

SVMの 2次計画問題に関する解法の考察

東京理科大学工学部経営工学科

沼田研究室

4400079 戸田健一

発表構成

1. はじめに
2. クラス判別とSVM
 - 2-1. クラス判別問題
 - 2-2. サポートベクターマシン
 - 2-3. 線形分離不可能な場合のSVM
 - 2-4. 非線形SVM
3. Sequential Minimal Optimization(SMO)
4. 2段階SMO
5. 実験
6. 実験結果, 考察
7. まとめ

参考文献

1.はじめに

- パターン認識問題におけるクラス判別手法：
サポートベクターマシン(SVM)^{[1],抄録[3]}
- SVM:2次計画問題を解く
データが多くなるにつれて計算量が非常に多くなる
- Sequential Minimal Optimization (SMO)^{[2],[3],抄録[1],[2]}：
SVMによる2次計画問題の計算量を減少

- 本研究の目的：
2段階SMOの提案, 評価

2. クラス判別とSVM

2-1. クラス判別問題

学習データ: $\mathbf{x}_i = (x_{i_1}, x_{i_2}, \dots, x_{i_m})$

所属クラス: $y_i = \begin{cases} 1 & \mathbf{x}_i \in \mathcal{X}_1 \\ -1 & \mathbf{x}_i \in \mathcal{X}_2 \end{cases}$ から

判別境界: $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} - b = 0$ を作成し,

最終的な判別関数: $g(\mathbf{x}_i) = \text{sign}(f(\mathbf{x}_i)) = \text{sign}(\mathbf{w}^T \mathbf{x}_i - b) \quad \dots(1)$

$$\left[\text{sign}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases} \right]$$

が, なるべく $y_i = \begin{cases} 1 & \mathbf{x}_i \in \mathcal{X}_1 \\ -1 & \mathbf{x}_i \in \mathcal{X}_2 \end{cases}$ と一致するような

