

# サポート・ベクター・マシンの幾何学的解釈

矢島 安敏

サポート・ベクター・マシン (Support Vector Machine, SVM) は、分類問題に対する非常に優れた手法として、近年さまざまな分野で活用されている。本稿では2クラス分類を行う SVM を考え、その幾何学的な解釈を試みよう。

$N$ 次元空間  $\mathbf{R}^N$  に  $M$  個のデータ (点) が与えられており、各データは「正例」および「負例」いずれか一方のクラスに属しているとする。以降では、正例 (負例) のデータ数を  $M_+$  ( $M_-$ ) とする。また、 $A_+$  を  $M_+$  行  $N$  列の行列とし、正例のクラスに属するデータを行ベクトルにもつもの、また同様に、 $A_-$  を  $M_-$  行  $N$  列の行列とし、負例のデータに対応したものとする。SVM では、これらのデータに基づいて線形の判別関数を構築する。

まず始めに、図1に示すようにデータが黒丸と白丸の2つのクラスに完全に分離されている場合を考える。この例の場合、2つのクラスを分ける線形関数 (直線) は(a), (b), (c)のように無数に考えられる。しかし、(a)や(c)の直線をクラスの判別境界としたのでは、直線付近のデータが少し変化しただけで境界を越えてしまい、あまり良い境界とは考えられない。SVM では、できるだけ両方のクラスのデータから離れた(b)の判別関数が良いものと考えられている。

このような判別関数を求める問題は、最適化問題として定式化される。Vapnik 等[2]によるオリジナルのSVMでは、法線ベクトル  $\mathbf{w} \in \mathbf{R}^N$  と実数  $b$  で定められる2本の平行な直線

$$d(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = \pm 1 \quad (1)$$

を考え、図2のようにクラスを分離する。すなわち、 $\mathbf{w}$ ,  $b$  は次の不等式系:

$$A_+ \mathbf{w} + e\mathbf{b} \geq e, \quad A_- \mathbf{w} + e\mathbf{b} \leq -e, \quad (2)$$

を満たすものとする。ただし、 $e$  は要素がすべて1のベクトルとし、その次元は前後の関係から定められる

ものとする。このとき、式(1)の平行な2直線の間隔は  $2/\|\mathbf{w}\|$  なので、 $\|\mathbf{w}\|^2$  を不等式系(2)を制約として最小化する次の2次計画問題:

$$\begin{array}{ll} \text{最小化} & \|\mathbf{w}\|^2 \\ \text{制約} & A_+ \mathbf{w} + e\mathbf{b} \geq e, \\ & A_- \mathbf{w} + e\mathbf{b} \leq -e \end{array} \quad (3)$$

を解けば、最も間隔が広い2直線が得られる。この問題の最適解を  $\mathbf{w}^*$ ,  $b^*$  とすれば、2直線の中央に位置する次の直線:

$$d^*(\mathbf{x}) = \mathbf{w}^{*T} \mathbf{x} + b^* = 0$$

が求めたい判別関数となる。このように定式化されるSVMは、ハードマージンSVMと呼ばれている。

ここで、新たに実数  $\rho$  を導入し、式(1)の2直線をより一般的に、

$$\mathbf{w}^T \mathbf{x} + b = \rho, \quad \mathbf{w}^T \mathbf{x} + b = -\rho$$

