

1 クラス SVM と近傍サポートによる領域判別

高畠 泰斗, 香田 正人

領域判別問題 (domain description problem) とその判別手法である 1 クラス SVM (OC-SVM) について紹介する。特にガウシアンカーネルを用いた OC-SVM について、その性質を整理し、また密度推定や近傍サポート (Support Neighbor) との関係性を明らかにすることで、どのように領域判別を実現しているのかを考察する。

キーワード: 領域判別, 1 クラス SVM, 近傍サポート, ガウシアンカーネル

1. はじめに

パターン認識においては、データ分布の密度関数を知ることができれば全ての問題は解決する。しかしながら有限の訓練パターンから密度関数を推定する問題は簡単ではない。そのような中で、密度関数の推定問題の緩和問題とも言える、高密度領域を推定する領域判別 (domain description) が外れ値検出などの様々な分野で利用されている。本稿ではこの領域判別問題および、その解法の 1 つである 1 クラス SVM (One-Class Support Vector Machine, OC-SVM) について考えていく。以下、第 2 節では領域判別の定義を述べ、第 3 節で OC-SVM について概要および性質などを整理し、最後に近傍サポートとして OC-SVM を理解することを考える。そして第 4 節では領域判別の応用について述べる。

2. 領域判別

領域判別問題は与えられた訓練パターンから高密度領域を推定する問題である。この高密度領域について以下に概説する。次のような確率空間 (X, \mathcal{C}, P) について考える。変数 $\mathbf{x}_i (i=1, \dots, n)$ を、分布 P を持つ集合 $X \in \mathbb{R}^d (d \geq 1)$ に含まれる、互いに独立に同じ分布に従う確率変数であるとする。また、 \mathcal{C} を X の可測部分集合のクラスとし、 λ を \mathcal{C} 上で定義された実関数であるとする。このとき、次の集合 $C_p(\eta)$ をレベル $\eta (\geq 0)$ における密度等高クラスタと呼ぶ[8]。

$$C_p(\eta) = \{\mathbf{x} | p(\mathbf{x}) \geq \eta\}. \quad (1)$$

分位関数 (quantile function) U を定義し、これを用いて密度関数の高密度領域について説明を行う。分位関数は以下のように定義される[3]。

$$U(\beta) = \inf\{\lambda(C) | P(C) \geq \beta, C \in \mathcal{C}\}, 0 < \beta \leq 1. \quad (2)$$

分布 P が経験分布, $P_n = \frac{1}{n} \sum_{i=1}^n I(\mathbf{x}_i \in C)$ であるときには経験分位関数が得られる。ここで、 I は指示関数 (indicator function) であり、 $I(\cdot)$ の中身が真のときに 1、偽のときに 0 を返す関数である。 P や P_n について、下限を達成する $C \in \mathcal{C}$ が存在したとき、これを $C(\beta)$ や $C_n(\beta)$ のように記述する。 λ は一般的に \mathbb{R}^d 上のルベグ測度が用いられるが、この場合、 $C(\beta)$ は少なくともデータの β の割合を含む集合 $C \in \mathcal{C}$ の中で“体積”が最小な集合を、分位関数 $U(\beta)$ はその体積を意味する。 $P(C_p(\eta)) = P(C(\beta)) = \beta$ であるとき、 $C(\beta)$ と $C_p(\eta)$ は一致する。このとき、 $C(\beta)$ は少なくともデータの β の割合を含む集合 $C \in \mathcal{C}$ の中で、最も密度が高い領域であると考えられる。この領域を一般に β 高密度領域と呼ぶ。

例を挙げる。図 1 は 1 次元正規分布 $N(0, 1)$ の密度関数である。このとき、 $\beta=0.5$ とする高密度領域 $C(0.5)$ は閉区間 $[-0.67, 0.67]$ になり、分位関数は式(2)より区間長として $U(0.5) = 1.34$ となる。 $p(-0.67) = p(0.67) = 0.32$ であるから、 $\beta=0.5$ とする高密度領域 $C(0.5)$ はレベル $\eta=0.32$ における密度等高クラスタ $C_p(0.32)$ と一致する。

3. 1 クラス SVM

3.1 1 クラス SVM の概要

高密度領域を推定する手法として、1 クラス SVM (OC-SVM) [8] について概説する。OC-SVM では、適切なカーネルに対応する写像 ϕ によって写された

