b = 0$ として，$f(\mathbf{x}) \geq 0$ ならば $\mathbf{x} \in \chi_1$，$f(\mathbf{x}) < 0$ ならば $\mathbf{x} \in \chi_2$ と判別する．$\mathbf{w}$ は重みベクトル，b は閾値と呼ばれるパラメータである．

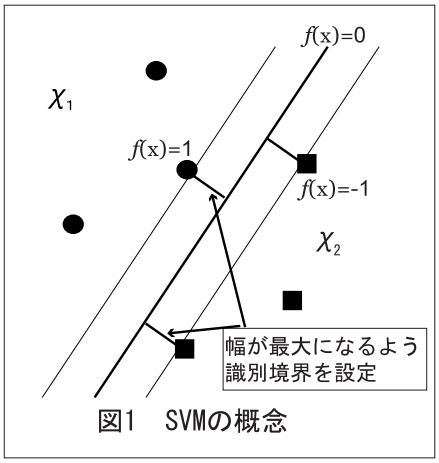

図1　SVMの概念

まず，与えられている学習データが線形関数により誤りなく判別できる場合を考える．このとき，判別境界 $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} - b = 0$ がクラス χ_1, χ_2 を正しく分離するという条件のもとで，最も望ましい $\mathbf{w}, b$ を次のように決定する．いま，それぞれのクラスについて判別境界に最も近い点の判別関数値を ± 1 とする．これは $\mathbf{w}$ の方向を変えずに設定することができる．このとき，$\mathbf{x}_i \in \chi_1$ ならば $f(\mathbf{x}) \geq 1$，$\mathbf{x}_i \in \chi_2$ ならば $f(\mathbf{x}) \leq -1$ となる．判別関数値が ± 1 である点はサポートベクターと呼ばれる．点 $\mathbf{x}$ と超平面 $f(\mathbf{x}) = 0$ との距離は $\frac{|\mathbf{w}^T\mathbf{x} - b|}{\|\mathbf{w}\|}$ となり，$f(\mathbf{x}) = \pm 1$ となる点（サポートベクター）との距離は $\frac{1}{\|\mathbf{w}\|}$ である．また，2つの超平面 $f(\mathbf{x}) = \pm 1$ の間の距離（マージン）は $\frac{2}{\|\mathbf{w}\|}$ であり，マージンをできるだけ大きくする．以上より「最適な」判別関数の $\mathbf{w}, b$ を求める問題は，各点の関数値を制約条件，マージンの逆数最小化を目的関数とした，次のような凸2次計画問題として定式化される．

$$\left| \begin{array}{ll} \text{minimize} & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{subject to} & y_i(\mathbf{w}^T\mathbf{x}_i - b) \geq 1 \quad (i = 1, \ldots, n) \end{array} \right. \tag{2}$$

上記の問題を解いて，求まった解 $\mathbf{w}^*, b^*$ を用いて最終的な判別関数を以下のように作成する．

$$g(\mathbf{x}_i) = \mathrm{sign}(\mathbf{w}^{*T}\mathbf{x} - b^*) = \begin{cases} 1 & \mathbf{x}_i \in \chi_1 \\ -1 & \mathbf{x}_i \in \chi_2 \end{cases} \tag{3}$$

ここで，$\mathrm{sign}(x)$ は，$x \geq 0$ のとき 1 を，$x < 0$ のとき -1 を返す関数である．

以上は線形分離可能な場合であるが，誤判別が避けられない（線形分離不可能）場合は次のように考える．すなわち，誤判別するデータに対し，補正量として非負変数 ξ_i を導入した上で制約条件を満たすように問題を定式化する．

$$\left| \begin{array}{ll} \text{minimize} & \dfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \\ \text{subject to} & y_i(\mathbf{w}^T\mathbf{x}_i - b) \geq 1 - \xi_i \quad (i = 1, \ldots, n) \end{array} \right. \tag{4}$$

