

# サポートベクターマシンにおける最適解の性質

城西大学 理学部 石川 咲絵  
札幌学院大学 経済学部・総合教育センター 中村 永友  
城西大学 理学部 土屋 高宏

要旨：本論文ではサポートベクターマシンにおける最適解の性質と解の非一意性について考察する。最適な分離超平面を挟む2つの超平面上に学習データがあっても、それが必ずしもサポートベクターになるとは限らないことが知られている。このような状況がどのような形で現れるのかを、あるデータの配置を通して理論的に検証する。また、これにより最適解の非一意性について触れ、数式処理ソフトウェアによる解析を行う。

キーワード：サポートベクター，分離超平面，最適化問題

## 1 はじめに

Vapnik (1998) によって提唱されたサポートベクターマシンは文字・音声認識などのパターン認識問題に対して優れた識別能力を発揮し、自然科学・社会科学における様々な分野で応用研究がなされ、新たな非線形識別・判別問題への適用が試みられつつある。統計的学習理論として知られる階層型ニューラルネットワークでは、パラメータ推定の際、初期値によって、局所最適解へ収束するという数値計算上の問題が生じた。一方、サポートベクターマシン (SVM) の学習理論においては凸2次計画問題として定式化され、目的関数の最適解が唯一決まるという利点があり、数値計算における1つの難点を克服した有効な手法といえる。目的関数の最適解において、Abe(2002)は、最適な分離超平面を挟む2つの超平面上のデータが、必ずしもサポートベクターになるとは限らないことを示し、さらに非一意な解の存在を指摘した。凸2次計画問題の目的関数の値は一意に求められるため、実質的には問題とならないが、非一意解の存在は数値計算上、留意すべき点であるといえる。

本論文では、Abe (2002)、阿部 (2011) に基づき、サポートベクターマシンにおける目的関数の最適解の性質について考察する。特に、サポートベクターにならない2つの超平面上のデータがどのような形で現れるのかを、あるデータの配置を通して理論的に調べるとともに、最適解の非一意性について検討する。さらに、数式処理ソフトウェアによる解析を試みる。

2節では、サポートベクターマシンの基礎理論を小西 (2010) に基づき、線形分離可能な場合と線形分離可

能でない場合に分けて述べる。3節では、サポートベクターマシンにおける最適解の性質について、ある特殊なデータの配置を通して理論的に検証し、解の非一意性について考察する。4節では、3節で取り上げた非一意な解が存在するデータについて、数式処理ソフトウェアによる解析を試みる。

## 2 サポートベクターマシンの基礎理論

個体の特徴づける  $p$  個の変数  $\mathbf{x}=(x_1, x_2, \dots, x_p)^T$  に関して観測された  $n$  個の学習データを  $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$  とする。ただし、 $y_i (i=1, 2, \dots, n)$  は 2 つの群  $G_1, G_2$  の所属を表すために導入されたラベル変数で

$$y_i = \begin{cases} 1 & (\mathbf{x}_i \in G_1) \\ -1 & (\mathbf{x}_i \in G_2) \end{cases}$$

とする。

サポートベクターマシンの目的は、学習データに基づいて 2 群を最もよく分離するような超平面  $\mathbf{w}^T \mathbf{x} + b = 0$  を構成することにある。ここで、 $\mathbf{w}=(w_1, w_2, \dots, w_p)^T$  は重みベクトル、 $b$  は切片である。

### 2.1 線形分離可能な場合

2 群の学習データが超平面によって完全に分離されている、すなわち線形分離可能な場合には、 $\mathbf{x}_i \in G_1 \Leftrightarrow \mathbf{w}^T \mathbf{x}_i + b > 0$  と  $\mathbf{x}_i \in G_2 \Leftrightarrow \mathbf{w}^T \mathbf{x}_i + b < 0$  が成り立っているとすると、すべてのデータに対して  $y_i(\mathbf{w}^T \mathbf{x}_i + b) > 0$  ( $i=1, 2, \dots, n$ ) が成立している。この線形分離可能な学習データに対して、最適な分離超平面を構成するための基準として用いられるのが、データと超平面との距離である。一般に、 $p$  次元データ  $\mathbf{x}_i$  と超平面  $\mathbf{w}^T \mathbf{x} + b = 0$  の距離は  $|\mathbf{w}^T \mathbf{x}_i + b| / \|\mathbf{w}\|$  で与えられる。ただし、 $\|\mathbf{w}\|$  は重みベクトルのノルムを表す。

線形分離可能な場合は 2 群を分離する 1 つの超平面  $H$  と、それを中間に挟む形で 2 つの超平面  $H_+$  と  $H_-$  が存在する。ここで、両端の超平面上には少なくとも 1 つのデータが存在し、その間にはデータは存在しないとする。この超平面  $H_+$  上のデータ  $\mathbf{x}_+$  あるいは  $H_-$  上のデータ  $\mathbf{x}_-$  と超平面  $H$  の距離

$$d = \frac{y_i(\mathbf{w}^T \mathbf{x}_+ + b)}{\|\mathbf{w}\|} = \frac{y_i(\mathbf{w}^T \mathbf{x}_- + b)}{\|\mathbf{w}\|}$$

