

最小楕円に基づく領域判別

後藤 順哉, 武田 朗子

本稿では, CVaR 最小化のテクニックを用い, 最小包囲楕円問題を拡張した最適化問題を提示する. その定式化が正規分布の最尤推定法に対しても一般化となっていることを指摘すると同時に, 領域判別の問題を含む, 他の類似の問題との関係についても紹介する. また, 正規分布の最尤推定の一般形を与える事実を基に, 提案する最小楕円問題の多クラス判別分析への適用を試みる.

キーワード: 最小包囲楕円, 最尤推定, CVaR 最小化

1. はじめに

本稿では, 領域判別や1クラスSVMといった, データの面的な広がり(領域)を捉える話題に関連して, 領域の形状を楕円形に限った議論に焦点を当ててみたい. そこで, やや唐突であるが, 楕円の定義から始めよう.

一般に, n 次元実数空間における楕円体は, $n \times n$ 実対称正定値行列 Q , および n 次元実ベクトル γ によって,

$$E(Q, \gamma) := \{x \in \mathbb{R}^n : \|Qx - \gamma\|^2 \leq n\}$$

と表すことができる. 特に, 上式の等号を満たす点の集合は楕円の境界面を形成する. ちなみに, 中心(位置)ベクトル $c \in \mathbb{R}^n$, および形状行列 D を用いた,

$$\tilde{E}(D, c) := \{x \in \mathbb{R}^n : \langle x - c, D(x - c) \rangle \leq n\}$$

という表現の方が馴染みがある方もおられるであろうが, $D = Q^2$, $c = Q^{-1}\gamma$ なる変数変換を経ることで, 2つの表現は互いに行き来が可能である.

さて, \mathbb{R}^n 上の点集合 $X := \{x^1, \dots, x^m\}$ が与えられたとき, これらから n 次元楕円 $E(Q, \gamma)$ を求める問題は様々な文脈で現れる. 有名で重要な例を2つ, 以下に紹介しよう.

例1) 最小包囲楕円問題 $X \subset \mathbb{R}^n$ が与えられたとき, これら全てをその内側に含み, かつ最小体積を持った楕円を求める問題は, 最小包囲楕円問題と呼ばれる(図1). この問題は比較的歴史が古く, 実験計画, 計

算幾何, ロバスト統計など, 多数の異なる文脈において現れる問題である.

既述の楕円表現を用いると, X の最小包囲楕円は, 以下の最適化問題の解 (Q^*, γ^*) によって, $E(Q^*, \gamma^*)$ として与えられる:

$$\begin{cases} \text{minimize} & -\ln \det[Q] \\ & Q > 0, \gamma \\ \text{subject to} & \|Qx^i - \gamma\|^2 \leq n, i=1, \dots, m. \end{cases} \quad (1)$$

ただし, $Q > 0$ は実対称行列 Q が正定値であることを, $\det[\cdot]$ は行列式を表すものとする.

ここで定式化(1)の意味を簡単に説明しておこう. まず, m 本の制約は, 所与の m 個のデータ点が全て楕円 $E(Q, \gamma)$ に含まれることを要請したものである.

一方, 目的関数は, 楕円 $E(Q, \gamma)$ の体積が

$$\frac{(n\pi)^{n/2}}{\Gamma(n/2+1)} \frac{1}{\det[Q]}$$

で与えられること(および, 対数関数の単調性)から, 体積の最小化を表したものである($\Gamma(\cdot)$ はガンマ関数). 問題(1)は, いわゆる凸計画問題であり, 多くのアルゴリズムが提案されている(既存研究のリストについては[7]などの参考文献欄を, 問題の凸性については[2]などを参照されたい).