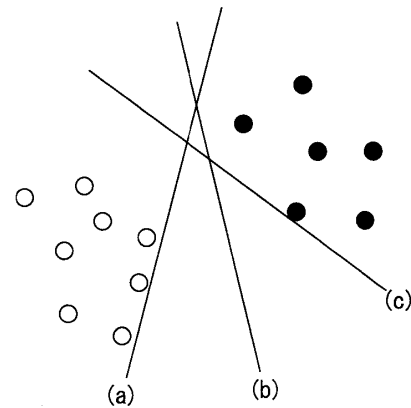


図1 さまざまな線形判別関数

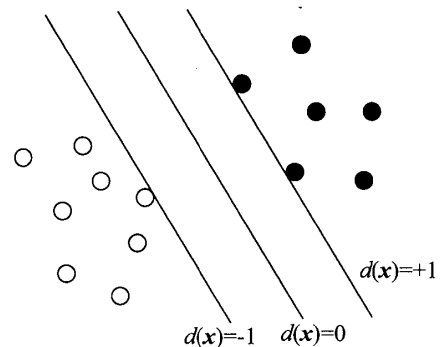


図2 間隔が最大となる2直線

やじま やすとし

東京工業大学 大学院社会理工学研究科

〒152-8552 目黒区大岡山 2-12-1

と記述しよう。この場合、2直線の間隔は $2\rho/\|w\|$ と表される。そこで、間隔の最大化を、 $\|w\|^2$ と $-\rho$ の和の最小化と考え、次のように定式化しよう。

$$\begin{cases} \text{最小化} & \frac{1}{2}\|w\|^2 - 2\rho \\ \text{制約} & A_+w + eb \geq e\rho, \\ & A_-w + eb \leq -e\rho. \end{cases} \quad (4)$$

このあと示す幾何学的な解釈の都合上、 $\|w\|^2$ と $-\rho$ の項には上のように係数を付与した。

ここで、問題(4)の Wolfe 双対問題[3]を考えると、図2とは異なる幾何学的な解釈が可能である。まず、 $u \in \mathbf{R}^{M_+}$ 、 $v \in \mathbf{R}^{M_-}$ を問題(4)の不等式制約に対応したラグランジュ乗数とし、次のようにラグランジュ関数を定める。

$$\begin{aligned} L(w, b, \rho; u, v) = & \frac{1}{2}\|w\|^2 - 2\rho \\ & + u^T(-A_+w - eb + e\rho) \\ & + v^T(A_-w + eb + e\rho) \end{aligned}$$

関数 $L$ を $w$ 、 $b$ 、 $\rho$ でそれぞれ微分して0とすれば、次の方程式：

$$\frac{\partial L}{\partial w} = w - A_+^T u + A_-^T v = 0, \quad (5)$$

$$\frac{\partial L}{\partial b} = -e^T u + e^T v = 0, \quad (6)$$

$$\frac{\partial L}{\partial \rho} = -2 + e^T u + e^T v = 0 \quad (7)$$

が得られ、これらが双対問題の制約式となる。特に、式(5)は

$$w = A_+^T u - A_-^T v \quad (8)$$

となり、また、式(6)と(7)は、次の同値な関係：

$$e^T u = 1, \quad e^T v = 1 \quad (9)$$

に書き換えることができる。そこで、式(8)と(9)をラグランジュ関数に代入すれば、変数 $w$ 、 $b$ 、 $\rho$ が消去でき、

$$L(w, b, \rho; u, v) = -\frac{1}{2}\|A_+^T u - A_-^T v\|^2$$

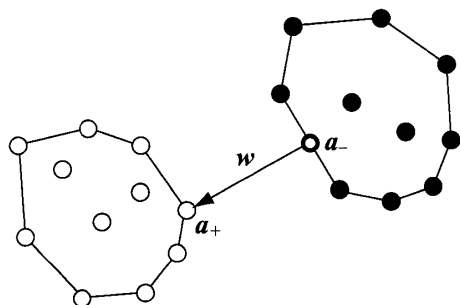


図3 双対問題の幾何学的な解釈

を得る。したがって、問題(4)の双対問題は、次のように定式化される。

$$\begin{cases} \text{最小化} & \frac{1}{2}\|A_+^T u - A_-^T v\|^2 \\ \text{制約} & e^T u = 1, u \geq 0, \\ & e^T v = 1, v \geq 0. \end{cases} \quad (10)$$

次に、この双対問題を幾何学的に考えてみよう。ベクトル $u$ は、要素が非負で和が1となるベクトルであることから、 $A_+^T u$ は「正例」のクラスに属する点の凸包の点となる。ベクトル $v$ についても同様で、 $A_-^T v$ は「負例」の凸包の点である。ゆえに、双対問題(10)は、それぞれのクラスの凸包の点で、もっとも距離の近い2点を求める問題であると考えられる。問題(10)の最適解を $u^*$ 、 $v^*$ とし、それぞれに対応する2点を $a_+ = A_+^T u^*$ および $a_- = A_-^T v^*$ とすれば、最も近い点は図3に示す通りである。