をマージンとよび、2 群を最もよく分離する問題を、このマージンを最大とする分離超平面を求める問題へと帰着させる。すなわち、

$$\begin{cases} \max_{\mathbf{w}, b} d, \\ \text{制約条件} \quad \frac{y_i(\mathbf{w}^T \mathbf{x}_i + b)}{\|\mathbf{w}\|} \geq d, \quad i=1, 2, \dots, n \end{cases}$$

と定式化できる。一般に、超平面  $\mathbf{w}^T \mathbf{x} + b = 0$  の両辺に、任意の実数  $r > 0$  を掛けても超平面は不変であるから、マージン最大化の制約条件を

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i=1, 2, \dots, n$$

と表すことができる。特に、2 つの超平面  $H_+, H_-$  上にあるデータに対しては等号が成立し、 $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$  となるから、マージンは  $d = 1 / \|\mathbf{w}\|$  である。したがって、マージン最大化問題は次のように定式化される。

$$\begin{cases} \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2, \\ \text{制約条件} \quad y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i=1, 2, \dots, n. \end{cases}$$

サポートベクターマシンでは、上式を主問題としてラグランジュ関数

$$L(\mathbf{w}, b, \alpha_1, \alpha_2, \dots, \alpha_n) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\} \quad (1)$$

を導入して、双対問題とよばれる最適化問題に帰着させて解く。ラグランジュ関数(1)式を重みベクトル  $\mathbf{w}$  と切片  $b$  で偏微分して0とおくと、

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i, \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (2)$$

を得る。(2)式を(1)式へ代入した目的関数は

$$L_D(\alpha_1, \alpha_2, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

となるので、双対問題は次のように定式化される。

$$\begin{cases} \max_{\alpha_1, \dots, \alpha_n} L_D(\alpha_1, \dots, \alpha_n) = \max_{\alpha_1, \dots, \alpha_n} \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\}, \\ \text{制約条件} \quad \alpha_i \geq 0 \quad (i=1, 2, \dots, n), \quad \sum_{i=1}^n \alpha_i y_i = 0. \end{cases}$$

この双対問題の解を  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_n$  とすると、重みベクトルの最適解は

$$\hat{\mathbf{w}} = \sum_{i=1}^n \hat{\alpha}_i y_i \mathbf{x}_i$$

で与えられる。また、分離超平面  $\hat{\mathbf{w}}^T \mathbf{x} + b = 0$  を挟む2つの超平面  $H_+, H_-$  上のデータに対して、

$$H_+ : \hat{\mathbf{w}}^T \mathbf{x}_+ + b = 1, \quad H_- : \hat{\mathbf{w}}^T \mathbf{x}_- + b = -1$$

が成り立つことから、切片  $b$  の最適解は

$$\hat{b} = -\frac{1}{2}(\hat{\mathbf{w}}^T \mathbf{x}_+ + \hat{\mathbf{w}}^T \mathbf{x}_-)$$

で与えられる。

したがって、最適な分離超平面を判別関数とする2群  $G_1, G_2$  の判別法は

$$\hat{\mathbf{w}}^T \mathbf{x} + \hat{b} = \sum_{i=1}^n \hat{\alpha}_i y_i \mathbf{x}_i^T \mathbf{x} + \hat{b} \begin{cases} \geq 0 \Rightarrow \mathbf{x} \in G_1 \\ < 0 \Rightarrow \mathbf{x} \in G_2 \end{cases}$$

で与えられる。

ところで、双対問題の最適解は最適化問題の制約条件として知られる Karush-Kuhn-Tucker 条件 (KKT 条件) より

$$\alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\} = 0, \quad i=1, 2, \dots, n$$

が成り立つ。

データが最適な分離超平面を挟む2つの超平面  $H_+$ ,  $H_-$  上にない場合,  $y_i(\mathbf{w}^T \mathbf{x}_i + b) > 1$  であるから,  $\alpha_i = 0$  となる. 一方, データが超平面  $H_+$ ,  $H_-$  上にある場合,  $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$  が成立し,  $\alpha_i \geq 0$  となることがわかる.  $\alpha_i > 0$  となる実質的に判別関数を構成するデータ  $\mathbf{x}_i$  をサポートベクターとよぶ.

## 2.2 線形分離可能でない場合

線形分離可能な場合には, すべてのデータ  $\mathbf{x}_i$  に対して, 不等式  $y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1$  が満たされる. これに対して, 線形分離可能でない場合には, この不等式を満たさないデータが存在する. このような場合には, スラック変数とよばれる実数  $\xi_i \geq 0$  を導入して,

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \quad (3)$$

とする. スラック変数は, データ  $\mathbf{x}_i$  に対して次のように定義される.

$$\xi_i = \max\left\{0, d - \frac{y_i(\mathbf{w}^T \mathbf{x}_i + b)}{\|\mathbf{w}\|}\right\}.$$

スラック変数を導入したときのマージンをソフトマージンという.