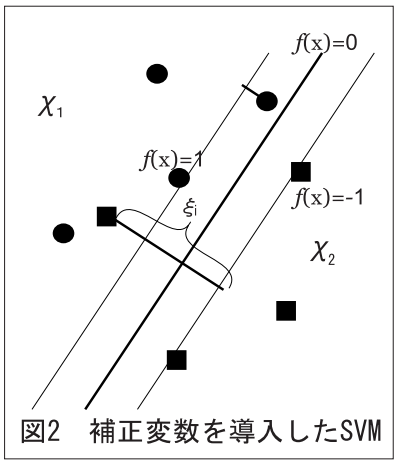

図2　補正変数を導入したSVM

ここで ξ_i は，あるサンプルが本来とは逆のクラスであると判別されてしまったときに，その値を補正する変数である．また C は，マージン最大化（マージンの逆数最小化）と誤判別を許す程度の間の重みを決めるパラメータである．また，式 (3) のラグランジュ双対問題は次のように書ける．ここで，双対変数 α_i はラグランジュ乗数に対応している．

$$\left| \begin{array}{ll} \text{maximize} & \displaystyle\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to} & 0 \leq \alpha_i \leq C \quad (i = 1, \ldots, n) \\ & \displaystyle\sum_{i=1}^{n} y_i \alpha_i = 0 \end{array} \right. \tag{5}$$

双対問題を解いたとき主問題の最適解 $\mathbf{w}^*, b^*$ は，$\mathbf{w}^*$ については $\mathbf{w}^* = \displaystyle\sum_{i=1}^{n} y_i \alpha_i \mathbf{x}_i$ であり，b^* は任意のサポートベクター $\mathbf{x}_i$ を用いて $b^* = y_i - \mathbf{w}^{*T}\mathbf{x}_i$ と表せる．したがって判別関数は

$$f(\mathbf{x}) = \sum_{i=1}^{n} y_i \alpha_i \mathbf{x}_i^T \mathbf{x} - b^* \tag{6}$$

となる．

また，SVM は本質的には線形関数による判別モデルであるが，式（5）の内積部分 $\mathbf{x}_i^T\mathbf{x}$ をカーネル関数 $K(\mathbf{x}_i, \mathbf{x}_j)$ に置換することで，非線形判別境界による判別が可能となる．

3　SMO

一般的な 2 次計画問題の解法（有効制約法など）を SVM に由来する 2 次計画問題に適用しようとすると行列演算が必要となり，サンプル数が多くなるにつれ多量のメモリと計算時間が必要となる．そこで，SMO では目的関数である式 (5) について，ある 2 つの変数 α_{i_1}，α_{i_2} のみを「動かす変数」とみなし，他の変数を固定した部分問題を考える．このとき目的関数は，

$$\begin{aligned} W(\alpha_{i_1}, \alpha_{i_2}) &= \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{n=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &= \alpha_{i_1} + \alpha_{i_2} + \sum_{\substack{i=1 \\ i \neq i_1, i_2}}^{n} \alpha_i - \frac{1}{2}K(\mathbf{x}_{i_1}, \mathbf{x}_{i_1})\alpha_{i_1}^2 - \frac{1}{2}K(\mathbf{x}_{i_2}, \mathbf{x}_{i_2})\alpha_{i_2}^2 \\ &\quad -\frac{1}{2}\sum_{\substack{i=1 \\ i \neq i_1, i_2}}^{n} y_i y_{i_1} K(\mathbf{x}_i, \mathbf{x}_{i_1})\alpha_i \alpha_{i_1} - \frac{1}{2}\sum_{\substack{i=1 \\ i \neq i_1, i_2}}^{n} y_i y_{i_2} K(\mathbf{x}_i, \mathbf{x}_{i_2})\alpha_i \alpha_{i_2} \end{aligned}$$

$$-\frac{1}{2}y_{i_1}y_{i_2}K(\mathbf{x}_{i_1},\mathbf{x}_{i_2})\alpha_{i_1}\alpha_{i_2}-\frac{1}{2}\sum_{\substack{i=1\\ i\neq i_1,i_2}}^{n}\sum_{\substack{j=1\\ j\neq i_1,i_2}}^{n}y_iy_jK(\mathbf{x}_i,\mathbf{x}_j)\alpha_i\alpha_j$$

$$= \alpha_{i_1}+\alpha_{i_2}+f_1(\alpha_{i_1}^2)+f_2(\alpha_{i_2}^2)+f_3(\alpha_{i_1}\alpha_{i_2})+f_4(\alpha_{i_1})+f_5(\alpha_{i_2})+W_{Const.}$$