たかばたけ たいと, こうだ まさと
筑波大学 大学院システム情報工学研究科
〒305-8573 つくば市天王台 1-1-1

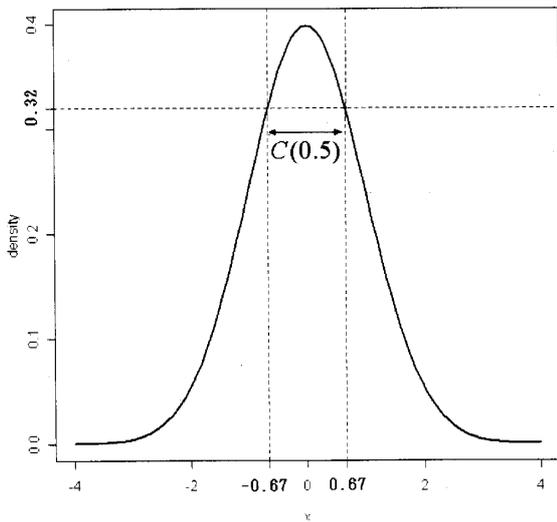


図1 正規分布における例

先の特徴空間上で、超平面： $\langle \mathbf{w}, \phi(\mathbf{x}) \rangle - \rho = 0$ ($\rho \in R$) によって領域判別を行う。ここで、 $\langle \cdot \rangle$ は内積を示す。多くの訓練パターンが判別超平面を挟んで特徴空間上の原点との反対側にあることを維持しながら超平面と原点の距離（マージン）を大きくすることで高密度領域の推定を目指す（詳しくは3.3節で記述する）。未知のパターン \mathbf{x} が高密度領域に含まれるか否かは判別関数

$$f(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \phi(\mathbf{x}) \rangle - \rho) \quad (3)$$

によって評価し、 $f(\mathbf{x}) = +1$ のとき高密度領域に含まれると判別する。このとき、集合 $\{\mathbf{x} | f(\mathbf{x}) = +1\}$ が高密度領域の推定量となる。ここで、符号関数 $\text{sign}(\cdot)$ は中身が非負のとき1、それ以外のとき-1を返す関数とする。

訓練パターンを $\mathbf{x}_i (i=1, \dots, n)$ とすると、マージンの最大化とペナルティ項を調整するパラメータ $\nu \in (0, 1]$ を用いて、上記の問題は以下の2次計画問題として定式化できる。

$$\begin{aligned} \min_{\mathbf{w}, \rho, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 - \rho + \frac{1}{\nu n} \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & (\mathbf{w}, \phi(\mathbf{x}_i)) \geq \rho - \xi_i \\ & \xi_i \geq 0 \quad (i=1, \dots, n). \end{aligned} \quad (4)$$

ここで $\xi_i (i=1, \dots, n)$ はスラック変数である。これは2クラス判別手法である ν -SVM[7]において全訓練パターンがクラス1に属し、原点を唯一のクラス2に属するパターンとみなして準（半）学習することと同じである。

次に最適化問題(4)を主問題としたときの双対問題を導出するため、非負のラグランジュ乗数 $\alpha_i, \gamma_i (i=1, \dots, n)$ を導入し、以下のようなラグランジュ関数を定

義する。

$$\begin{aligned} L(\mathbf{w}, \rho, \xi, \alpha, \gamma) = & \frac{1}{2} \|\mathbf{w}\|^2 - \rho + \frac{1}{\nu n} \sum_{i=1}^n \xi_i \\ & - \sum_{i=1}^n \alpha_i \{ \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \rho + \xi_i \} - \sum_{i=1}^n \gamma_i \xi_i. \end{aligned} \quad (5)$$

最適解において、主問題の各変数 \mathbf{w}, ρ, ξ の偏微分係数は0になることにより、

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^n \alpha_i \phi(\mathbf{x}_i) = \mathbf{0} \quad \therefore \mathbf{w} = \sum_{i=1}^n \alpha_i \phi(\mathbf{x}_i) \quad (6)$$

$$\frac{\partial L}{\partial \rho} = -1 + \sum_{i=1}^n \alpha_i = 0 \quad \therefore \sum_{i=1}^n \alpha_i = 1 \quad (7)$$

$$\frac{\partial L}{\partial \xi_i} = \frac{1}{\nu n} - \alpha_i - \gamma_i = 0 \quad \therefore 0 \leq \alpha_i \leq \frac{1}{\nu n} \quad (i=1, \dots, n) \quad (8)$$