線形分離可能でない学習データに基づく超平面を求める問題は, (3)式の制約条件の下で, マージン  $1/\|\mathbf{w}\|$  をできるだけ大きく, すなわち,  $\|\mathbf{w}\|^2$  をできるだけ小さくする超平面を求める問題に帰着される. しかしながら, マージンを小さくしようとする, スラック変数の総和  $\sum_{i=1}^n \xi_i$  は大きくなるため, 調整パラメータ  $\lambda > 0$  を用いて, 次の最適化問題へと帰着できる.

$$\begin{cases} \min_{\mathbf{w}, \xi} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + \lambda \sum_{i=1}^n \xi_i \right\}, \\ \text{制約条件 } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i=1, 2, \dots, n. \end{cases} \quad (4)$$

線形分離可能な場合と同様に, 上式を主問題としてラグランジュ関数

$$L(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2} \|\mathbf{w}\|^2 + \lambda \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 + \xi_i\} - \sum_{i=1}^n \beta_i \xi_i \quad (5)$$

を用いて, 双対問題へと置き換える. ただし,  $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ ,  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ ,  $\beta = (\beta_1, \beta_2, \dots, \beta_n)^T$  である. ラグランジュ関数(5)式を  $\mathbf{w}$ ,  $b$ ,  $\xi$  で偏微分して0とおくと,

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i, \quad \sum_{i=1}^n \alpha_i y_i = 0, \quad \lambda = \alpha_i + \beta_i, \quad i=1, 2, \dots, n \quad (6)$$

を得る. (6)式を(5)式へ代入すると, 線形分離可能な場合と同様に, 目的関数は

$$L_D(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

となる.  $\beta_i = \lambda - \alpha_i \geq 0$  および  $\alpha_i \geq 0$  より, 制約条件は  $0 \leq \alpha_i \leq \lambda$  となるので, 双対問題は次のように定式化される.

$$\begin{cases} \max_{\alpha} L_D(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\}, \\ \text{制約条件} \quad \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq \lambda, \quad i=1, 2, \dots, n. \end{cases} \quad (7)$$

この双対問題の解を  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_n$  とすると、重みベクトルの最適解は

$$\hat{\mathbf{w}} = \sum_{i=1}^n \hat{\alpha}_i y_i \mathbf{x}_i$$

で与えられる。また、切片  $b$  の最適解は分離超平面  $\hat{\mathbf{w}}^T \mathbf{x} + b = 0$  を挟む 2 つの超平面  $H_+, H_-$  上のデータを用いて、

$$\hat{b} = -\frac{1}{2} (\hat{\mathbf{w}}^T \mathbf{x}_+ + \hat{\mathbf{w}}^T \mathbf{x}_-)$$

で与えられる。

したがって、最適な分離超平面を判別関数とする 2 群  $G_1, G_2$  の判別法は、線形分離可能な場合と同様に

$$\hat{\mathbf{w}}^T \mathbf{x} + \hat{b} = \sum_{i=1}^n \hat{\alpha}_i y_i \mathbf{x}_i^T \mathbf{x} + \hat{b} \begin{cases} \geq 0 \Rightarrow \mathbf{x} \in G_1 \\ < 0 \Rightarrow \mathbf{x} \in G_2 \end{cases}$$

で与えられる。

KKT 条件より、線形分離可能でない場合は次の 2 つの式が成り立つ。

$$\alpha_i \{-y_i(\mathbf{w}^T \mathbf{x}_i + b) - \xi_i + 1\} = 0, \quad \beta_i (-\xi_i) = 0, \quad i=1, 2, \dots, n.$$

2 つの超平面  $H_+, H_-$  上およびその間の領域にない線形分離可能なデータに対しては、 $y_i(\mathbf{w}^T \mathbf{x}_i + b) > 1$  となり、 $-y_i(\mathbf{w}^T \mathbf{x}_i + b) - \xi_i + 1 < 0$  であるから、 $\alpha_i = 0$  となる。

超平面  $H_+, H_-$  の間の領域にあるデータおよび他の群の領域へ入り込んだ線形分離可能でないデータに対しては、 $y_i(\mathbf{w}^T \mathbf{x}_i + b) < 1$  となり、 $a = -y_i(\mathbf{w}^T \mathbf{x}_i + b) + 1$  とおくと、 $\alpha_i(a - \xi_i) = 0$  であるから、 $\alpha_i = 0$  または  $\xi_i = a > 0$  となる。ここで、 $\alpha_i = 0$  は不適である。なぜならば、条件より  $a - \xi_i \leq 0$  であり、 $a > 0$  であるから、 $\xi_i > 0$  となる。このとき、 $\beta_i = 0$  となるので、(6)式より  $\alpha_i = \lambda > 0$  が成立する。

超平面  $H_+, H_-$  上にあるデータに対しては、 $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$  が成立し、 $\alpha_i \xi_i = 0$  であるから、 $\alpha_i = 0$  または  $\xi_i = 0$  となる。 $\xi_i = 0$  のとき  $\beta_i \geq 0$  より、 $0 \leq \alpha_i \leq \lambda$  となる。

以上をまとめると、次のようになる。

$$\begin{aligned} y_i(\mathbf{w}^T \mathbf{x}_i + b) > 1 &\Rightarrow \alpha_i = 0, \\ y_i(\mathbf{w}^T \mathbf{x}_i + b) < 1 &\Rightarrow \alpha_i = \lambda, \\ y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1 &\Rightarrow 0 \leq \alpha_i \leq \lambda. \end{aligned}$$

したがって、データが 2 つの超平面  $H_+, H_-$  上にある場合、 $\alpha_i = 0$  となるデータ  $\mathbf{x}_i$  が存在する可能性があることがわかる。このことは、そのようなデータ  $\mathbf{x}_i$  がサポートベクターにならないことを意味する。