例2) 楕円分布のパラメータ推定問題 X が n 次元楕円分布に従うサンプル・データであるとき, これら

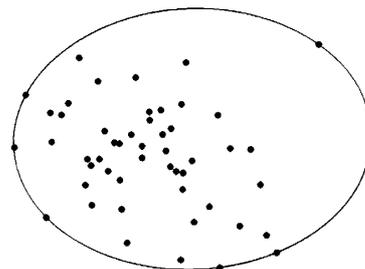


図1 2次元における最小包囲楕円の例

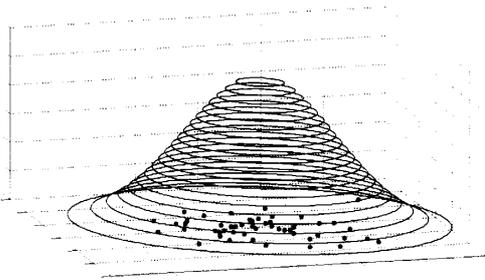


図2 2次元データに対する正規密度の最尤推定

から未知のパラメータを推定し、元の分布形を推定する問題も集合 X と楕円を結びつける問題であると言える。ここで、楕円分布とは $p(\mathbf{x}) := \kappa \det[\mathbf{Q}]q(\|\mathbf{Q}\mathbf{x} - \boldsymbol{\gamma}\|^2)$ なる密度関数を持つ分布である (ただし、 $\kappa > 0$ は定数、 q は \mathbb{R} 上の関数)。この問題が楕円と結びつくのは、楕円分布の密度関数の等高線が、(その名が示す通り) 楕円を与えるからである (図2)。これらのパラメータ ($\mathbf{Q}, \boldsymbol{\gamma}$) を推定する方法についてはいくつもありうるが、最も有名なものは最尤推定法であろう。

楕円分布の最も有名な例は正規分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ であり、 X に対するこの方法は以下の最大化問題として定式化される (例えば、[9] など) :

$$\begin{aligned} & \left| \begin{array}{l} \text{maximize} \quad -m \ln \det[\boldsymbol{\Sigma}] \\ \boldsymbol{\Sigma} > 0, \boldsymbol{\mu} \end{array} \right. \\ & \quad - \sum_{i=1}^m \langle \mathbf{x}^i - \boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}(\mathbf{x}^i - \boldsymbol{\mu}) \rangle. \end{aligned}$$

この問題はデータ X が、ある同一の正規分布から独立にサンプリングされたものと仮定して、尤度 $\prod_{i=1}^m \exp\{-\frac{1}{2}(\mathbf{x}^i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}^i - \boldsymbol{\mu})\} / \sqrt{(2\pi)^n \det[\boldsymbol{\Sigma}]}$ を最大にするパラメータ ($\boldsymbol{\Sigma}, \boldsymbol{\mu}$) を求めるという発想に基づいている。よく知られているように、この問題は (適当な仮定の下で) 陽に解くことができ、その解 ($\boldsymbol{\Sigma}^*, \boldsymbol{\mu}^*$) は以下のように与えられる :

$$\begin{aligned} \boldsymbol{\Sigma}^* &= \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^i - \bar{\mathbf{x}})(\mathbf{x}^i - \bar{\mathbf{x}})^\top; \\ \boldsymbol{\mu}^* &= \frac{1}{m} \sum_{i=1}^m \mathbf{x}^i =: \bar{\mathbf{x}}. \end{aligned}$$

冒頭に記した楕円表記法の差異を考慮すれば、この正規尤度最大化問題は以下のように表すことができることに注意しよう :

$$\text{maximize}_{\mathbf{Q} > 0, \boldsymbol{\gamma}} \quad m \ln \det[\mathbf{Q}] - \frac{1}{2} \sum_{i=1}^m \|\mathbf{Q}\mathbf{x}^i - \boldsymbol{\gamma}\|^2. \quad (2)$$

また、その解 ($\mathbf{Q}^*, \boldsymbol{\gamma}^*$) は

$$\mathbf{Q}^* = \left(\frac{1}{m} \sum_{i=1}^m (\mathbf{x}^i - \bar{\mathbf{x}})(\mathbf{x}^i - \bar{\mathbf{x}})^\top \right)^{-\frac{1}{2}}; \quad \boldsymbol{\gamma}^* = \mathbf{Q}^* \bar{\mathbf{x}} \quad (3)$$

と表すことができる。

このように、データ集合 $X \subset \mathbb{R}^n$ から楕円 $E(\mathbf{Q}, \boldsymbol{\gamma})$ を構築する問題には (これらの他にも) 基本的で重要な問題が含まれる。本稿では、金融工学でよく知られた CVaR (Conditional Value-at-Risk) 最小化のテクニック [5] を援用することで、例1で示した最小包囲楕円問題を拡張した、楕円構築問題の新しい定式化を提示し、その含意、および、領域判別の問題を含む、いくつかの定式化との関連について指摘する。とりわけ、例2に示した正規尤度最大化に対しても最小楕円の文脈から幾何的な意味づけを与えられることを示し、その事実の応用として、多クラス判別問題への応用を試みる。なお、議論の詳細については [4] をご参照いただきたい。