\mathbf{w}, b を求める問題: **クラス判別問題**

2-2. サポートベクターマシン

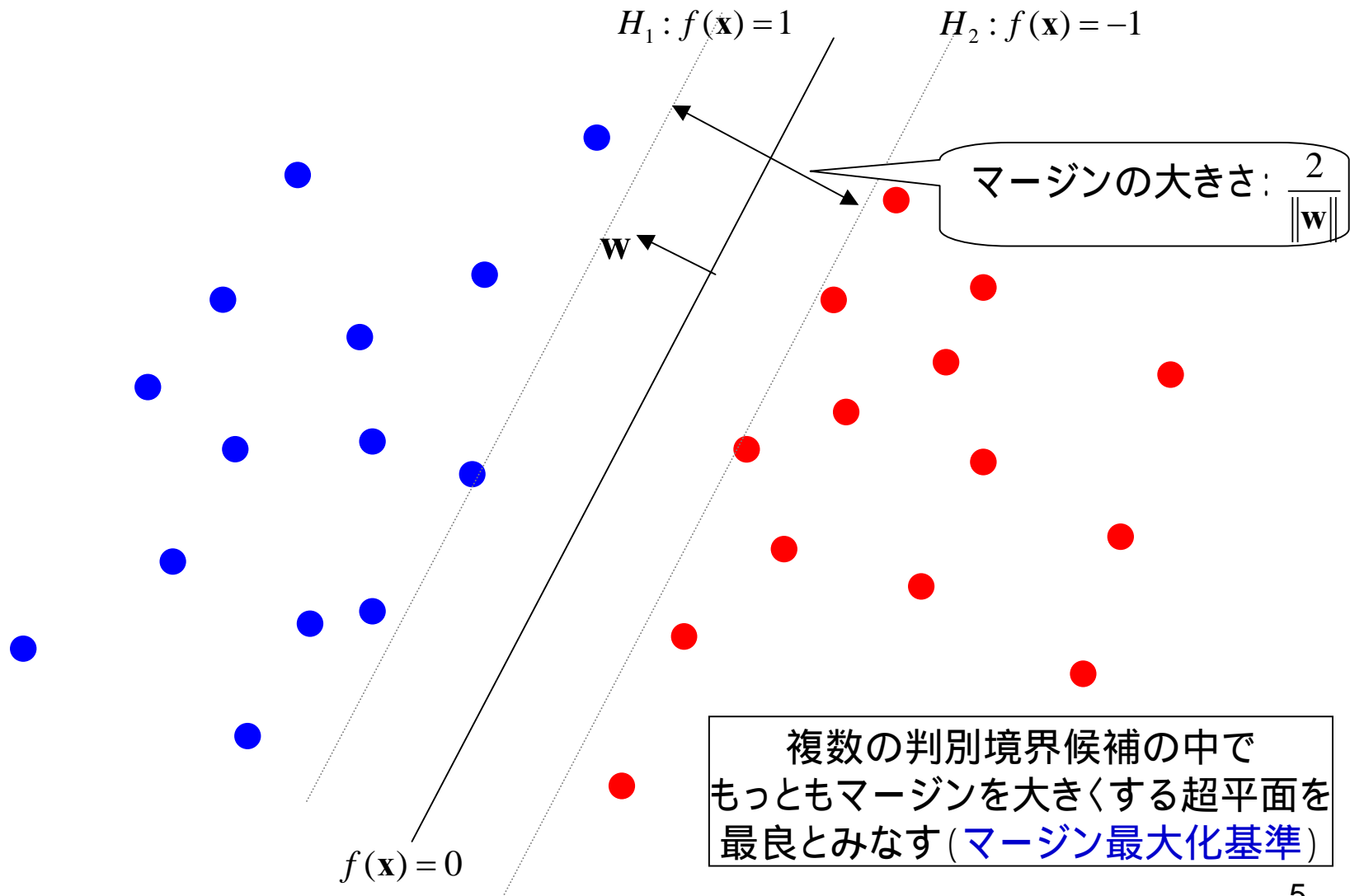


図2 サポートベクターマシン

2-2. SVM (定式化)

制約条件: $\mathbf{w}^T \mathbf{x}_i - b \begin{cases} \geq 1 & \mathbf{x}_i \in \mathcal{X}_1 \\ \leq -1 & \mathbf{x}_i \in \mathcal{X}_2 \end{cases} \Rightarrow y_i (\mathbf{w}^T \mathbf{x} - b) \geq 1$

目的関数: $\frac{2}{\|\mathbf{w}\|}$ 最大化 $\Rightarrow \frac{1}{2} \|\mathbf{w}\|^2$ 最小化

2-2. SVM (定式化)

SVMの数理計画問題:

$$\begin{cases} \underset{\mathbf{w}, b}{\text{minimize}} & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{subject to} & y_i \cdot (\mathbf{w}^T \mathbf{x}_i - b) \geq 1 \end{cases} \quad \dots(2)$$

これまでは線形分離可能を仮定

線形分離不可能な場合はどうするか？

2-3. 線形分離不可能な場合のSVM

線形分離不可能な場合の
SVMによる数理計画問題:

$$(P) \begin{cases} \text{minimize}_{\mathbf{w}, b, \xi} & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to} & y_i \cdot (\mathbf{w}^T \mathbf{x}_i - b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{cases} \quad \dots (3)$$

補正した分だけ
ペナルティを加算

誤判別しているとき
判別関数値を補正

C : マージン最大化の評価に対する
誤判別の程度の重みを決めるパラメータ

2-3. 線形分離不可能な場合のSVM

ラグランジュ関数 L [4],[5] を導入:

$$L(\mathbf{w}, b, \xi, \alpha, \gamma) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$$

制約式を足しこむ

$$- \sum_{i=1}^n \alpha_i \{y_i (\mathbf{w}^T \mathbf{x}_i - b) - (1 - \xi_i)\} - \sum_{i=1}^n \gamma_i \xi_i \quad \dots (4)$$

$L(\mathbf{w}, b, \xi, \alpha, \gamma)$ を, , について最大化する問題: **双対問題**

$$(D) \begin{cases} \underset{\alpha}{\text{maximize}} & L(\mathbf{w}, b, \xi, \alpha, \gamma) \\ \text{subject to} & \alpha_i \geq 0 (i = 1, \dots, n) \\ & \gamma_i \geq 0 (i = 1, \dots, n) \end{cases}$$

2-3. 線形分離不可能な場合のSVM

主問題(P)と双対問題(D)について

Karush-Kuhn-Tucker条件(KKT条件):