次節では、2 つの超平面  $H_+, H_-$  上のデータが必ずしもサポートベクターになるとは限らないことを、あるデータの配置を通して検証する。

### 3 最適解の性質と解の非一意性

Abe (2002) は 2 次計画問題の目的関数に対する最適解の性質について調べ、L1 ソフトマージンサポートベクターマシン (L1 SVM) に対して、 $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$  を満たすデータが、必ずしもサポートベクターになるとは限らない、すなわち  $\alpha_i = 0$  となるデータ  $\mathbf{x}_i$  が存在することを 3 つのデータが  $\mathbf{R}^2$  における長方形の頂点に配置している場合に示した。さらに、特殊なデータ配置について、最適解の非一意性を指摘した (Abe, 2002; 阿部, 2011)。ここで、L1 ソフトマージンサポートベクターマシンとはソフトマージンに基づく主問題において、調整パラメータ  $\lambda$  を  $C/p$ 、 $\xi_i$  を  $\xi_i^p$  と表したときの  $p=1$  の場合をいう。

本節では、4 つのデータが  $\mathbf{R}^n$  における長方形の各頂点に配置している場合と、平行四辺形の各頂点に配置している場合に、 $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$  を満たすデータが、必ずしもサポートベクターになるとは限らないことを理論的に示す。また、この場合に  $\alpha_i = 0$  となるデータがどのような形で現れるかを調べ、最適解の非一意性について考察する。

**定理 1.** 4 つの  $n$  次元データ  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$  が  $\mathbf{R}^n$  における長方形の各頂点に配置しているものとし、

$$\mathbf{x}_1, \mathbf{x}_2 \in G_1 (y_1, y_2 = 1), \quad \mathbf{x}_3, \mathbf{x}_4 \in G_2 (y_3, y_4 = -1)$$

とする。ただし、 $\mathbf{x}_2 - \mathbf{x}_1 = \mathbf{x}_4 - \mathbf{x}_3$  とし、 $y_i (i=1, \dots, 4)$  はラベル変数とする。このとき、双対問題の最適解  $\alpha_i (i=1, \dots, 4)$  は

$$\alpha_1 = \frac{2}{\|\mathbf{x}_3 - \mathbf{x}_1\|^2} - \beta, \quad \alpha_2 = \beta, \quad \alpha_3 = \frac{2}{\|\mathbf{x}_3 - \mathbf{x}_1\|^2} - \beta, \quad \alpha_4 = \beta$$

で与えられる。ただし、 $\beta$  は  $\lambda$  をソフトマージンの調整パラメータとして、 $0 \leq \alpha_i \leq \lambda$  を満たすような実数をとるものとする。

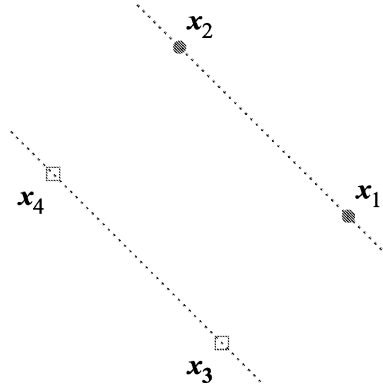


図 1 : 4 つのデータが長方形の各頂点に配置

**証明.** (7) より、双対問題は以下のようなになる。

$$\left\{ \begin{array}{l} \max_{\alpha} L_D(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^4 \alpha_i - \frac{1}{2} (\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 - \alpha_3 \mathbf{x}_3 - \alpha_4 \mathbf{x}_4)^T (\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 - \alpha_3 \mathbf{x}_3 - \alpha_4 \mathbf{x}_4) \right\} \\ \text{制約条件} \quad \alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0, \quad 0 \leq \alpha_i \leq \lambda, \quad i=1, \dots, 4. \end{array} \right.$$

$$A=(a_1\mathbf{x}_1+a_2\mathbf{x}_2-a_3\mathbf{x}_3-a_4\mathbf{x}_4)^T(a_1\mathbf{x}_1+a_2\mathbf{x}_2-a_3\mathbf{x}_3-a_4\mathbf{x}_4)$$

とおくと、制約条件より、

$$\begin{aligned} A &= \{a_2(\mathbf{x}_2-\mathbf{x}_1)-a_3(\mathbf{x}_3-\mathbf{x}_1)-a_4(\mathbf{x}_4-\mathbf{x}_1)\}^T \{a_2(\mathbf{x}_2-\mathbf{x}_1)-a_3(\mathbf{x}_3-\mathbf{x}_1)-a_4(\mathbf{x}_4-\mathbf{x}_1)\} \\ &= a_2^2\|\mathbf{x}_2-\mathbf{x}_1\|^2 + a_3^2\|\mathbf{x}_3-\mathbf{x}_1\|^2 + a_4^2\|\mathbf{x}_4-\mathbf{x}_1\|^2 \\ &\quad - 2a_2a_3(\mathbf{x}_2-\mathbf{x}_1)^T(\mathbf{x}_3-\mathbf{x}_1) - 2a_2a_4(\mathbf{x}_2-\mathbf{x}_1)^T(\mathbf{x}_4-\mathbf{x}_1) + 2a_3a_4(\mathbf{x}_3-\mathbf{x}_1)^T(\mathbf{x}_4-\mathbf{x}_1) \end{aligned}$$