という 2 変数の 2 次関数として表される．さらに，制約条件

$$\sum_{i=1}^{n}y_i\alpha_i=0 \quad \Leftrightarrow \quad y_{i_1}\alpha_{i_1}+y_{i_2}\alpha_{i_2}=-\sum_{\substack{i=1\\ i\neq i_1,i_2}}^{n}y_i\alpha_i(=\text{Const.}) \tag{7}$$

を代入して α_{i_1} を消去ことにより，目的関数は 1 変数関数となる．各変数のとりうる値は $0\leq\alpha_i\leq C$ であるから，2 変数の存在領域は図 3 のようになる．こうして生成した部分問題は放物線の頂点が最適解である.（ただし，放物線の頂点が双対変数の実行可能領域を外れてしまった場合，実行可能領域の範囲でもっとも頂点に近い点が最適解である.）部分問題を繰り返し解いていき，すべての α_i がクラス y_i，式 (5) の判別関数 $f(\mathbf{x})$ によって

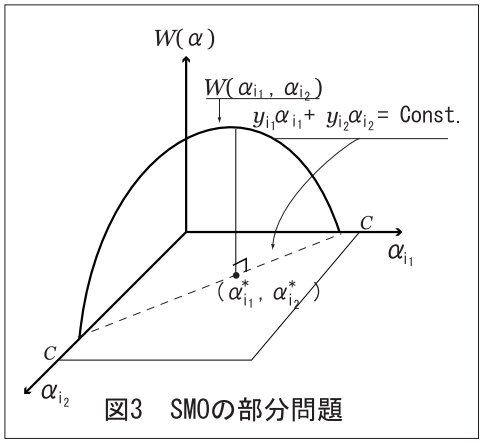

図3　SMOの部分問題

$$\begin{aligned}\alpha_i=0 \quad &\Leftrightarrow \quad y_if(\mathbf{x}_i)\geq 1\\ 0<\alpha_i<C \quad &\Leftrightarrow \quad y_if(\mathbf{x}_i)=1\\ \alpha_i=C \quad &\Leftrightarrow \quad y_if(\mathbf{x}_i)\leq 1\end{aligned} \tag{8}$$

を満たせば（KKT 条件），最適解となり終了する．

SMO の手順を以下に示す.

Step1　KKT 条件を満たしていない α_{i_2} を順番に選ぶ.
Step2　α_{i_2} に対して，以下の優先度で α_{i_1} を選ぶ.
Step2-1　$0<\alpha_i<C$ なる学習データが 2 つ以上ある場合，それらについて，α_{i_2} に対して最も更新幅が大きくなるような α_{i_1} を選ぶ.
Step2-2　Step2-1 でなかった場合，$\alpha_i=0$ または $\alpha_i=C$ なる学習データについて，ランダムな順番で α_{i_1} を選ぶ.
Step2-3　Step2-2 でなかった場合，すべての学習データについて，ランダムな順番で α_{i_1} を選ぶ.
Step3　選んだ 2 変数 $\alpha_{i_1},\alpha_{i_2}$ に対する部分問題を解く．更新した場合，Step1 へ.
Step2-1 を経て更新していない場合，Step2-2 へ.
Step2-2，2-3 を経てすべての α_{i_1} について更新していない場合，Step4 へ.
Step4　すべての α_{i_2} について Step2 を経ていて，更新しなかった場合，最適解となり終了.
そうでない場合，α_{i_2} を変更して Step2 へ.

SMO を利用した SVM 問題を $n=100$,200,500，正規分布

$$\chi_1 \quad : \quad \mu_1=\begin{pmatrix}100\\100\end{pmatrix},\quad \Sigma_1=\begin{pmatrix}60^2&0\\0&40^2\end{pmatrix}$$

$$\chi_2 \quad : \quad \mu_2=\begin{pmatrix}300\\300\end{pmatrix},\quad \Sigma_2=\begin{pmatrix}60^2&0\\0&80^2\end{pmatrix}$$

に従う乱数で生成した．プログラム環境は Boland 社の Delphi 6 Personal Edition を利用した．図 4 に実行画面を示す．また結果は 6 節の表 1，表 2 で示す.