を得る。式(6)~(8)をラグランジュ関数(5)に適用することで主問題(4)の双対問題は以下のように表される。

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.} \quad & \sum_{i=1}^n \alpha_i = 1 \\ & 0 \leq \alpha_i \leq \frac{1}{\nu n} \quad (i=1, \dots, n). \end{aligned} \quad (9)$$

ただし、 k は ϕ に対応するカーネル関数で $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ である。

最適化問題の解において、KKT条件より次の2式が成り立つ。

$$\alpha_i \{ \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \rho + \xi_i \} = 0 \quad (i=1, \dots, n) \quad (10)$$

$$\left(\frac{1}{\nu n} - \alpha_i \right) \xi_i = 0 \quad \therefore \gamma_i \xi_i = 0 \quad (i=1, \dots, n) \quad (11)$$

そして式(8), (10), (11)より $i=1, \dots, n$ について以下の条件が成立する。

$$\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \rho + \xi_i > 0 \text{ かつ } \xi_i = 0, \text{ つまり,}$$

$$\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \rho > 0 \text{ のとき, } \alpha_i = 0. \quad (12)$$

$$\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \rho + \xi_i = 0 \text{ かつ } \xi_i > 0, \text{ つまり,}$$

$$\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \rho < 0 \text{ のとき, } \alpha_i = \frac{1}{\nu n}. \quad (13)$$

$$\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \rho + \xi_i = 0 \text{ かつ } \xi_i = 0, \text{ つまり,}$$

$$\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - \rho = 0 \text{ のとき, } 0 \leq \alpha_i \leq \frac{1}{\nu n}. \quad (14)$$

式(6)より、重みベクタ \mathbf{w} は最適化問題(9)によって得られるラグランジュ乗数 α_i を用いて、訓練パターンの線型結合として表されるが、条件(12)~(14)より、得られる超平面上のパターンおよび、原点側にあるパターン（外れ点と呼ぶことにする）のみが超平面を記述するのに必要なパターンとなることがわかる。これらのパターンを合わせてサポートベクタ (support vector) と呼ぶ。また、超平面上のパターンを \mathbf{x}_B とすると ρ は次式のように求めることができる。

$$\rho = \langle \mathbf{w}, \phi(\mathbf{x}_B) \rangle = \sum_{i=1}^n a_i k(\mathbf{x}_B, \mathbf{x}_i). \quad (15)$$

したがって、判別関数 $f(\mathbf{x})$ は次式のように書ける。

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^n a_i k(\mathbf{x}, \mathbf{x}_i) - \sum_{i=1}^n a_i k(\mathbf{x}_B, \mathbf{x}_i) \right). \quad (16)$$

3.2 ν プロパティ

パラメータ ν についての重要な性質である ν プロパティ (ν -property) について説明する。

命題 3.1 (ν プロパティ [8])

最適化問題(4)の解において $\rho \neq 0$ であると仮定する。このとき以下の性質が成り立つ。

1. ν は外れ点の割合の上限である。
2. ν はサポートベクタの割合の下限である。
3. $\mathbf{x}_i (i=1, \dots, n)$ が連続な分布 P から互いに独立に生成されたもので、カーネルが解析的かつ定数でないとき、 $n \rightarrow \infty$ のとき確率 1 で、外れ点の割合とサポートベクタの割合は ν に収束する。

【証明】

性質の 1 および 2 について 最適化問題(4)の最適解において、集合 $I_{in}, I_{bound}, I_{out} \subseteq \{1, \dots, n\}$ を、 $I_{in} = \{i : \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle > \rho\}$, $I_{bound} = \{i : \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle = \rho\}$, $I_{out} = \{i : \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle < \rho\}$ と定義する。式(7)を両辺 ν 倍すると、

$$\nu = \sum_{i=1}^n \nu a_i = \sum_{i \in I_{in}} \nu a_i + \sum_{i \in I_{bound}} \nu a_i + \sum_{i \in I_{out}} \nu a_i$$

となる。これは条件(12), (13)より、

$$\nu = \sum_{j \in I_{in}} 0 + \sum_{i \in I_{bound}} \nu a_i + \sum_{i \in I_{out}} \frac{1}{n} \quad (17)$$