2. 条件付き最小楕円問題

2.1 問題の定式化

まず、以下の最適化問題を導入しよう :

$$\begin{aligned} & \left| \begin{array}{l} \text{minimize} \quad -\ln \det[\mathbf{Q}] \\ \mathbf{Q} > 0, \boldsymbol{\gamma}, a \end{array} \right. \\ & \left| \text{subject to} \quad F_\beta(\mathbf{Q}, \boldsymbol{\gamma}, a) \leq n. \end{array} \quad (4) \right. \end{aligned}$$

ただし、 $\beta \in [0, 1)$ とし、制約式左辺は

$$F_\beta(\mathbf{Q}, \boldsymbol{\gamma}, a) := a + \frac{1}{(1-\beta)m} \sum_{i=1}^m [\|\mathbf{Q}\mathbf{x}^i - \boldsymbol{\gamma}\|^2 - a]^+$$

なる関数であり、 $[w]^+ := \max\{w, 0\}$ と約束する。

本稿では、定式化(4)の解 ($\mathbf{Q}^*, \boldsymbol{\gamma}^*$) によって得られる楕円 $E(\mathbf{Q}^*, \boldsymbol{\gamma}^*)$ を条件付き最小楕円 (β -Conditional Minimum Volume Ellipsoid; 以下、 β -CMVE) と呼ぶ。

以下では所与のデータ X について次を仮定する :

仮定1 $\mathbf{x}^1, \dots, \mathbf{x}^m$ のアフィン包は \mathbb{R}^n を張る。

定理1 仮定1の下で(4)は解を持つ。また、最適解において、制約 $F_\beta(\mathbf{Q}, \boldsymbol{\gamma}, a) \leq n$ は等号で満たされる。

実際、(4)の1本の微分不可能な制約は、複数の微分可能な制約 :

$$a + \frac{1}{(1-\beta)m} \sum_{i=1}^m z_i \leq n,$$

$$z_i \geq \|\mathbf{Q}\mathbf{x}^i - \boldsymbol{\gamma}\|^2 - a, \quad z_i \geq 0, \quad i=1, \dots, m$$

で置き換えられることに注意すると、問題(4)は微分可能な凸計画問題に帰着でき、効率的な内点法アルゴリズムを構築することができる (詳細は [4])。

2.2 定式化の含意

以下では問題(4)の幾何的な解釈を説明していく。このことを示すために、仮定1の下で問題(4)と等価な、次の問題を導入しよう。

$$\begin{cases} \text{minimize} & -\ln \det[\mathbf{Q}] \\ & \mathbf{Q} > 0, \boldsymbol{\gamma} \\ \text{subject to} & \phi_\beta(\mathbf{Q}, \boldsymbol{\gamma}) \leq n. \end{cases} \quad (5)$$

ただし,

$$\phi_\beta(\mathbf{Q}, \boldsymbol{\gamma}) := \min_\alpha F_\beta(\mathbf{Q}, \boldsymbol{\gamma}, \alpha).$$

問題(4)と問題(5)の違いは、制約式における α を $\mathbf{Q}, \boldsymbol{\gamma}$ と同時に決定するか、後で決定するかでしかない。実際、[5]に拠れば、(4)と(5)は解が存在すれば等価である。そこで、以下では関数値 ϕ_β の意味を解釈することで、問題(4)の含意を説明していきたい。

まず、楕円 $E(\mathbf{Q}, \boldsymbol{\gamma})$ に対して、点 \mathbf{x}^i の楕円スコア f^i を以下の様に定義する ($i=1, \dots, m$):

$$f^i(\mathbf{Q}, \boldsymbol{\gamma}) := f(\mathbf{x}^i | \mathbf{Q}, \boldsymbol{\gamma}) := \|\mathbf{Q}\mathbf{x}^i - \boldsymbol{\gamma}\|^2.$$

このスコアは、楕円 $E(\mathbf{Q}, \boldsymbol{\gamma})$ が決まれば、各点 \mathbf{x}^i 毎に計算される数値であり、中心 $\mathbf{Q}^{-1}\boldsymbol{\gamma}$ から点 \mathbf{x}^i への、計量 \mathbf{Q}^2 の下での距離の2乗を示している。特に、(3)で与えられる \mathbf{Q}^* で、 $1/m$ を $1/(m-1)$ とした $(\mathbf{Q}^*, \boldsymbol{\gamma}^*)$ を用いれば、いわゆる (不偏推定量に基づく) マハラノビス距離の2乗に等しいことに注意しよう。