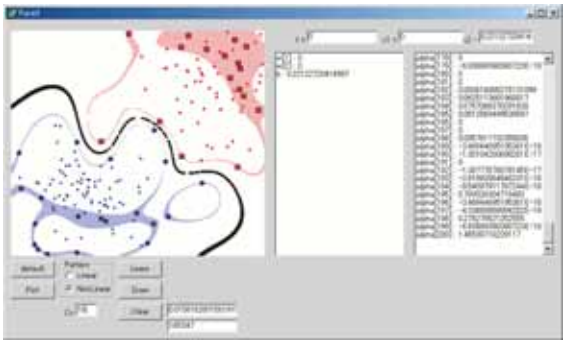
図4　実行画面

SMO は 2 変数の選び方次第で実行速度が大きく変わってくる．そこで，以下の方法でより効率の良い変数の選び方を行い，実行時間の短縮を試みる.

4　予備問題による2段階SMO

SVMは，χ_1, χ_2の境界付近のデータのみが判別境界に関与する．このため，境界付近のデータ，すなわちサポートベクターになると考えられるサンプルだけを抽出して問題を解ければ，サンプル数の減少による2次計画問題の計算量の減少が見込める．しかし，境界付近のデータのみを事前に抽出することは難しい．そこで，本研究では予備問題を作成し暫定的なサポートベクターを作り，その後本問題を解くという方法を提案する．その手順を以下に示す．

Step1 χ_1, χ_2に属するデータのそれぞれの平均値を算出し，μ_1, μ_2とする．
Step2 μ_1, μ_2を対角頂点とする超直方体に含まれるサンプルのみについて，SMOを実行する．
Step3 Step2により求まった解を初期解として，全サンプルについてSMOを実行する．

Step2ではサンプル数を減少させた上で，少ない計算量で仮のサポートベクターを作成し，Step3では，Step2で使われなかったサンプルによる初期解の修正を行う．以上により，全体の実行時間を短くすることを試みる．

5　実験

SMOに対する2段階SMOの実行速度を評価するため，プログラムによる数値実験を行った．条件は3節で行ったものと同様である．同じサンプルデータに対して，どの程度実行時間の減少を見込めるか検証を行った．

表1　線形SVMによる実行結果

学習データ数	SMO	2段階SMO	短縮時間比率
	実行時間(s)		
100	1.4322	1.3316	**0.9298**
200	4.092	3.8412	**0.9387**
500	7.6386	7.2482	**0.9489**

パラメータ：C=10

表2　非線形SVMによる実行結果

学習データ数	SMO	2段階SMO	短縮時間比率
	実行時間(s)		
100	1.8604	1.2678	**0.6815**
200	13.311	11.9574	**0.8983**
500	67.1788	60.1366	**0.8952**

パラメータ：ガウシアンカーネル，
δ=100，C=10

6　実験結果および考察

表1，表2の結果から，2段階SMOはSMOに対して，線形判別，非線形判別のいずれに対しても10%前後の実行時間減少にとどまった．思うように速度が向上しなかった理由としては，サンプル数の絞込みが思ったほどできなかったことと，2回目のSMOで解を修正するときに誤判別しているデータが多くなると，その修正作業に多量の時間を費やしてしまったためと考えられる．よって，より効率の良いサンプル数の減少法を考案するべきであった．

7　まとめ

本研究ではSVMに由来する2次計画問題を2段階SMOで解くことにより実行速度の向上を試みたが，いずれの場合においても速度向上はわずかなものにとどまった．また今回はサンプルデータとして正規分布に従う乱数を用いたが，今回の手法が，正規分布のように整然としてはいない特殊な例題にも対応できるものであるとはいいがたい．今後の課題としては，2段階SMOにおける予備問題をより効果的に解くための問題設定や，サポートベクターを初期のうちに指定できるような解法の開発などが挙げられる．

参考文献

[1] J.Platt： Sequential Minimal Optimization - A Fast Algorithm for Training Support Vector Machines；Microsoft Technical Report，1998.

[2] G.Mak： The Implementation Of Support Vector Machines Using The Sequential Minimal Optimization Algorithm； McGill University Master's thesis， 2000.

[3] 前田英作： 痛快！サポートベクトルマシン； 情報処理，42巻7号，pp．676-683，2001.