# ４．サポートベクターマシン
# ４．１．サポートベクターマシンの最適化

08T4067L　横田翔

# 4.1 サポートベクターマシンの最適化

## 4.1.1 双対問題とサポートベクターマシン

### a. サポートベクターマシンの双対問題

$$\min_{w_i, b, \xi_i} \frac{1}{2}\sum_{i,j=1}^{N} w_i w_j K_{ij} + C\sum_{i=1}^{N} \xi_i$$

$$subject \quad to \quad \begin{cases} Y_i\left(\sum_{j=1}^{N} K_{ij} w_j + b\right) \geq 1 - \xi_i & (\forall i) \\ \xi_i \geq 0 \quad (\forall i) \end{cases}$$

ここで、$K_{ij} = k(X_i, X_j)$ と表す。

これに対する双対問題は次のように与えられる

# サポートベクターマシンの双対問題

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j Y_i Y_j K_{ij}$$

$$subject \quad to \quad \begin{cases} 0 \le \alpha_i \le C \quad (\forall i) \\ \sum_{i=1}^{N} \alpha_i Y_i = 0 \end{cases} \qquad (4.1)$$

主問題から双対問題を導く。

SVMの主問題に対するLagrange双対問題は

$$g(\alpha, \beta) = \min_{\omega, b, \xi} L(w, b, \xi, \alpha, \beta) \qquad （4.2）$$

ただし $\alpha_i \geq 0, \beta_i \geq 0$

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2}\sum_{i.j=1}^{N} w_i w_j K_{ij} + C\sum_{i=1}^{N}\xi_i$$
$$+\sum_{i=1}^{N}\alpha_i \left\{ 1 - Y_i(\sum_{j=1}^{N} w_j K_{ij} + b) - \xi_i \right\} + \sum_{i=1}^{N}\beta_i(-\xi_i)$$

により定義される。ここで式(4.2)は制約なしの最小化問題であるので、最適解 $(w^*, b^*, \xi^*)$ は微分を実行し

$$\nabla_w : \sum_{j=1}^{N} K_{ij} w_j^* - \sum_{j=1}^{N} \alpha_j Y_j K_{ij} = 0 \quad (\forall i)$$

$$\nabla_b : \sum_{j=1}^{N} \alpha_j Y_j = 0$$

$$\nabla_\xi : C - \alpha_i - \beta_i = 0 \quad (\forall i)$$

を満たす。

これらの関係式を合わせると

$$\frac{1}{2}\sum_{i,j=1}^{N} w_i^* w_j^* K_{ij} = \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j Y_i Y_j K_{ij},$$

$$\sum_{i=1}^{N} \alpha_i \{1 - Y_i(\sum_{j=1}^{N} K_{ij} w_j^* + b^*)\} = \sum_{i=1}^{N} \alpha_i - \sum_{i,j=1}^{N} \alpha_i \alpha_j Y_i Y_j K_{ij}$$

が得られ、$C - \alpha_i - \beta_i = 0$と合わせると、

Lagrange双対問題は

$$g(\alpha,\beta) = L(w^*, b^*, \xi^*, \alpha, \beta,) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j Y_i Y_j K_{ij}$$

と書き直される。

# b.KKT条件とサポートベクター

KKT条件ー微分可能な目的関数を持つ凸最適化問題の解の必要十分条件を与える。

ここで

$$h^*(x) = \sum_{i=1}^{N} w_i^* k(x, X_i) + b^*$$

とおく。

# SVMのKKT条件

(1) $1 - Y_i h^*(X_i) - \xi_i^* \le 0,\ (\forall i)$

(2) $-\xi_i^* \le 0,\ (\forall i)$

(3) $\alpha_i^* \ge 0,\ (\forall i)$

(4) $\beta_i^* \ge 0,\ (\forall i)$

(5) $\alpha_i^*(1 - Y_i h^*(X_i) - \xi_i^*) = 0,\ (\forall i)$

(6) $\beta_i^* \xi_i^* = 0\ \ (\forall i)$

(7) $\nabla_w : \sum_{j=1}^{N} K_{ij} w_j^* - \sum_{j=1}^{N} \alpha_j Y_j K_{ij} = 0$

$$\nabla_b : \sum_{j=1}^{N} \alpha_j Y_j = 0$$

$$\nabla_\xi : C - \alpha_i - \beta_i = 0\ \ (\forall i)$$

まずKKT条件（7）の第１式を用いると、解は

$$h^*(x) = \sum_{i=1}^{N} \alpha_i^* Y_i k(x, X_i) + b^*$$

の形で書ける。また、（7）の第3式より

$\beta_i^* = C - \alpha_i^*$ であるので、（5）、（6）の式は、

$$\alpha_i^*(1 - Y_i h^*(X_i) - \xi_i^*) = 0 \quad (\forall i) \tag{4.4}$$

$$(C - \alpha_i^*)\xi_i^* = 0 \quad (\forall i) \tag{4.5}$$

と表すことができる。（3）、（4）より $0 \le \alpha_i^* \le C$ であるが各 $i = 1, \cdots, N$ に対し、$\alpha_i^*$ の値に従ってデータと識別平面の位置関係は3つに分かれる。