さて、固定された $\mathbf{Q}, \boldsymbol{\gamma}$ に対して、楕円スコアの (経験的) 分布関数 $\Phi(\cdot | \mathbf{Q}, \boldsymbol{\gamma})$ を考えよう:

$$\Phi(\alpha | \mathbf{Q}, \boldsymbol{\gamma}) := \frac{1}{m} |\{i \in \{1, \dots, m\} : f^i(\mathbf{Q}, \boldsymbol{\gamma}) \leq \alpha\}|.$$

さらに、これを用いて楕円スコアの β -分位点 (quantile) を以下の様に定義する ($\beta \in [0, 1]$):

$$\alpha_\beta(\mathbf{Q}, \boldsymbol{\gamma}) := \min\{\alpha \geq 0 : \Phi(\alpha | \mathbf{Q}, \boldsymbol{\gamma}) \geq \beta\}.$$

図3は分布関数 Φ とそれに対応したヒストグラム、そして β -分位点 α_β を大雑把に描いたものである。この図が示すように、 β -分位点 α_β は、分布する楕円スコアのうち、おおよそ下位 $100\beta\%$ (あるいは上位 $100(1-\beta)\%$) 目のスコアを示したものである。

一般に、 ϕ_β と α_β の間に以下の関係が成り立つ ([5]):

$$0 \leq \alpha_\beta \leq E[f | f \geq \alpha_\beta] \leq \phi_\beta \leq E[f | f > \alpha_\beta].$$

ただし、 $E[\cdot]$ は Φ の下での期待値を表し、 $\mathbf{Q}, \boldsymbol{\gamma}$ へ

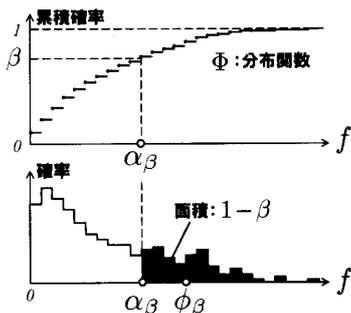


図3 分布関数 Φ (上) とヒストグラム (下)

の依存関係を省略してある。この関係式を眺めると、 ϕ_β が2つの似通った条件付き期待値で挟まれていることが分かる。実際、もし分布関数が連続であれば、これら2つの条件付き期待値は一致する。したがって、 ϕ_β は近似的に、 β -分位点 α_β より大きな (つまり、上位 $(1-\beta)100\%$ に入っている) 楕円スコアの平均値を意味していると解釈できる (図3)。

これより、定式化(5)、ひいては(4)は、おおよそ、楕円スコアの上位 $100(1-\beta)\%$ の平均値を境界面が通る楕円のうち、最小体積を持つものを求める最適化問題であることが分かる。図4は3つの β に対し、実際に計算した β -CMVE (左) と、 β による境界面の位置の差異を描いたもの (右) である。

とりわけ、 β が十分1に近い場合と、 $\beta=0$ の場合には、定式化(4)、(5)より α を除くことが出来、その含意がより鮮明になる:

命題1 (i) $\beta > 1 - \frac{1}{m}$ のとき、問題(5)は最小包圍楕円問題(1)と等価である。

(ii) $\beta=0$ のとき、問題(5)は以下で定義される問題と等価である:

$$\begin{cases} \text{minimize} & -\ln \det[\mathbf{Q}] \\ & \mathbf{Q} > 0, \boldsymbol{\gamma} \\ \text{subject to} & \frac{1}{m} \sum_{i=1}^m \|\mathbf{Q}\mathbf{x}^i - \boldsymbol{\gamma}\|^2 \leq n. \end{cases} \quad (6)$$