$\frac{\partial L}{\partial w_i} = 0 (i = 1, \dots, n),$	$\frac{\partial L}{\partial b} = 0,$	$\frac{\partial L}{\partial \xi_i} = 0 (i = 1, \dots, n)$	$\dots(5)$	(停留点)
$\frac{\partial L}{\partial \alpha_i} \leq 0 (i = 1, \dots, n),$	$\frac{\partial L}{\partial \gamma_i} \leq 0 (i = 1, \dots, n)$	$\dots(6)$		(主実行可能性)
$\alpha_i \geq 0 (i = 1, \dots, n),$	$\gamma_i \geq 0 (i = 1, \dots, n)$	$\dots(7)$		(双対実行可能性)
$\alpha_i \frac{\partial L}{\partial \alpha_i} = 0 (i = 1, \dots, n),$	$\alpha_i \frac{\partial L}{\partial \gamma_i} = 0 (i = 1, \dots, n)$	$\dots(8)$		(相補性)

を満たすとき、(P),(D)は最適解である

2-3. 線形分離不可能な場合のSVM

停留点の条件:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}_i} = 0 \quad (i = 1, \dots, n) &\longrightarrow \mathbf{w} = \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i \\ \frac{\partial L}{\partial b} = 0 &\longrightarrow \sum_{i=1}^n y_i \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \quad (i = 1, \dots, n) &\longrightarrow \alpha_i = C - \xi_i \quad (i = 1, \dots, n) \end{aligned}$$

をラグランジュ関数: $L(\mathbf{w}, b, \xi, \alpha, \gamma)$ に代入

$$\longrightarrow \text{(D)} \left\{ \begin{array}{l} \text{maximize} \quad \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to} \quad 0 \leq \alpha_i \leq C \quad (i = 1, \dots, n) \quad \dots (9) \\ \sum_{i=1}^n y_i \alpha_i = 0 \end{array} \right.$$

2-3. 線形分離不可能な場合のSVM

双対変数を用いた判別関数:

$$f(\mathbf{x}) = \sum_{i=1}^n y_i \alpha_i \mathbf{x}_i^T \mathbf{x} - b \quad (10)$$

KKT条件による最適解の条件:

$$\alpha_i = 0 \quad \Rightarrow \quad y_i f(\mathbf{x}_i) \geq 1$$

$$0 < \alpha_i < C \quad \Rightarrow \quad y_i f(\mathbf{x}_i) = 1$$

$$\alpha_i = C \quad \Rightarrow \quad y_i f(\mathbf{x}_i) \leq 1$$

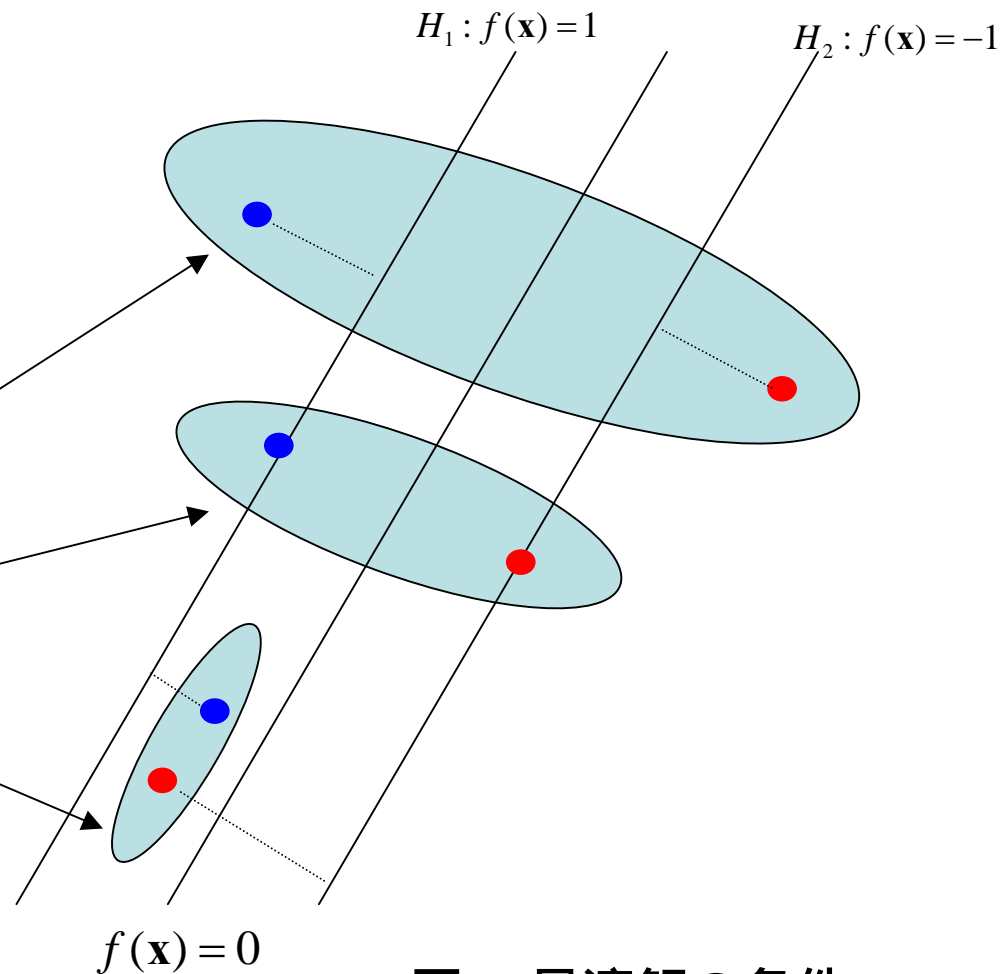


図3 最適解の条件

2-4. 非線形SVM

学習データを高次元に写像, 高次元空間で線形判別

もとの次元の空間では非線形判別境界が作成可能

$$\left\{ \begin{array}{l} \text{maximize} \quad \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j \mathbf{h}(\mathbf{x}_i)^T \mathbf{h}(\mathbf{x}_j) \\ \text{subject to} \quad 0 \leq \alpha_i \leq C \quad (i=1, \dots, n) \\ \sum_{i=1}^n y_i \alpha_i = 0 \end{array} \right. \quad \dots(9)'$$

$\mathbf{h}(\mathbf{x}_i)^T \mathbf{h}(\mathbf{x}_j)$ をカーネル関数 $K(\mathbf{x}_i, \mathbf{x}_j)$ で置換することにより
個々の学習データに対する写像が不要に

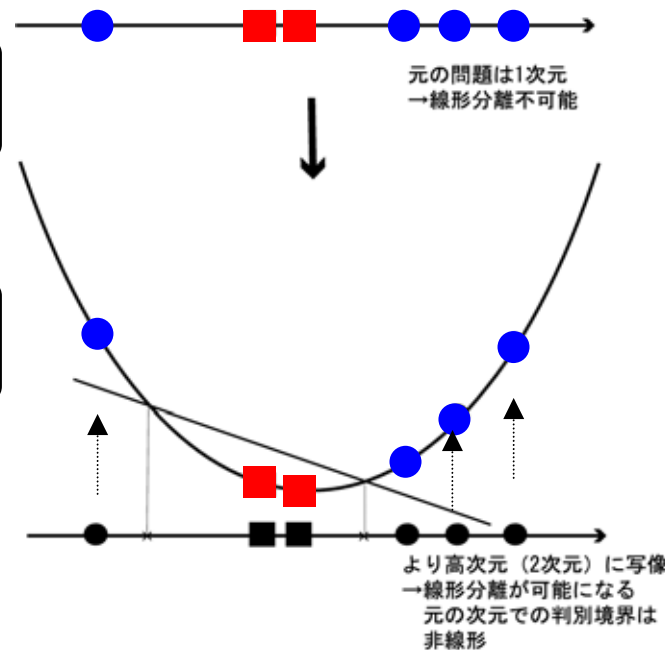


図4 非線形SVM

2-4. 非線形SVM

カーネル関数を用いた定式化:

$$(D) \begin{cases} \text{maximize} & \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to} & 0 \leq \alpha_i \leq C \\ & \sum_{i=1}^n y_i \alpha_i = 0 \end{cases} \quad \dots(11)$$