と変形できる。ここで、 $\mathbf{x}_4-\mathbf{x}_1=(\mathbf{x}_4-\mathbf{x}_2)+(\mathbf{x}_2-\mathbf{x}_1)$  と表し、 $\mathbf{x}_4-\mathbf{x}_2=\mathbf{x}_3-\mathbf{x}_1$  と  $(\mathbf{x}_2-\mathbf{x}_1)^T(\mathbf{x}_3-\mathbf{x}_1)=0$  を用いると、

$$A=(a_2-a_4)^2\|\mathbf{x}_2-\mathbf{x}_1\|^2+(a_3+a_4)^2\|\mathbf{x}_3-\mathbf{x}_1\|^2$$

となる。ゆえに、目的関数は

$$\begin{aligned} L_D(\boldsymbol{\alpha}) &= 2(a_3+a_4) - \frac{1}{2}\{(a_2-a_4)^2\|\mathbf{x}_2-\mathbf{x}_1\|^2+(a_3+a_4)^2\|\mathbf{x}_3-\mathbf{x}_1\|^2\} \\ &= -\frac{m}{2}(a_2-a_4)^2 - \frac{n}{2}(a_3+a_4)^2 + 2(a_3+a_4) \end{aligned}$$

と表される。ここで、 $\|\mathbf{x}_2-\mathbf{x}_1\|^2=m$ 、 $\|\mathbf{x}_3-\mathbf{x}_1\|^2=n$  とおいた。

$L_D(\boldsymbol{\alpha})$  を  $a_2$  についての 2 次関数とみなすと、 $m>0$  より  $L_D(\boldsymbol{\alpha})$  は  $a_2=a_4$  のとき最大値をとる。このとき、 $a_3+a_4=l$  とおき、 $L_D(\boldsymbol{\alpha})$  を  $l$  についての 2 次関数とみなすと、

$$L_D(\boldsymbol{\alpha})=-\frac{n}{2}l^2+2l=-\frac{n}{2}\left(l-\frac{2}{n}\right)^2+\frac{2}{n}$$

と表されるから、 $L_D(\boldsymbol{\alpha})$  は  $l=2/n$  のとき最大値をとる。

以上より、 $a_2=a_4=\beta$  とおくと、

$$a_3=l-a_4=\frac{2}{n}-a_4=\frac{2}{\|\mathbf{x}_3-\mathbf{x}_1\|^2}-\beta$$

となる。また制約条件より、

$$a_1=-a_2+a_3+a_4=a_3=\frac{2}{\|\mathbf{x}_3-\mathbf{x}_1\|^2}-\beta$$

を得る。 ■

定理 1 より、長方形配置のデータに対して最適解の非一意性が示された。また、4 つのデータについて、 $\mathbf{x}_1, \mathbf{x}_2$  は超平面  $H_+$  上に、 $\mathbf{x}_3, \mathbf{x}_4$  は超平面  $H_-$  上にあるが、 $\lambda \geq 2/\|\mathbf{x}_3-\mathbf{x}_1\|^2$  のとき、 $\beta=0$  とすると、 $a_1=a_3=2/\|\mathbf{x}_3-\mathbf{x}_1\|^2$ 、 $a_2=a_4=0$  となり、 $\mathbf{x}_2$  と  $\mathbf{x}_4$  はサポートベクターにはならないことがわかる。

定理 1 において、 $\beta$  のとりうる値の範囲は、 $0 \leq a_i \leq \lambda$  ( $i=1, \dots, 4$ ) より、 $0 \leq \beta \leq \lambda$  かつ  $2/\|\mathbf{x}_3-\mathbf{x}_1\|^2 - \lambda \leq \beta \leq 2/\|\mathbf{x}_3-\mathbf{x}_1\|^2$  を満たすから、 $\lambda \geq 2/\|\mathbf{x}_3-\mathbf{x}_1\|^2$  のとき  $0 \leq \beta \leq 2/\|\mathbf{x}_3-\mathbf{x}_1\|^2$ 、 $1/\|\mathbf{x}_3-\mathbf{x}_1\|^2 \leq \lambda \leq 2/\|\mathbf{x}_3-\mathbf{x}_1\|^2$  のとき  $2/\|\mathbf{x}_3-\mathbf{x}_1\|^2 - \lambda \leq \beta \leq \lambda$  となる。

**例 1.**  $\mathbf{x}_1=(1, 0)^T$ 、 $\mathbf{x}_2=(0, 1)^T$ 、 $\mathbf{x}_3=(0, -1)^T$ 、 $\mathbf{x}_4=(-1, 0)^T$  とし、 $\mathbf{x}_1, \mathbf{x}_2 \in G_1$ 、 $\mathbf{x}_3, \mathbf{x}_4 \in G_2$  とする (Abe, 2002; p.94, Example 1)。定理 1 の結果を用いると、 $\|\mathbf{x}_3-\mathbf{x}_1\|=\sqrt{2}$  より、

$$a_1=1-\beta, \quad a_2=\beta, \quad a_3=1-\beta, \quad a_4=\beta$$

となる。ただし、 $\beta$  のとりうる値の範囲は、 $\lambda \geq 1$  のとき  $0 \leq \beta \leq 1$ 、 $1/2 \leq \lambda \leq 1$  のとき  $1 - \lambda \leq \beta \leq \lambda$  である。特に、 $\lambda = 1/2$  のとき  $\beta = 1/2$  となるので、 $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1/2$  となる。