また、仮定1の下で(6)の解 $(\mathbf{Q}^*, \boldsymbol{\gamma}^*)$ は(3)で与えられる。

1つ目の命題は β -CMVE が最小包圍楕円問題 (例1) の拡張になっていることを示している。これは最小包圍楕円問題(1)の m 本の制約が、

$$\max_{i=1, \dots, m} \{f^i(\mathbf{Q}, \boldsymbol{\gamma})\} \leq n$$

と書き換えられることから、先に示した ϕ_β の意味づけを通して直感的に納得できるであろう (図4右上)。

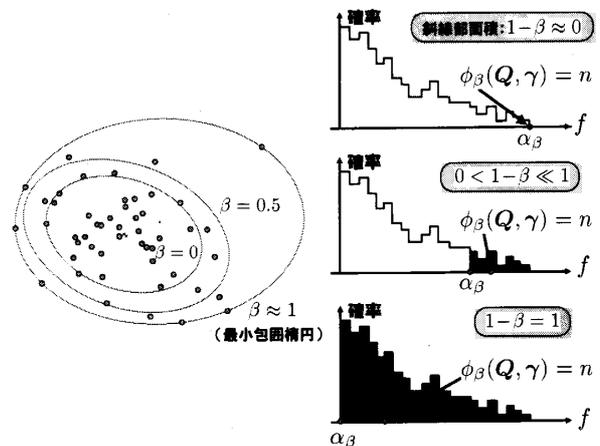


図4 3つの β -CMVE と境界の位置 (イメージ図)

2つ目の命題についても、 $\beta=0$ のとき、 ϕ_β は(条件付きでない)全体の平均値を意味することから納得できるものである(図4右下)。一方、その解(3)については、正規分布の(対数)尤度最大化(2)と同じ解になっていることに注意されたい。このことから、CMVEは正規尤度最大化についての拡張にもなっており、裏を返せば、正規尤度最大化に対し、最小楕円という文脈での幾何的な解釈を与えることが分かる。

3. 条件付き尤度最大化

前節では正規分布の最尤法に対する、最小楕円による幾何的な解釈を示したが、本節では逆に、 β -CMVEを尤度最大化の文脈に位置づけてみたい。まず、 $\beta \in [0, 1)$ に対し、次の一般化した尤度最大化問題を導入する：

$$\begin{cases} \text{maximize}_{Q>0, \gamma, a} & m \ln \det[\mathbf{Q}] \\ & -\frac{1}{2} \sum_{i=1}^m \left(\alpha + \frac{1}{1-\beta} [f^i(\mathbf{Q}, \boldsymbol{\gamma}) - \alpha]^+ \right). \end{cases} \quad (7)$$

簡単に分かるように、 $\beta=0$ とすれば、通常の正規尤度最大化(2)と等価である。また、この問題は自明に次のように書き換えることができることに注意しておく：

$$\begin{cases} \text{minimize}_{Q, \gamma, a, z} & -\ln \det[\mathbf{Q}] + \frac{1}{2} \left\{ \alpha + \frac{\mathbf{e}^T \mathbf{z}}{(1-\beta)m} \right\} \\ \text{subject to} & z_i \geq \|\mathbf{Q}\mathbf{x}^i - \boldsymbol{\gamma}\|^2 - \alpha, i=1, \dots, m, \\ & \mathbf{z} \geq \mathbf{0}, \mathbf{Q} > \mathbf{O}. \end{cases} \quad (8)$$

以下の命題はこの問題と β -CMVEの等価性を示している。

命題2 問題(7)と問題(4)は同じ楕円 $E(\mathbf{Q}, \boldsymbol{\gamma})$ を出力する。

ここで定式化(7)について詳しく見ておこう。目的関数を指数関数 $\exp\{\cdot\}$ で変換すると、(7)は次の関数の最大化と同じであることが理解できる： $\prod_{i=1}^m l^i(\mathbf{Q}, \boldsymbol{\gamma})$ 。ただし、 $l^i(\mathbf{Q}, \boldsymbol{\gamma}) :=$

$$\begin{cases} \det[\mathbf{Q}] \exp\left\{ -\frac{1}{2} \left(\alpha + \frac{1}{1-\beta} (f^i(\mathbf{Q}, \boldsymbol{\gamma}) - \alpha) \right) \right\} \\ \quad \text{for } i \text{ such that } f^i(\mathbf{Q}, \boldsymbol{\gamma}) > \alpha, \\ \det[\mathbf{Q}] \exp\left\{ -\frac{1}{2} \alpha \right\} \\ \quad \text{for } i \text{ such that } f^i(\mathbf{Q}, \boldsymbol{\gamma}) \leq \alpha. \end{cases}$$