問題(4)で得られる最適な法線ベクトルを $w^*$ とすれば、式(8)より

$$w^* = A_+^T u^* - A_-^T v^* = a_+ - a_-$$

を満たす。ゆえに、求めたい判別平面の法線ベクトルは、最も近い2点の差ベクトルと一致することがわかる。さらに、相補性の条件より、最適解は

$$u^{*T}(A_+w^* + eb^* - e\rho^*) = 0,$$

$$v^{*T}(A_-w^* + eb^* + e\rho^*) = 0,$$

を満たす。この2式を加えると

$$u^{*T}(A_+w^* + eb^*) + v^{*T}(A_-w^* + eb^*) = 0,$$

$$w^{*T}\left(\frac{a_+ + a_-}{2}\right) + b^* = 0,$$

となることより、判別の境界は $a_+$ と $a_-$ の中点を通る直線となる。

次にデータが線形関数では分離できない場合を考えよう。ここでは、問題(4)にスラック変数ベクトル $\xi_+ \in \mathbf{R}^{M_+}$ 、 $\xi_- \in \mathbf{R}^{M_-}$ を導入し、線形関数で分離されない場合のエラー項とする。そこで、次の最適化問題：

$$\begin{cases} \text{最小化} & \frac{1}{2}\|w\|^2 - 2\rho + C(e^T \xi_+ + e^T \xi_-) \\ \text{制約} & A_+w + eb + \xi_+ \geq e\rho, \xi_+ \geq 0, \\ & A_-w + eb - \xi_- \leq -e\rho, \xi_- \geq 0, \end{cases} \quad (11)$$

を考える。ただし、 $C$ は正のパラメータで、2直線の間隔の最大化とエラーの最小化とのバランスを決める役割を果たしている。これは、ソフトマージンSVMと呼ばれるものである。なお、 $\nu$ -SVM[4]と呼ばれる定式化では、目的関数の係数とパラメータの付け方を変え、

$$\frac{1}{2}\|w\|^2 - C\rho + \frac{1}{M}(e^T\xi_+ + e^T\xi_-)$$

を最小化しているが、本質的な考え方は問題(11)と変わらない。

問題(11)では、ソフトマージン SVM の幾何学的な解釈や、パラメータ  $C$  の役割を直感的に理解することは難しい。しかし、その双対問題を用いると幾何学的な解釈が可能となる。詳しい式変形は省略するが、問題(11)の Wolfe 双対問題は

$$\begin{cases} \text{最小化} & \frac{1}{2}\|A^T u - A^T v\|^2 \\ \text{制約} & e^T u = 1, C e \geq u \geq 0, \\ & e^T v = 1, C e \geq v \geq 0 \end{cases} \quad (12)$$

となる。問題(10)の場合と同様に考えれば、この問題は、次の2つの集合：

$$X_+(C) \equiv \{x | x = A^T u, e^T u = 1, C e \geq u \geq 0\},$$

$$X_-(C) \equiv \{x | x = A^T v, e^T v = 1, C e \geq v \geq 0\},$$

に属する点の中で最も近いものを求める問題と考えることができる。

ここで、集合  $X_+(C)$  や  $X_-(C)$  は reduced convex hull (RC-Hull) [1] と呼ばれている。例えば、図4に示した5つの点  $A_1$  から  $A_5$  を行ベクトルとする5行4列の行列を  $A$ 、また  $u$  を5次元のベクトルとし、

$$X(C) \equiv \{x \in \mathbb{R}^2 | x = A^T u, e^T u = 1, C e \geq u \geq 0\}$$

を定める。このとき、 $C=1, 0.5$  および  $0.25$  それぞれに対応する RC-Hull を描くと図4中の多角形が得られる。特に、 $C=0.2$  に対応する RC-Hull は5点の重心1点となり、さらに  $C$  を  $0.2$  より小さくすれば  $X(C)$  は空集合になってしまう。