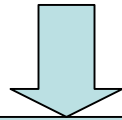
双対変数を用いた判別関数: $f(\mathbf{x}) = \sum_{i=1}^n y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) - b \quad (12)$

3. Sequential Minimal Optimization

一般的な2次計画問題の解法:

学習データが多くなるにつれて問題のサイズが増大

→ 多くの計算量が必要, 実行時間が長くなる



Sequential Minimal Optimization(SMO):

部分問題を生成し、逐次最適化

→ 行列演算が不要になり、高速に解くことができる

$$\left\{ \begin{array}{l} \text{maximize} \quad W(\boldsymbol{\alpha}) = \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to} \quad 0 \leq \alpha_i \leq C \\ \sum_{i=1}^n y_i \alpha_i = 0 \end{array} \right. \quad \dots(10)$$

3. SMO

2変数 $\alpha_{i_1}, \alpha_{i_2}$ を選び

他の変数を定数とみなして部分問題に

→ 目的関数 W : 2変数の2次関数

制約条件:

$$\sum_{i=1}^n y_i \alpha_i = 0 \Leftrightarrow y_{i_1} \alpha_{i_1} + y_{i_2} \alpha_{i_2} + \sum_{\substack{i=1 \\ i \neq i_1, i_2}}^n \alpha_i y_i = 0$$

$$\Leftrightarrow y_{i_1} \alpha_{i_1} + y_{i_2} \alpha_{i_2} = \text{Const.}$$

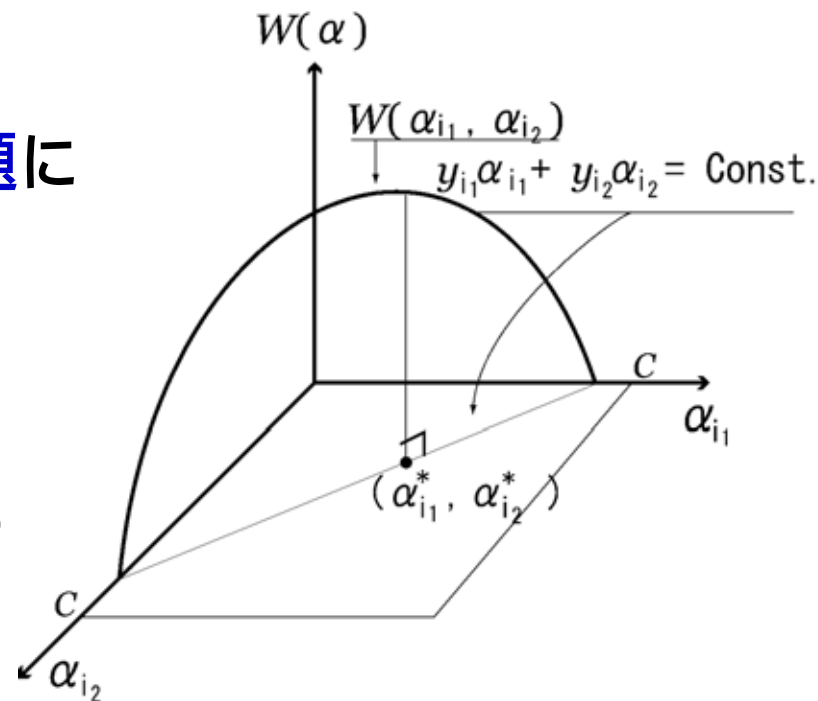


図5 部分問題の最適解

を代入 → 目的関数 W : 1変数の2次関数に

3. SMO

双対変数を用いた判別関数:

$$f(\mathbf{x}) = \sum_{i=1}^n y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) - b \quad (10)$$