と書ける。また条件(14)より、

$$0 \leq \sum_{i \in I_{bound}} \nu a_i \leq \sum_{i \in I_{bound}} \frac{1}{n}$$

が成り立つので、これを式(17)に代入すると、

$$\sum_{i \in I_{out}} \frac{1}{n} \leq \nu \leq \sum_{i \in I_{bound} \cup I_{out}} \frac{1}{n} \quad (18)$$

が成立する。よって ν は外れ点の割合の上限であり、サポートベクタの割合の下限である。□

性質の 1 および 2 については主問題から証明する方法もある [13], [15]。こちらはより直感的な理解を与える。

性質の 3 について 詳細は省略するが ([7] に詳しい)、 $\mathbf{x}_i (i=1, \dots, n)$ が連続な分布 P から互いに独立に生成されたもので、カーネルが解析的かつ定数でないとき、 $\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle = \rho$ となる \mathbf{x}_i は孤立点となる。したがって $n \rightarrow \infty$ のとき、確率 1 で、 $\sum_{i \in I_{bound}} 1/n$ は 0 に収束する。このとき、式(18)より、外れ点の割合とサポートベクタの割合は ν に収束する。□

この ν プロパティより、パラメータ ν が外れ点の割合を調整することがわかる。さらに、 $\nu = 1 - \beta$, $0 < \beta \leq 1$ とすると、OC-SVM は第 2 節で定義した β 高密度領域の推定を目指していると言える。

3.3 ガウシアンカーネルの特性とマージン最大化

本節では OC-SVM が ϕ で写した先の特徴空間において、原点と判別超平面とのマージンを大きくすることがどのように高密度領域推定に関係しているのかを考える。SVM とその派生手法において、よく使われるカーネル関数 $k(\mathbf{x}, \mathbf{y})$ には、

$$\text{多項式カーネル} : (1 + \langle \mathbf{x}, \mathbf{y} \rangle)^q \quad (19)$$

$$\text{ガウシアンカーネル} : \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{l^2}\right) \quad (20)$$

$$\text{シグモイドカーネル} : \tan h(\gamma \langle \mathbf{x}, \mathbf{y} \rangle - \delta) \quad (21)$$

などがある。ただし、 q は自然数のパラメータ、 $h(>0)$, γ, δ は実数のパラメータである。OC-SVM では適切なカーネルを選択しなければ意味のある推定量は得られない [14]。本節では OC-SVM によく使われる(20)式のガウシアンカーネルを取り上げ、詳しく見ていく。

ガウシアンカーネルは 2 点間の距離 $\|\mathbf{x}_i - \mathbf{x}_j\|$ に依存した RBF カーネルであり、任意の \mathbf{x}_i に対して、 $k(\mathbf{x}_i, \mathbf{x}_i) = 1$ となる。このことから、各 \mathbf{x}_i はノルム 1 のベクター $\phi(\mathbf{x}_i)$ に写像されることがわかる。さらに付け加えると、2 つのベクター $\phi(\mathbf{x}_i)$ と $\phi(\mathbf{x}_j)$ 間の偏角 θ は

$$\cos \theta = \frac{\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle}{\|\phi(\mathbf{x}_i)\| \|\phi(\mathbf{x}_j)\|} = k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$$

によって与えられる。このことは全ての訓練パターンが ϕ によって半径 1 の超球の四分球の表面上に写像されることを意味する [9]。

図 2 はガウシアンカーネルに対応する写像で写した先を 2 次元の概念図で表現したものである。大きな部分円は各パターンが写像される半径 1 の超球である。そして、右上に位置する円は、全て (ほとんど) の訓練パターンを含む最小の超球を求める Support Vector Domain Description (SVDD) [11], [12] と呼ばれる手法によって得られる超球を表す。多くの訓練パターンが判別超平面を挟んで原点との反対側にあることを維持しながらマージンを大きくすることで、高密度領域を推定しようとしていることを図 2 は示す。

また、判別超平面: $\langle \mathbf{w}, \phi(\mathbf{x}) \rangle - \rho = 0$ は、 $k(\mathbf{x})$ を i 番目の要素が $k(\mathbf{x}, \mathbf{x}_i)$ となる n 次元ベクタとして、 $\langle \mathbf{a}, k(\mathbf{x}) \rangle - \rho = 0$ と書くことができ、 $k(\mathbf{x})$ の空間 (図