**定理 2.** 4 つの  $n$  次元データ  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$  が  $\mathbf{R}^n$  における平行四辺形の各頂点に配置しているとき、定理 1 と同様の条件の下で、双対問題の最適解  $\alpha_i (i=1, \dots, 4)$  は

$$\begin{aligned}\alpha_1 &= \frac{2\{\|\mathbf{x}_2 - \mathbf{x}_1\|^2 - (\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1)\}}{\|\mathbf{x}_2 - \mathbf{x}_1\|^2\|\mathbf{x}_3 - \mathbf{x}_1\|^2 - \{(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1)\}^2} - \beta, \\ \alpha_2 &= \frac{2(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1)}{\|\mathbf{x}_2 - \mathbf{x}_1\|^2\|\mathbf{x}_3 - \mathbf{x}_1\|^2 - \{(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1)\}^2} + \beta, \\ \alpha_3 &= \frac{2\|\mathbf{x}_2 - \mathbf{x}_1\|^2}{\|\mathbf{x}_2 - \mathbf{x}_1\|^2\|\mathbf{x}_3 - \mathbf{x}_1\|^2 - \{(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1)\}^2} - \beta, \\ \alpha_4 &= \beta\end{aligned}$$

で与えられる。ただし、 $\beta$  は  $\lambda$  をソフトマージンの調整パラメータとして、 $0 \leq \alpha_i \leq \lambda$  を満たす実数とする。

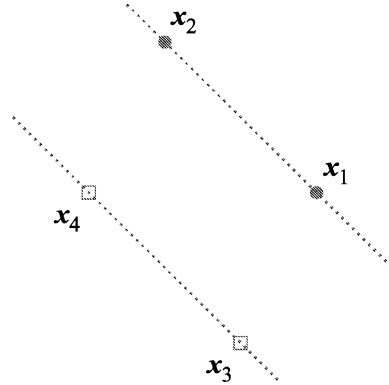


図 2 : 4 つのデータが平行四辺形の各頂点に配置

**証明.** 双対問題

$$\begin{cases} \max_{\alpha} L_D(\alpha) = \max_{\alpha} \left\{ \sum_{i=1}^4 \alpha_i - \frac{1}{2} (a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 - a_3 \mathbf{x}_3 - a_4 \mathbf{x}_4)^T (a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 - a_3 \mathbf{x}_3 - a_4 \mathbf{x}_4) \right\} \\ \text{制約条件} \quad a_1 + a_2 - a_3 - a_4 = 0, \quad 0 \leq \alpha_i \leq \lambda, \quad i=1, \dots, 4 \end{cases}$$

において、

$$A = (a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 - a_3 \mathbf{x}_3 - a_4 \mathbf{x}_4)^T (a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 - a_3 \mathbf{x}_3 - a_4 \mathbf{x}_4)$$

とおくと、制約条件および  $\mathbf{x}_4 - \mathbf{x}_2 = \mathbf{x}_3 - \mathbf{x}_1$  より、

$$\begin{aligned}A &= \{a_2(\mathbf{x}_2 - \mathbf{x}_1) - a_3(\mathbf{x}_3 - \mathbf{x}_1) - a_4(\mathbf{x}_4 - \mathbf{x}_1)\}^T \{a_2(\mathbf{x}_2 - \mathbf{x}_1) - a_3(\mathbf{x}_3 - \mathbf{x}_1) - a_4(\mathbf{x}_4 - \mathbf{x}_1)\} \\ &= (a_2 - a_4)^2 \|\mathbf{x}_2 - \mathbf{x}_1\|^2 + (a_3 + a_4)^2 \|\mathbf{x}_3 - \mathbf{x}_1\|^2 - 2(a_2 - a_4)(a_3 + a_4)(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1)\end{aligned}$$

となる。ゆえに、



$$\begin{aligned}
L_D(\boldsymbol{\alpha}) &= 2(\alpha_3 + \alpha_4) - \frac{1}{2} \{ (\alpha_2 - \alpha_4)^2 \|\mathbf{x}_2 - \mathbf{x}_1\|^2 + (\alpha_3 + \alpha_4)^2 \|\mathbf{x}_3 - \mathbf{x}_1\|^2 - 2(\alpha_2 - \alpha_4)(\alpha_3 + \alpha_4)(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1) \} \\
&= -\frac{m}{2} \left[ \alpha_2 - \frac{1}{m} \{ m\alpha_4 + X(\alpha_3 + \alpha_4) \} \right]^2 + \frac{X^2 - mn}{2m} (\alpha_3 + \alpha_4)^2 + 2(\alpha_3 + \alpha_4)
\end{aligned}$$

と表される。ここで、 $\|\mathbf{x}_2 - \mathbf{x}_1\|^2 = m$ ,  $\|\mathbf{x}_3 - \mathbf{x}_1\|^2 = n$ ,  $(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1) = X$ とおいた。

$L_D(\boldsymbol{\alpha})$  を  $\alpha_2$  についての2次関数とみなすと、 $m > 0$  より、 $L_D(\boldsymbol{\alpha})$  は

$$\alpha_2 = \frac{1}{m} \{ m\alpha_4 + X(\alpha_3 + \alpha_4) \}$$

のとき最大値をとる。このとき、 $\alpha_3 + \alpha_4 = l$  とおくと、

$$\begin{aligned}
L_D(\boldsymbol{\alpha}) &= \frac{X^2 - mn}{2m} l^2 + 2l \\
&= \frac{X^2 - mn}{2m} \left( l + \frac{2m}{X^2 - mn} \right)^2 - \frac{2m}{X^2 - mn}
\end{aligned}$$