問題(12)の場合では、 $X_+(C)(X_-(C))$  は  $M_+(M_-)$  個の点の RC-Hull であるので、パラメータ  $C$  は

$$C \geq \underline{C} \equiv \max\{1/M_+, 1/M_-\}$$

となる必要がある。一方、 $C$  が大きく  $X_+(C)$  と  $X_-(C)$  に共通部分が存在すれば、問題(12)の最適解  $u^*, v^*$  は、

$$A^T u^* = A^T v^* \in X_+(C) \cap X_-(C)$$

となることから、対応した主問題(11)の最適解も

$$w^* = A^T u^* - A^T v^* = 0$$

となり、判別関数を求めることができない。特に、正例と負例の数が同数 ( $M_+ = M_- = M/2$ ) で、かつ重心が等しい場合には、 $C \geq \underline{C} = 2/M$  であるすべての  $C$  で、 $X_+(C)$  と  $X_-(C)$  はその重心を共通部分にもつため、主問題(11)の最適解は  $w^* = 0$  となってしまう、判別関数を得ることはできない。

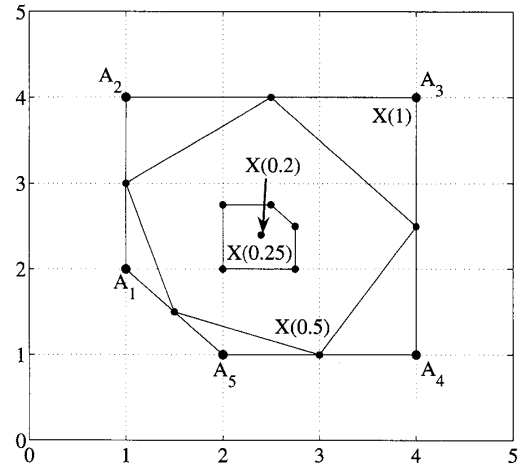


図4 RC-Hull

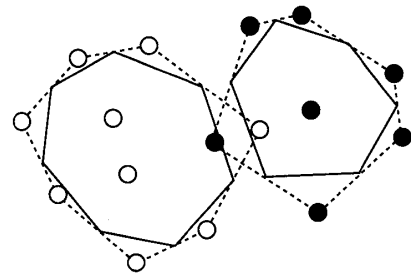


図5  $C=0.5$  の場合の2つの RC-Hull

そこで、パラメータ  $C$  は  $\underline{C}$  より大きく、かつ図5のように  $X_+(C)$  と  $X_-(C)$  が共通部分をもたないと仮定しよう。この場合、問題(12)の最適解  $u^*, v^*$  に対応する2点を

$$a_+ = A^T u^* \in X_+(C), \quad a_- = A^T v^* \in X_-(C)$$

とすれば、判別平面の法線ベクトルは、ハードマージンの場合と同様に、 $w^* = a_+ - a_-$  と求められる。

一方、ハードマージンの場合では、判別平面は2点  $a_+, a_-$  の中点を通るものであった。しかし、ソフトマージンの場合では、必ずしも中点を通らない。すなわち、次の相補条件：

$$u^{*T}(A_+ w^* + e b^* + \xi^* - e \rho^*) = 0,$$

$$v^{*T}(A_- w^* + e b^* - \xi^* + e \rho^*) = 0,$$

を加えると、

$$w^{*T} \left( \frac{a_+ + a_-}{2} \right) + b^* = \frac{C(e^T \xi^* - e^T \xi^*)}{2}$$

となることより、エラーの和が等しい、すなわち  $e^T \xi^* = e^T \xi^*$  の場合に限り、判別平面は2点  $a_+, a_-$  の中点を通ることとなる。

#### 参考文献

- [1] K. P. Bennett and E. J. Bredensteiner: Duality and

- geometry in SVM classifiers. In *Proc. 17th International Conf. on Machine Learning*, pp.57-64, 2000.
- [2] N. Cristianini and J. Shawe-Taylor eds: *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, U. K., 2000.
- [3] O. L. Mangasarian: *Nonlinear Programming*, Vol. 10 of *Classics in Applied Mathematics*.SIAM, Philadelphia, PA, 1994.
- [4] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett: New support vector algorithms. *Neural Computation*, 12 : 1207-1245, 2000.