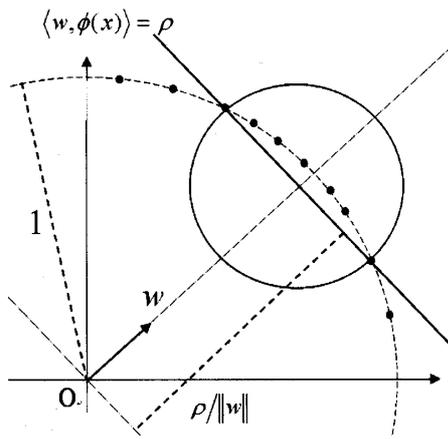


図2 $\phi(\mathbf{x})$ の空間の概念図

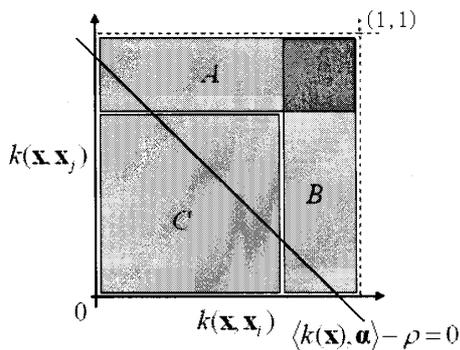


図3 $k(\mathbf{x})$ の空間の概念図: A は \mathbf{x}_j に近いパターンがプロットされる領域, B は \mathbf{x}_i に近いパターンがプロットされる領域, C は $\mathbf{x}_i, \mathbf{x}_j$ ともに遠いパターンがプロットされる領域をイメージしたもの.

3 参照) で原点と訓練パターンを分離しているとみることができ、 $k(\mathbf{x}, \mathbf{x}_i)$ は $[0, 1]$ の範囲で \mathbf{x} と \mathbf{x}_i の近さを表す。したがって、 \mathbf{x} がどの訓練パターンにも近くなければ、 $k(\mathbf{x})$ のほとんどの要素が 0 に近い値となり、原点に近いところに分布することがわかる。このことから、原点と訓練パターンを分離するように判別超平面を決定することが高密度領域の推定になることが理解できよう。

3.4 カーネル密度推定と近傍サポート

本節では、ガウシアンカーネルを用いた OC-SVM とカーネル密度推定の関連性について説明し、近傍サポート (Support Neighbor) として OC-SVM を理解することを考える。

高密度領域を推定する方法として次のようなカーネル密度推定量 \hat{f}_n を用いる方法がある [10]。

$$\hat{f}_n(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^n k(\mathbf{x}, \mathbf{x}_i) \quad (22)$$

ここで $k(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{h^2}\right)$ とおけば、ガウシ

アンカーネルを用いたカーネル密度推定量になる。ただし d は訓練パターンの次元である。これはガウシアンカーネルを用いた OC-SVM で得られる超平面

$$\langle \mathbf{w}, \phi(\mathbf{x}) \rangle = \sum_{i=1}^n a_i k(\mathbf{x}, \mathbf{x}_i) \quad (23)$$

と同じ形をしている。特に $\nu=1$ として OC-SVM を解くと最適解において、 $a_i = 1/n \propto 1/nh^d (i=1, \dots, n)$ となり、得られる超平面はカーネル密度推定量と一致することがわかる。

続いて、一般の $\nu \in (0, 1]$ においてガウシアンカーネルを用いた OC-SVM とカーネル密度推定量の関係性について考察する。ガウシアンカーネルに対応した写像 ϕ によって写した特徴空間上での距離 $d_{ij}^* = \|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|$ を考える。このとき $d_{ij}^{*2} = 2 - 2k(\mathbf{x}_i, \mathbf{x}_j)$ から $k(\mathbf{x}_i, \mathbf{x}_j) = 1 - d_{ij}^{*2}/2$ と計算できる。したがって、OC-SVM の双対問題(9)は次のように書き換えられる。

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} + \frac{1}{4} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_{ij}^{*2} \\ \text{s.t.} \quad & \sum_{i=1}^n \alpha_i = 1 \\ & 0 \leq \alpha_i \leq \frac{1}{\nu n} \quad (i=1, \dots, n) \end{aligned} \quad (24)$$