と表される。ここで、 $L_D(\boldsymbol{\alpha})$  を  $l$  についての2次関数とみなすと、Cauchy-Schwarzの不等式より、

$$X^2 - mn = \{ (\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1) \}^2 - \|\mathbf{x}_2 - \mathbf{x}_1\|^2 \|\mathbf{x}_3 - \mathbf{x}_1\|^2 < 0$$

となるから、 $L_D(\boldsymbol{\alpha})$  は

$$l = -\frac{2m}{X^2 - mn}$$

のとき最大値をとる。

以上より、 $\alpha_4 = \beta$  とおくと、

$$\begin{aligned}
\alpha_3 &= l - \alpha_4 = \frac{2m}{mn - X^2} - \beta \\
&= \frac{2\|\mathbf{x}_2 - \mathbf{x}_1\|^2}{\|\mathbf{x}_2 - \mathbf{x}_1\|^2 \|\mathbf{x}_3 - \mathbf{x}_1\|^2 - \{ (\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1) \}^2} - \beta, \\
\alpha_2 &= \frac{1}{m} \{ m\alpha_4 + X(\alpha_3 + \alpha_4) \} = \frac{X}{m} l + \alpha_4 \\
&= \frac{2(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1)}{\|\mathbf{x}_2 - \mathbf{x}_1\|^2 \|\mathbf{x}_3 - \mathbf{x}_1\|^2 - \{ (\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1) \}^2} + \beta, \\
\alpha_1 &= -\alpha_2 + \alpha_3 + \alpha_4 = l - \alpha_2 \\
&= \frac{2\{ \|\mathbf{x}_2 - \mathbf{x}_1\|^2 - (\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1) \}}{\|\mathbf{x}_2 - \mathbf{x}_1\|^2 \|\mathbf{x}_3 - \mathbf{x}_1\|^2 - \{ (\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1) \}^2} - \beta
\end{aligned}$$

を得る。 ■

定理2より、平行四辺形配置のデータに対して最適解の非一意性が示された。また、 $\lambda$ の値のとり方によっては、 $\beta = 0$  とすることができ、このとき、 $\mathbf{x}_4$  はサポートベクターにはならないことがわかる。

**例2.**  $\mathbf{x}_1 = (2, 0)^T$ ,  $\mathbf{x}_2 = (0, 1)^T$ ,  $\mathbf{x}_3 = (0, -1)^T$ ,  $\mathbf{x}_4 = (-2, 0)^T$  とし、 $\mathbf{x}_1, \mathbf{x}_2 \in G_1$ ,  $\mathbf{x}_3, \mathbf{x}_4 \in G_2$  とする。定理2の結果を用いると、 $\|\mathbf{x}_2 - \mathbf{x}_1\| = \|\mathbf{x}_3 - \mathbf{x}_1\| = \sqrt{5}$ ,  $(\mathbf{x}_2 - \mathbf{x}_1)^T(\mathbf{x}_3 - \mathbf{x}_1) = 3$  より、

$$\alpha_1 = \frac{1}{4} - \beta, \quad \alpha_2 = \frac{3}{8} + \beta, \quad \alpha_3 = \frac{5}{8} - \beta, \quad \alpha_4 = \beta$$

となる。ただし、 $\beta$ のとりうる値の範囲は、 $\lambda \geq 5/8$ のとき  $0 \leq \beta \leq 1/4$ 、 $1/2 \leq \lambda \leq 5/8$ のとき  $5/8 - \lambda \leq \beta \leq \lambda - 3/8$ である。例2のデータ配置では  $\lambda \geq 5/8$ のとき、 $\beta = 1/4$ あるいは  $\beta = 0$ とすると、 $\alpha_1 = 0$ あるいは  $\alpha_4 = 0$ となるが、 $0 \leq \beta \leq 1/4$ の範囲で  $\beta$ の値をどのようにとっても、 $\alpha_2 = 0$ 、 $\alpha_3 = 0$ とはなりえないことがわかる。

#### 4 Mapleによる解析

前節で述べた例1，例2について，数式処理ソフトウェア Mapleによる解析を試みる。

##### 例1の解析

学習データを  $\mathbf{x}_1 = (1, 0)^T$ ， $\mathbf{x}_2 = (0, 1)^T$ ， $\mathbf{x}_3 = (0, -1)^T$ ， $\mathbf{x}_4 = (-1, 0)^T$ とし， $\mathbf{x}_1, \mathbf{x}_2 \in G_1$ ， $\mathbf{x}_3, \mathbf{x}_4 \in G_2$ とする。このとき，目的関数  $L_D(\boldsymbol{\alpha})$ を最大とする解  $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$ を導出するための Mapleのコードは図3の通りである。

```
>with(LinearAlgebra)
>G1 := Matrix(2, 2, {(1, 1) = 1, (1, 2) = 0, (2, 1) = 0, (2, 2) = 1})
>G2 := Matrix(2, 2, {(1, 1) = 0, (1, 2) = -1, (2, 1) = -1, (2, 2) = 0})
>X := Matrix([[G1], [G2]])
>y := Vector(4, {(1) = 1, (2) = 1, (3) = -1, (4) = -1})
>Y := DiagonalMatrix(y)
>Z := Y.X
>ZT := Transpose(Z)
>H := Matrix(Z.ZT, datatype = float)
>with(Optimization)
>c := Vector([-1, -1, -1, -1], datatype = float)
>Aeq := Matrix(Transpose(y), datatype = float)
>beq := Vector([0], datatype = float)
>bl := 0.
>bu := 1.
>m := QPSolve([c, H], [NoUserValue, NoUserValue, Aeq, beq], [bl, bu],
initialpoint = Vector([0, 0, 0, 0]), output = solutionmodule)
>m:-Results()
```