すなわち、全てのデータが等しく正規尤度に寄与するのではなく、 α より大きな楕円スコアを有する(つまり、より小さい尤度を有する)もののみが全体の尤度に寄与する、そんな条件付きの尤度関数を最大化する

ことと等しい。実際、最適解において、 α は β -分位点 α_β の近似を与える([5])ことから、問題(7)、ひいては β -CMVEを求める(4)は、下位 $100(1-\beta)\%$ の尤度を最大化する(その意味でmax-min最尤法の拡張としての)条件付き最尤推定問題と見なせる。

4. 他の定式化との関連

4.1 Sun-Freund[7]の拡張との比較

これまでに β -CMVEが最小包圍楕円(1)の一般化になっていることを指摘したが、[7]では(1)を直接緩和する形で一般化した定式化が以下の形で示されている：

$$\begin{cases} \text{minimize}_{Q, \gamma, z} & -\ln \det[\mathbf{Q}] + \mathbf{P} \mathbf{e}^T \mathbf{z} \\ \text{subject to} & \|\mathbf{Q}\mathbf{x}^i - \boldsymbol{\gamma}\|^2 \leq 1 + z_i, i=1, \dots, m, \\ & \mathbf{z} \geq \mathbf{0}, \mathbf{Q} > \mathbf{O}. \end{cases} \quad (9)$$

ただし、 $P > 0$ はパラメータで、 $\mathbf{e} := (1, \dots, 1)^T$ 。(9)も β -CMVE同様、1つのパラメータを有している。実際、両者の間には次のような対応を見いだすことができる。

命題3 $(\mathbf{Q}^*, \boldsymbol{\gamma}^*, \alpha^*, \mathbf{z}^*)$ を問題(8)の解で、 $\alpha^* > 0$ とする。このとき、 $\left(\frac{1}{\sqrt{\alpha^*}} \mathbf{Q}^*, \frac{1}{\sqrt{\alpha^*}} \boldsymbol{\gamma}^*, \frac{1}{\alpha^*} \mathbf{z}^* \right)$ は $P = \frac{\alpha^*}{2(1-\beta)m}$ とした問題(9)の解である。

実際、[7]では最小包圍楕円が外れデータ(例えば、図4左の右上の点)に対し、極めて敏感な解を出力する点を指摘し、それに対する一つの解決策として(9)を提示している。しかしながら、 P は正数である以外に制限はなく、その意味合いも曖昧である。それに比べると、 β -CMVEは、外れデータに対する免疫を加えるという点では、同様の効果を示すことが出来る上に、パラメータ β の意味合いは、既に述べてきたように、 β -分位点以上の期待値が境界面になるという、幾何的な解釈を伴うという御利益がある点でより良いパラメータ化であると言える。

このようなパラメータの導入に関わる議論は、SVMの分野においてC-SVMと ν -SVMの差異として知られるものと同じである(詳細は[6]など)。

4.2 SVDD[8]の拡張としてのCMVE

ここで話を一旦領域判別の手法として知られるSupport Vector Domain Description (SVDD) [8]へと移そう。これは、サンプル点集合 $X \subset \mathbf{R}^n$ が与えられたときに、それが暗示するもとの分布の密度を推定する方法の1つで、その基本的な定式化が以下の

形で与えられるものである (以下は[6]を参考):

$$\begin{cases} \underset{c, r, z}{\text{minimize}} & r^2 + \frac{1}{\nu m} e^{-z} \\ \text{subject to} & \|x^i - c\|^2 \leq r^2 + z_i, i=1, \dots, m, \\ & z \geq 0. \end{cases} \quad (10)$$

ただし, $\nu \in (0, 1]$ はパラメータである. $\nu = 1 - \beta$, $r^2 = \alpha$ と置き換え, この定式化を眺めると, (10)は(8)において Q を単位行列に固定したものと等しく, 楕円を r を半径とする n 次元超球に限定した, 条件付き最小球問題と見なすことができることが分かる. このことから, β -CMVE は SVDD (10) の (球から楕円へという意味で) 自然な拡張である.

5. 多クラス判別への応用

節 2, 3 において, β -CMVE の求解が正規分布の最尤推定に対する拡張として見なすことができることを指摘した. 本節では, この事実から, 従来正規性の仮定に基づいて展開されてきた議論に β -CMVE を適用することを試みる. 具体的には, フィッシャーの判別分析として知られる方法論に対して, この一般化が結果を改善できるかどうかについて比較を示す.