KKT条件による最適解の条件:

$$\alpha_i = 0 \quad \Rightarrow \quad y_i f(\mathbf{x}_i) \geq 1$$

$$0 < \alpha_i < C \quad \Rightarrow \quad y_i f(\mathbf{x}_i) = 1$$

$$\alpha_i = C \quad \Rightarrow \quad y_i f(\mathbf{x}_i) \leq 1$$

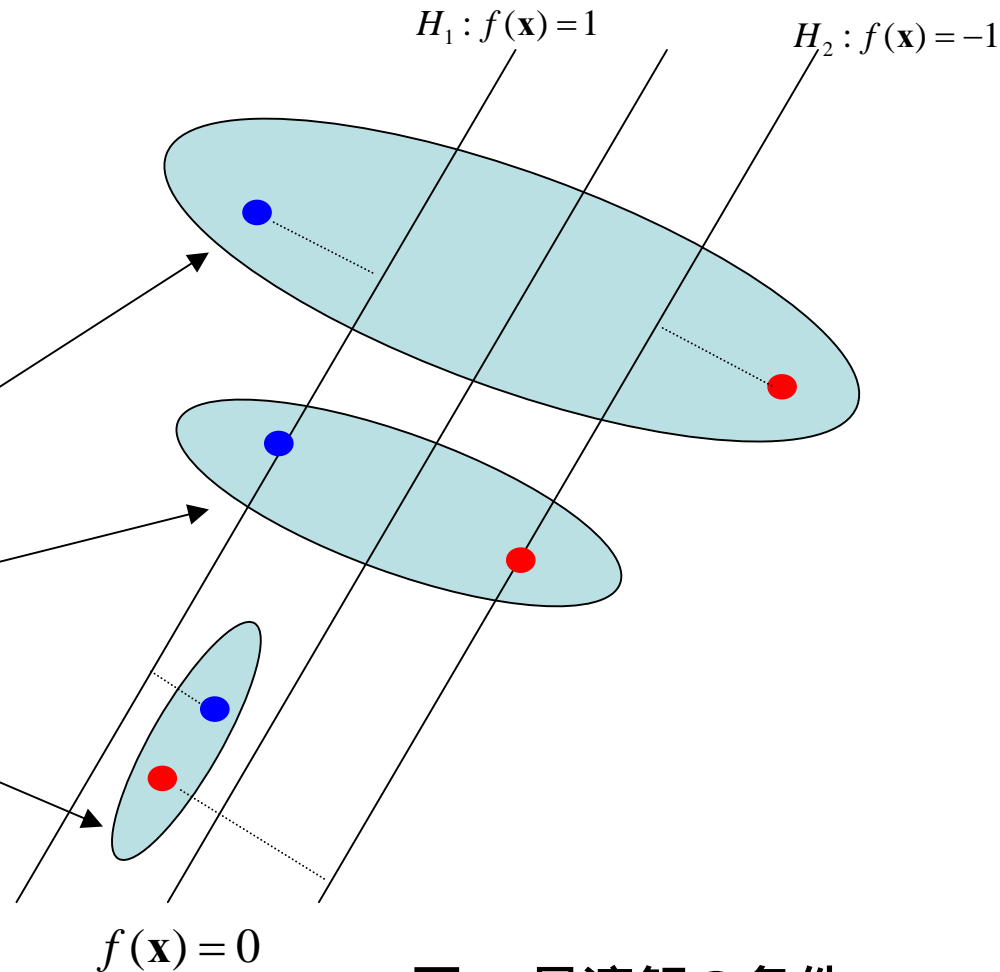


図6 最適解の条件

3. SMO

SMOのアルゴリズム

Step1. KKT条件を満たしていない i_2 を順番に選ぶ.

すべての i_2 がKKT条件を満たしていれば、最適解となり終了.

Step2. 以下の優先度で i_1 を選ぶ.

2-1. $0 < i_1 < C$ である i_1 について,

i_1, i_2 の更新幅がなるべく大きくなるような i_1 を選ぶ.

2-2. $i_1 = 0, i_1 = C$ である i_1 について,

i_1, i_2 の更新幅がなるべく大きくなるような i_1 を選ぶ.

2-3. ランダムに i_1 を選ぶ.

Step3. i_1, i_2 に関する部分問題を解く.

部分問題の最適解を得たらStep1へ.