図3：目的関数の最適解を求める Mapleのコード

Mapleにおける最適化パッケージは **Optimization** であり，**QPSolve** コマンドにより，2次計画問題を解くことができる。初期値は **QPSolve** コマンドの **initialpoint** オプションで設定できる。また，ソフトウェアの調整パラメータ  $\lambda$  は **bu** により設定する。ここでは  $\lambda = 1$ とした。

双対問題において非一意な最適解  $(\hat{a}_1, \dots, \hat{a}_4)^T$ が存在するため，Mapleでは結果を表示するコマンド **m:-Results()**の実行後，警告メッセージ

**Warning, necessary conditions met but sufficient conditions not satisfied**

が表示される。

初期値を変更すると、反復計算によって得られる最適解は以下のように収束した。このときの  $L_D(\alpha)$  の値は不変であり、その最大値  $L_D(\hat{\alpha})$  は  $-1$  である。得られた解はいずれも例 1 の結果を満たしている。

初期値	(0, 0, 0, 0)	(1, 1, 1, 1)	(0.3, 0.6, 0.4, 0.3)
最適解 $\hat{\alpha}$	(1, 0, 1, 0)	(0, 1, 0, 1)	(0.5, 0.5, 0.5, 0.5)
$L_D(\hat{\alpha})$	-1		

## 例 2 の解析

学習データを  $\mathbf{x}_1=(2, 0)^T$ ,  $\mathbf{x}_2=(0, 1)^T$ ,  $\mathbf{x}_3=(0, -1)^T$ ,  $\mathbf{x}_4=(-2, 0)^T$  とし,  $\mathbf{x}_1, \mathbf{x}_2 \in G_1$ ,  $\mathbf{x}_3, \mathbf{x}_4 \in G_2$  とする。このとき,  $L_D(\alpha)$  を最大とする  $\alpha = \hat{\alpha}$  は上記コードで,

```
>G1 := Matrix(2, 2, {(1, 1) = 2, (1, 2) = 0, (2, 1) = 0, (2, 2) = 1})
>G2 := Matrix(2, 2, {(1, 1) = 0, (1, 2) = -1, (2, 1) = -2, (2, 2) = 0})
```

として実行することにより導出できる。この場合も、例 1 と同様の警告メッセージが表示される。初期値を変更し、解の収束状況を調べた結果、以下のようになった。得られた解はいずれも例 2 の結果を満たしている。

初期値	(0, 0, 0, 0)	(1, 1, 1, 1)	(0.3, 0.6, 0.4, 0.3)
最適解 $\hat{\alpha}$	(0.25, 0.375, 0.625, 0)	(0, 0.625, 0.375, 0.25)	(0.03125, 0.59375, 0.40625, 0.21875)
$L_D(\hat{\alpha})$	-0.625		

## 5 おわりに

本論文では、サポートベクターマシンの最適解について、4 つのデータが  $\mathbf{R}^n$  における長方形の各頂点に配置している場合と、平行四辺形の各頂点に配置している場合に、 $y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$  を満たすデータが、必ずしもサポートベクターになるとは限らないことを理論的に示し、 $\alpha_i = 0$  となるデータがどのような形で現れるかを検討した。また、これらの場合に、最適解の非一意性について考察した。さらに、非一意な解が存在するデータについて、数式処理ソフトウェアによる解析を試みた。この結果、初期値の設定により、解が異なる値に収束することがわかった。今後の課題として、最適解の非一意性とソフトマージンの調整パラメータ  $\lambda$  の関係を調べることが挙げられる。

## 参考文献

- Abe, S. (2002). Analysis of support vector machines. *Neural Networks for Signal Processing XII. Proceedings of the 2002 IEEE Signal Processing Society Workshop*, 89-98.
- 阿部重夫 (2011). パターン認識のためのサポートベクトルマシン入門, 森北出版.
- 小西貞則 (2010). 多変量解析入門 (第 8 章), 岩波書店.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley.

# Properties of optimal solutions in support vector machines

by

Sakie ISHIKAWA \*

Nagatomo NAKAMURA \*\*

Takahiro TSUCHIYA \*\*\*

## Abstract

This paper discusses the properties of optimal solutions for the objective function of a quadratic programming problem and the non-uniqueness of the solutions in support vector machines. Even if there exists a training data on two hyperplanes constructing the optimal margin, it is known that the data may not be support vector. In other words, there may be cases where the optimal solutions are non-unique. We theoretically examine the non-uniqueness of the solutions through two arrangements of data points in  $\mathbf{R}^n$ . Further, by using the computer algebra software Maple, we apply support vector classification for the above data and solve the optimization problems.

**Keywords:** Support Vectors, Optimal Separating Hyperplane, Optimization Problem.

---

\*Josai University

\*\*Sapporo Gakuin University, nagatomo@sgu.ac.jp

\*\*\*Josai University, takahiro@math.josai.ac.jp