ここで扱うのはいわゆる多クラス判別と呼ばれる問題で, 各データ $x^i \in \mathbb{R}^n$ がいずれか 1 つのクラス $k^i \in K$ に属しているものとする. このとき, 手許にあるクラス・ラベル付きデータの集合 (学習データと呼ぶ) $\{(x^i, k^i)\}$ から, なるべく良い判別規則を学習する (=作成する) ことがここでの目標となる. ここで, 判別規則作成に用いなかったクラス・ラベル付きデータ (テスト・データと呼ぶ) を, 作成した判別規則に当てはめたときに, どれだけ正確にクラス・ラベルを言い当てることができるかという観点で判別規則の良し悪しを測る.

具体的には, 10 分割のクロス・バリデーションを用いる. この方法は手許にあるデータ集合を 10 分割し, そのうち 9 分割分のデータで判別規則を作成し, 残り 1 分割分のデータでその判別規則の予測精度をテストするということを, (全てのデータが 1 度テスト・データになるように) 10 回行い, その総計 (もしくはは平均値) で評価するというものである.

本実験では β -CMVE を利用して, 以下の手順に従い判別規則を構成する:

1. 各クラス k 毎に β_k を固定し, 各クラスの学習データ $\{(x^i, k^i) : k^i = k\}$ に対し, β_k -CMVE を求める.

2. 全てのクラスの β_k -CMVE が求まったら, 次の (a), (b), (c), いずれかの基準に則って, テスト・データ \bar{x} のクラス・ラベル \bar{k} を予測する.
 - (a) ベイズ判別基準

$$\bar{k} \in \arg \max_{k \in K} \left\{ \ln \det[Q^k] - \frac{1}{2} f(\bar{x} | Q^k, \gamma^k) + \ln m_k \right\}.$$

ただし, m_k は学習に用いたクラス k のサンプル・データ数. すなわち, $m_k = |\{(x^i, k^i) : k^i = k\}|$.

- (b) 正規尤度基準

$$\bar{k} \in \arg \max_{k \in K} \left\{ \ln \det[Q^k] - \frac{1}{2} f(\bar{x} | Q^k, \gamma^k) \right\}.$$

- (c) 変形マハラノビス距離 (楕円スコア) 基準

$$\bar{k} \in \arg \min_{k \in K} f(\bar{x} | Q^k, \gamma^k)$$

本実験では, 表 1 に示したデータ・セットに対し, 全てのクラス k について, 11 個の β_k -CMVE ($\beta_k = 0.0, 0.05, 0.15, \dots, 0.95$) を計算し, その全ての組み合わせに関するクロス・バリデーションを行った. 以下の結果表示ではその組み合わせの中で最良の結果のみ示している.

表 1 は上記で述べた β -CMVE を用いた判別, フィッシャーの判別, ν -SVM の 3 つの手法のテスト・データにおける誤判別率をまとめたものである. ここで言うフィッシャーの判別とは, いわゆる線形判別ではなく, 上述の 3 つの判別基準(a)-(c)に, (3)で $1/m_k$ を $1/(m_k - 1)$ として計算した (Q^k, γ^k) を採用したものを指している. m_k が大きい場合には(3)とほぼ違いがなくなることから, フィッシャー判別は近似的に β -CMVE の特殊ケースと見なせることに注意しよう.

一方, ν -SVM の結果は LIBSVM[3]に基づくものであり, 線形, および RBF カーネルを採用した¹. パラメータ ν については, CMVE の β と同様に, 11 種類, RBF カーネルについては, カーネルのパラメータ γ についても $\gamma = 2^{-25}, 2^{-21}, \dots, 2^{-7}$ の 10 個について計算し最良の結果を示してある. ただし, LIBSVM の制約から, 全てのクラス共通の ν, γ を採用している.

表から, ほとんど全てのデータ, および判別基準で,

¹ カーネルは SVM が高度に非線形な判別を行うことを可能にする仕組みであり, RBF カーネルはその中でも最も有名なもので, 事前に固定すべきパラメータを 1 つ有する. 一方, 線形カーネルとはカーネルを適用しないことと同義である.