【$\alpha_i^* = 0$ の場合】

式（4．5）より$\xi_i^* = 0$ で、KKT条件（1）から

$$Y_i h^*(X_i) \geq 1.$$

これはデータ点 $X_i$が $h^*$によって十分識別されている

【 $0 < \alpha_i^* < C$ の場合】

上式同様 $\xi_i^* = 0$で、さらに式（4．4）より

$$Y_i h^*(X_i) = 1.$$

すなわち $X_i$はマージンを定める超平面上に存在する。

【$\alpha_i^* = C$の場合】

式（4．4）より

$$Y_i h^*(X_i) \leq 1.$$

すなわち $X_i$はマージンを定める超平面の内側ないし
誤識別領域に存在する。

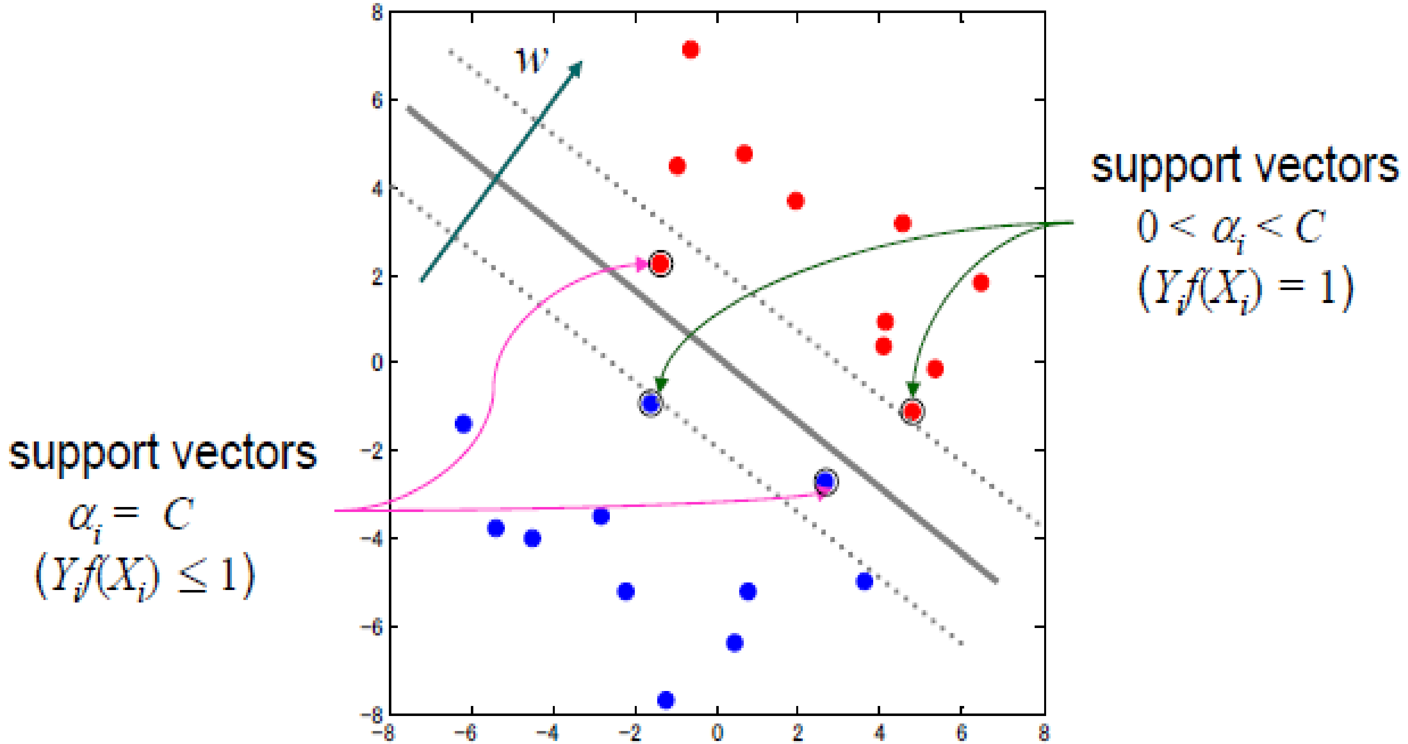


図4.1　サポートベクター

スパース表現－最適解がデータの一部だけで表現されること

これはカーネルリッジ回帰と対比して考えるとわかりやすい。
SVMとリッジ回帰の識別関数はともに

$$h(x) = \sum_i \alpha_i Y_i k(x, X_i) + b$$

と表現される。

カーネルリッジ回帰では一般に最適なhの表現にすべてのデータが必要であるが、SVMではサポートベクターに対応するデータ点のみで表現される。

式(4.1)の双対問題についてbを求める必要がある。

$0<\alpha_i^*<C$ を満たす任意のiに対して

$$Y_i(\sum_j k(X_i,X_j)Y_j\alpha_j^*+b^*=1$$

が成り立つので、両辺に $Y_i$ を掛ければ

$$b^*=Y_i-\sum_j \alpha_j^* k(X_i,X_j)Y_j$$

である。そこで $0<\alpha_i^*<C$ なるすべてのiに対して右辺の値を求め、それらの平均により $b^*$ を定めるのが一つの方法である。

# 4.1.2　効率的な解法

【a. Sequential Minimal Optimization (SMO)】

・２つの変数 $(\alpha_i, \alpha_j)$を選び、他の変数を固定して２変数の２次計画問題を解く手続きを繰り返す方法である。

・各ステップにおいて２個の変数を選ぶためにはKKT条件に基づく基準が用いられる。

# 【SVMに対するKKT条件】

（i） $\sum_{i=1}^{N} Y_i \alpha_i^* = 0,$

（ii） $\begin{cases} \alpha_i^* = 0 \text{かつ} Y_i f^*(X_i) \geq 1 & （このとき\ \xi_i^* = 0\ ） \\ 0 < \alpha_i^* < C \text{かつ} Y_i f^*(X_i) = 1 & （このとき\ \xi_i^* = 0\ ） \\ \alpha_i^* = C \text{かつ} Y_i f^*(X_i) \leq 1 & \end{cases}$

（このとき $\xi_i^* = 1 - Y_i f^*(X_i)$ ）

$$(4.6)$$

- まず各αに対し（ii）の条件が満たされるかどうか調べる。
- 少なくとも一方が条件を満たさない組（i,j）を次の最適化変数として選ぶ