最適化問題(24)は $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d_{ij}^{*2}$ の最大化を目指している。命題 3.1 より、 $n \rightarrow \infty$ のとき、 α_i は $1/\nu n$ または 0 になる。したがって、特徴空間上において、 $\alpha_i = 1/\nu n$ となるサポートベクタ間の距離の二乗和が最大となるようにサポートベクタを νn 個選ぶことになる。これは、特徴空間上において $\alpha_i = 0$ となる非サポートベクタ間の距離の二乗和が最小となるように非サポートベクタを $(1-\nu)n$ 個選ぶことに相当する。また、ユークリッド距離の計算から、 $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ としたとき、 $\sum_{i=1}^n \sum_{j=1}^n d^2(\mathbf{x}_i, \mathbf{x}_j) = 2n \sum_{i=1}^n d^2(\mathbf{x}_i, \bar{\mathbf{x}})$ となり、各パターン同士の距離の二乗和は、各パターンとその平均からの距離の二乗和に比例することがわかる。以上のことから次の定理が成り立つ。

定理 3.1 ([4])

ガウシアンカーネルを用いた OC-SVM を考える。このとき非サポートベクタとなるパターンの集合は、漸近的に、特徴空間において、それらの平均からの距離の二乗和を最小にするパターンの集合となる。

カーネル密度推定量 $\hat{f}_n(\mathbf{x})$ を用いた最頻値推定量は $\arg\max_{\mathbf{x}} \hat{f}_n(\mathbf{x})$ として与えられる [2]。 ϕ の逆像 ϕ^{-1}

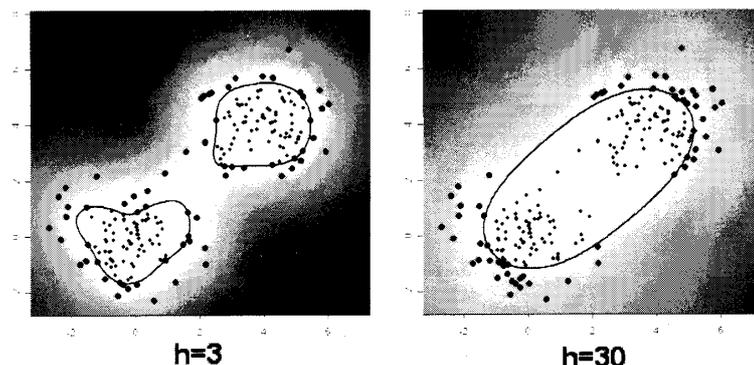


図4 ガウシアンカーネルを用いたOC-SVMの実行例： $\langle \mathbf{w}, \phi(\mathbf{x}) \rangle$ の値でグラデーションを付けてある。曲線は $\langle \mathbf{w}, \phi(\mathbf{x}) \rangle = \rho$ のデータ空間上での境界で、大きな黒点はサポートベクターである。

が存在する場合、 $\mathbf{x}^* = \phi^{-1}\left(\frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i)\right)$ は最頻値（モード）推定量である。実際、 $k(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{h^2}\right)$ とすると、

$$\begin{aligned} \hat{p}_n(\mathbf{x}) &= \frac{1}{nh^d} \sum_{i=1}^n k(\mathbf{x}, \mathbf{x}_i) \\ &= \frac{1}{nh^d} \sum_{i=1}^n \left(1 - \frac{1}{2} \|\phi(\mathbf{x}) - \phi(\mathbf{x}_i)\|^2\right) \end{aligned}$$

となり、 $\hat{p}_n(\mathbf{x}^*)$ が最大値となることが容易に理解できる。以上のことと定理3.1から次の定理が言える。

定理3.2([4])

ガウシアンカーネルを用いたOC-SVMを考える。このとき非サポートベクタとして得られるパターンの集合は、漸近的に、この非サポートベクタを用いたガウシアンカーネルによるカーネル密度推定量によって計算される最頻値推定量の最近傍にあるパターンの集合となる。

定理3.2より、OC-SVMはカーネル密度推定による最頻値推定量の近傍をサポートするものであると理解できる。

カーネル密度推定量の精度はカーネルの幅を与えるパラメータ h に依る。したがって、OC-SVMによって高密度領域の良い推定量を得られるかどうかは、このパラメータ h の設定に依存する。図4は h を変えたときの実行例である。訓練集合は平均の異なる2つの2次元正規分布から生成された100個ずつの訓練パターンを合わせた200個の訓練パターンからなるもので、パラメータ ν は0.3とした。図4から、望ましい推定量が得られるかどうかは h の選択に大きく依存することが確認できる。