表1 テスト・データにおける誤判別率[%]

データ名	K	サンプル数 (内訳)	n	CMVE			フィッシャー判別			ν-SVM[3]	
				(a)	(b)	(c)	(a)	(b)	(c)	線形	RBF
Heart	2	270 (150, 120)	13	16.67	16.67	16.30	17.04	17.04	17.04	15.56	14.44
Liver	2	345 (145, 200)	6	33.04	32.75	29.86	42.03	42.61	29.86	38.50	24.92
Diabetes	2	768 (500, 268)	8	23.70	23.44	22.92	26.30	27.08	32.90	23.96	22.66
Cancer	2	569 (357, 212)	30	4.04	4.22	11.25	4.04	4.39	12.83	5.98	4.40
Cancer (3)	2	569 (357, 212)	3	3.69	3.51	2.99	3.87	4.22	8.61	7.04	6.86
Iris	3	150 (50, 50, 50)	4	2.00	2.00	2.00	2.00	2.00	2.67	1.33	1.33
Wine	3	178 (59, 71, 48)	13	0.00	0.00	0.00	0.56	0.56	1.69	7.81	7.81
Vehicle	4	846 (218, 212, 217, 199)	18	13.95	13.83	14.07	14.89	14.89	15.37	28.59	17.86

(a) ベイズ判別規則; (b) 正規尤度基準; (c) 変形マハラノビス距離基準; "Cancer (3)" は既存研究を参考に属性 no.2, 24, 25 のみを用いたもの

CMVEの結果がフィッシャー判別の結果を上回っていることが分かる。また、いくつかのデータセットに対してはν-SVMの結果をも凌いでいる。例えば、WineデータについてはCMVEで適当にβを変えることで、0%の事後的な予測誤差を達成できている。これだけの結果で圧倒的な優位性を主張することはできないが、一般に通常のデータが厳密に正規分布に従うとは考えられないことから、本稿で示した一般化によって、正規性からの逸脱に対していくらかの対応が可能になっている様子がうかがえる。

また、SVMはRBFなどの非線形カーネルを用いることで、高度に非線形な判別ルールを作ることができるが、その結果はパラメータのチューニングに大きく依存し、場合によってはオーバー・フィッティングすることもあった。その意味ではフィッシャー判別や、それを少し拡張した本方法論の様に、予め領域の形状に強い（しかしながらそれなりに記述力を保持した）制限を掛けておく、古典的な方法論も捨てがたい側面を有しているように思える。

6. おわりに

本稿では最小包囲楕円問題（例1）を拡張した最適化問題を提示し、それが正規尤度最大化（例2）に対する一般化にもなっている点を指摘した。ここでは[5]でまとめられているCVaR最小化のテクニックを援用したが、SVDDやν-SVMにおいて同様のテクニックが用いられ、研究されていることは興味深い。

これまで最適化は統計学や経済学において1つの重要なツールとされてきたが、好まれるのは大抵、無制約か、全ての制約が等号で満たされ、解が陽に（あるいは直ちに）求まる（本稿で言う(3)の様な）形のものが主流である。SVMをはじめ、最適化のアプローチ

がその“らしさ”を発揮する1つの側面は、データの（全部でなく）一部に依存させた解を出力してくれる点にある。とりわけ、 ϕ_β のように、平均的位置から外れた部分（周辺部、あるいは分布の裾部分）に焦点を当てることの有効性は近年いろいろな場面で指摘されている。本方法論がそのような文脈で知識発見の一助となることを期待したい。

参考文献

- [1] C. L. Blake and C. J. Merz, UCI Repository of machine learning databases, 1998. [<http://www.ics.uci.edu/~mllearn/MLRepository.html>].
- [2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [3] C. C. Chang and C. J. Lin, LIBSVM: A Library for Support Vector Machines, 2001. [<http://www.csie.ntu.edu.tw/~cjlin/libsvm>].
- [4] J. Gotoh and A. Takeda, "Conditional Minimum Volume Ellipsoid with Application to Multiclass Discrimination," 社会システム・マネジメント専攻 Discussion Paper Series, (2006), 筑波大学.
- [5] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *Journal of Banking and Finance*, 26 (2002), 1443-1471.
- [6] B. Schölkopf and A. J. Smola, *Learning with Kernels*, The MIT Press, 2002.
- [7] P. Sun and R. M. Freund, "Computation of Minimum Volume Covering Ellipsoids," *Operations Research*, 52 (2004), 690-706.
- [8] S. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognition Letters*, 20 (1999), 1191-1199.
- [9] 宮川雅巳, 『統計技法』, 共立出版株式会社, 1998.