- 次に、選ばれた２変数に対してSVMの目的関数を最適化する。

ここでは2変数を簡単に（i,j）＝（1,2）とする。
このときの制約条件は

$$\alpha_1 + s_{12}\alpha_2 = \gamma, \qquad 0 \le \alpha_1, \alpha_2 \le C$$

と書ける。ここで $s_{12} = Y_1 Y_2$ と $\gamma = \sum_{l \ge 3} Y_1 Y_l \alpha_l$ は2変数の最適化においては定数である

さらに目的関数は

$$\alpha_1+\alpha_2-\frac{1}{2}\alpha_1^2K_{11}-\frac{1}{2}\alpha_2^2K_{22}-s_{12}\alpha_1\alpha_2K_{12}$$

$$-Y_1\alpha_1\sum_{j\geq 3}Y_j\alpha_jK_{1j}-Y_2\alpha_2\sum_{j\geq 3}Y_j\alpha_jK_{2j}+(\alpha_1,\alpha_2\text{によらない項})$$

と表現できる。

$(\alpha_1,\alpha_2)$に関する最適化は、区間制約のもとで1変数の二次関数を最小化する問題に還元される。

このSMOによってSVMは数千以上の大きいデータ数の問題にも適用可能。

# b.サポートベクターマシンの KKT条件再論

式（4.6）の条件を導く。

KKT条件（7）の第１，３式を用いると

$$\beta_i^* = C - \alpha_i \ (\forall i) \qquad \sum_{j=1}^{N} K_{ij} w_j^* = \sum_{j=1}^{N} \alpha_j Y_j K_{ij} \ (\forall i)$$

を得るこれにより $\beta_i^*, w_i^*$ を消去すると、条件（４）、（６）は

$\alpha_i^* \leq C, \ \xi_i^*(C - \alpha_i^*) = 0 \ (\forall i)$ に同値である。

よってKKT条件は以下の通りとなる。

$(a)\quad 1-Y_i h^*(X_i)-\xi_i^* \le 0,$

$(b)\quad \xi_i^* \ge 0,$

$(c)\quad 0 \le \alpha_i^* \le C,$

$(d)\quad \alpha_i^*(1-Y_i f^*(X_i)-\xi_i^*)=0,$

$(e)\quad \xi_i^*(C-\alpha_i^*)=0$

$(f)\quad \sum_{i=1}^{N} Y_i \alpha_i^* = 0$

さらに場合分けによって $\xi_i^*$ を消去できる。

・$\alpha_i^*=0$ の場合：(e)より$\xi_i^*=0$. したがって(a)から $Y_i f^*(X_i)\ge 1.$

・$0<\alpha_i^*<C$ の場合：(e)より$\xi_i^*=0$. ゆえに(d)から $Y_i f^*(X_i)=1.$

・$\alpha_i^*=C$ の場合：(b),(d)により $\xi_i^*=1-Y_i f^*(X_i)\ge 0.$

以上により、KKT条件は式（４．６）の条件と同値である。