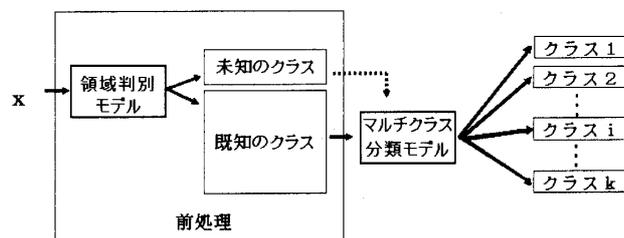


図5 階層構造：未知のクラスに分類されたデータに対しては、マルチクラス分類器を使うことでどの既存クラスに近いかという情報が得られる。

4. 領域判別問題の応用

領域判別問題の応用として、マルチクラス分類問題の前処理への適用を考える。マルチクラス分類は訓練パターンから事前に定義された複数のクラスに分類するモデルを学習し、新たに発生したパターンに適用する。しかし、クラスの定義自体が単純ではない場合、新たに発生したパターンが事前に定義されたクラスに属するパターンであるとは限らないことが考えられる。このとき、前処理としてパターンに領域判別モデルを適用することで定義されたクラスに属するのかわ、未知のクラスに属するのかわを判定できる。つまり、マルチクラス分類のクラスに未知のクラスを追加することができる。領域判別モデルをマルチクラス分類問題に効果的に利用する図5のような階層構造が提案されている[6]。

5. まとめ

本稿では領域判別問題とその判別手法の一つであるOC-SVMについて紹介した。ガウシアンカーネルに対応する写像 ϕ で写した先の特徴空間で、原点との

マージンを大きくするように判別超平面を決定することが高密度領域の推定に繋がること、カーネル密度推定との関係から OC-SVM はカーネル密度推定による最頻値推定量の近傍サポートとして理解できることを説明した。また、ガウシアンカーネルを用いた OC-SVM の性能はカーネルパラメータ h に大きく依存し、適切に選択することが重要であることも指摘した。応用例としては、マルチクラス分類問題の前処理への適用を説明したが、他にもクラスタリングへの応用[1]など様々な研究がある。

参考文献

- [1] Camastra, F. and Verri, A. "ANovel Kernel Method for Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (5): 801-804, 2005.
- [2] Devroye, L. "Recursive estimation of the mode of a multivariate density," *The Canadian Journal of Statistics*, 7 (2): 159-167, 1979.
- [3] Einmal, J. H. J. and Mason, D. M. "Generalized Quantile Processes," *The Annals of Statistics*, 20 (2): 1062-1078, 1992.
- [4] Muñoz, A. and Moguerza, J. M. "One-class Support Vector Machines and Density Estimation: The Precise Relation," *CIARP 2004*, 216-223, 2004.
- [5] Muñoz, A. and Moguerza, J. M. "Estimation of High-Density Regions Using One-Class Neighbor Machines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (3): 476-480, 2006.
- [6] Rocco, C. M. and Zio, E. "A support vector machine integrated system for the classification of operation anomalies in nuclear components and systems," *Reliability Engineering and System Safety*, in press, 2006.
- [7] Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L. "New Support Vector Algorithms," *Neural Computation*, 12 (5): 1207-1245, 2000.
- [8] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. "Estimating the Support of a High Dimensional Distribution," *Neural Computation*, 13 (7): 1443-1471, 2001.
- [9] Schölkopf, B. and Smola, A. J. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.
- [10] Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.
- [11] Tax, D. M. J. and Duin, R. P. W. *Data Domain Description using Support Vectors*. In Proceedings of European Symposium on Artificial Neural Networks '99, Brugge, 1999.
- [12] Tax, D. M. J. and Duin, R. P. W. "Support Vector Domain Description," *Pattern Recognition Letters*, 20: 1191-1199, 1999.
- [13] 佐治明. 1 クラスサポートベクターマシンにおける準教師付学習と自動モデル選択. 筑波大学大学院博士課程システム情報工学研究科修士論文, 2006.
- [14] 鈴木伸彦. *Domain Description based on Reliability Learning*. 筑波大学大学院博士課程システム情報工学研究科修士論文, 2005.
- [15] 高島慧. 周辺密度を用いた *One-Class SVM* の提案. 筑波大学大学院博士課程システム情報工学研究科修士論文, 2006.