4. 2段階SMO

SVM:

判別境界付近の学習データのみが判別境界に関係



2段階SMO:

予備問題で暫定的なサポートベクターを作り本問題を解く

4. 2段階SMO

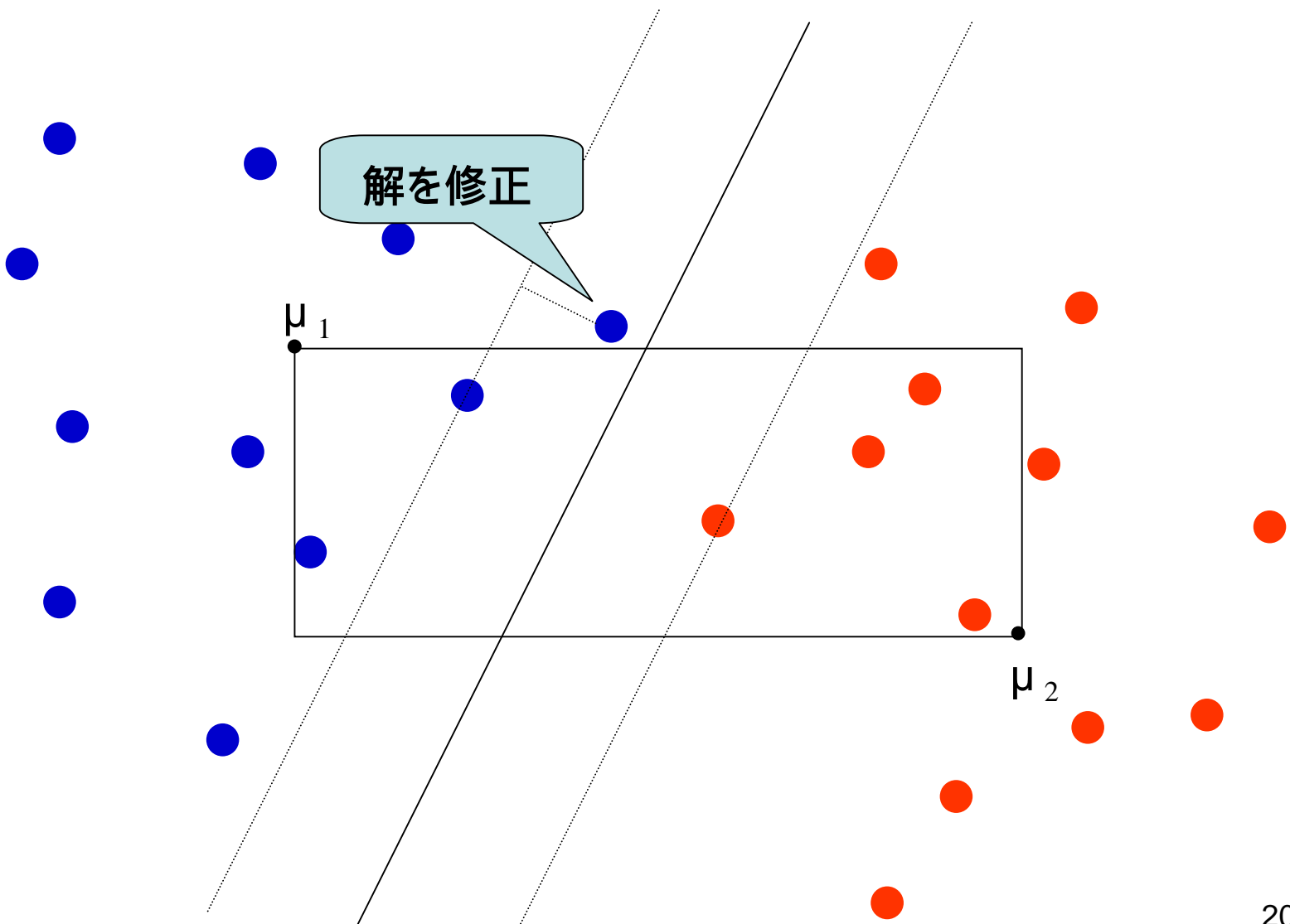


図7 2段階SMO

5. 実験

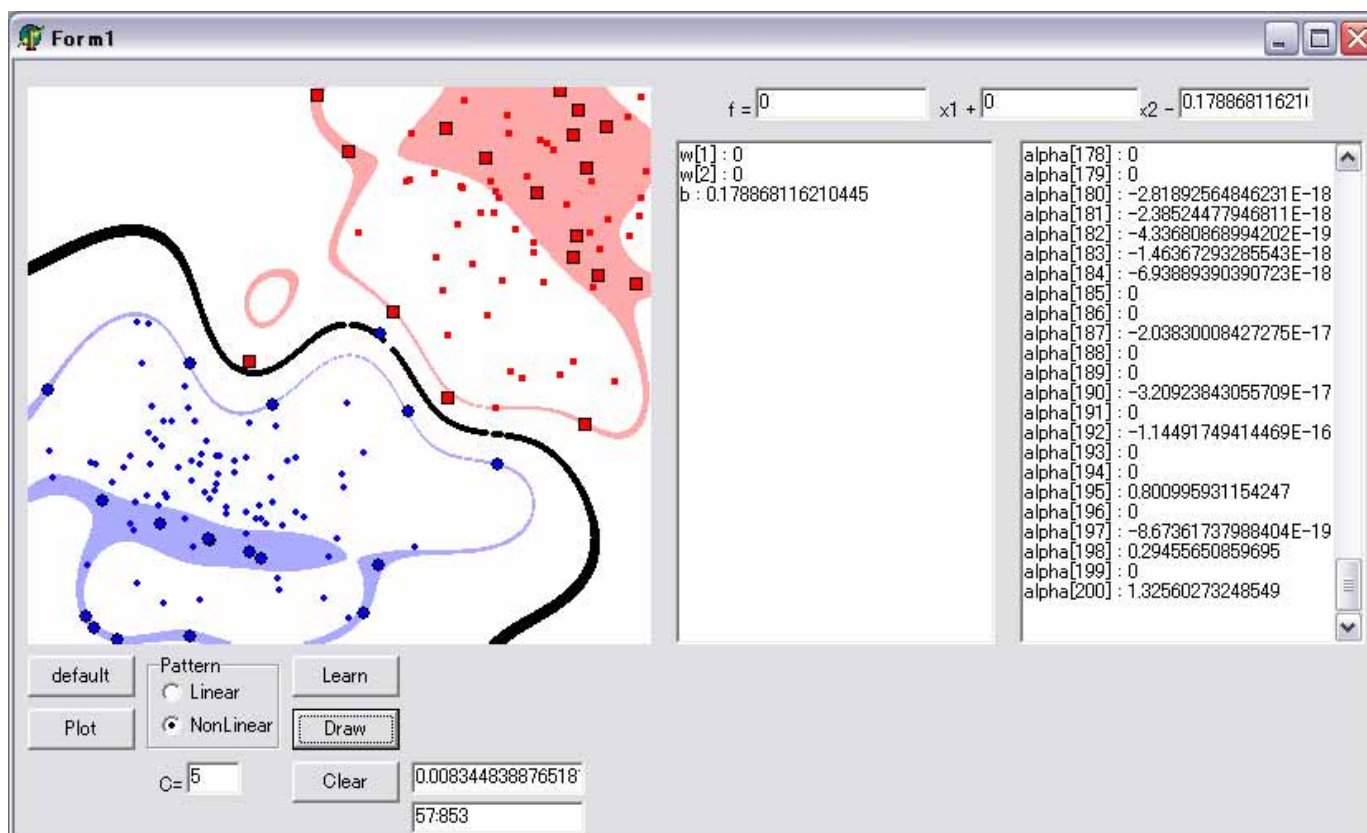


図8 実行画面

6. 実験結果，考察

表1 線形SVMによる実行時間

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

表2 非線形SVMによる実行時間

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

7.まとめ

SMOをより高速化させるべく2段階SMOを提案, 評価
速度向上はわずかなものにとどまった



2段階目でKKT条件を満たさない学習データが多かった



より効率の良い学習データの絞込み
整然としてない学習データに対する絞込み方法の考案が必要

参考文献

- [1] 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.
- [2] J.Platt：Sequential Minimal Optimization -A Fast Algorithm for Training Support Vector Machines；Microsoft Technical Report，1998.
- [3] Mak：The Implementation Of Support Vector Machines Using The Sequential Minimal Optimization Algorithm；McGill University Master's thesis，2000.
- [4] 矢部博，八巻直一：非線形計画法；朝倉書店，1999.
- [5] 福島雅夫：非線形最適化の基礎；朝倉書店，2001.

訂正

p.183,上から11行目 誤:式(5)の判別関数に...

————→ 正:式(6)の判別関数に...

p.183,下から5行目 誤: Boland社の...

————→ 正: Borland社の...