2008 年度前期　応用統計学　第９回

判別分析とパターン認識 **(3)** ー サポートベクタマシンとカーネル法

サポートベクタマシンは，空間中に配置された点の２つの集合を最適に分離する境界を，その集合に属する点の分布を表す確率分布モデルを考えることなく求める方法です．その基本的アイデアは大変簡単で，「境界を，それぞれの集合でもっとも境界に近い点のどちらからも，もっとも遠くなるように置く」というものです．この簡単な考え方はかなり古くからあるものですが，最近ふたたび脚光を浴びています．それは，空間をさらに高次元の空間に変換するのと同等の操作を行なう「カーネル法」という方法を導入することによって，線形でない「曲がった」境界を求めることができるようになり，より複雑な認識問題に対応できるようになったためです．

基本的なサポートベクタマシン

まず最初に，前回のニューラルネットワークについての講義で述べた「線形分離可能」な問題について考えます（図 1）．この問題について，それぞれの集合（○と●）を「最適に」分離する超平面を求めます．ここでいう「最適な」分離超平面とは，それぞれの集合に現在存在する点を完全に分離するだけでなく，それぞれの集合に存在するであろう「未知」の点をも分離することができる，という意味です．第８回で扱った判別分析は，それぞれの集合を形成する確率分布，すなわち「いま存在する点以外の点が，空間中のどのような場所に現れやすいか」を推定することで，ある程度「未知」の点を分離できるようにしています．しかし，第８回の最後で「次元のわな」として述べたように，通常のパターン認識問題では，空間の次元は既知の点の数よりもはるかに大きく，既知の点は空間中に非常に疎にしか分布していないので，確率分布の推定は実際は非常に難しいことになります．

そこで，確率分布の推定を必要としない，別の簡単な方法を考えます．この方法では，「最適」な境界を，「それぞれの集合のどちらからももっとも離れている境界」と考えます．言い換えると，この境界はそれぞれの集合の「ちょうど中間」を通ります．それぞれの集合の分布をあらわす確率分布は不明ですが，このような境界は，どちらの集合からももっとも離れているのですから，それぞれの集合の未知の点ももっともうまく分離できると期待されます．それぞれの集合に属する点のうち，この境界にもっとも近い点をサポートベクトルといいます．

このような境界は，それぞれの集合の凸閉包（集合に属するすべての点を囲む最小の凸図形）を結ぶ最短の線分の中点を通り，その線分に垂直な超平面となります．

$\boldsymbol{x}$ を，空間中のある点（ベクトル）とします．境界超平面は，下の式で表される超平面のひとつです．

$$\boldsymbol{w}^T\boldsymbol{x} + b = 0. \tag{1}$$

ここで，$\boldsymbol{w}$ は重みベクトル，b はバイアス項とよばれます．集合に属するあるベクトル $\boldsymbol{x}_i$ と境界超平面の距離はマージンとよばれ，つぎのように表されます．

$$\frac{|\boldsymbol{w}^T\boldsymbol{x}_i + b|}{\|\boldsymbol{w}\|}. \tag{2}$$

(1) 式で表される超平面は，$\boldsymbol{w}$ と b に共通の定数をかけても同じものになります．そこで，次のような制約を導入します．

$$\min_i |\boldsymbol{w}^T\boldsymbol{x}_i + b| = 1. \tag{3}$$

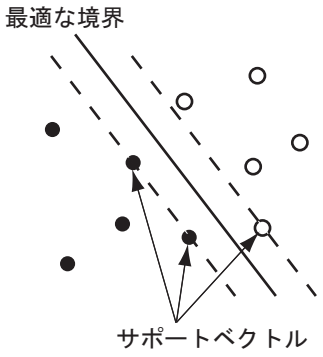

図 1: サポートベクタマシンによる最適な境界

最適な境界は，(2) 式の最小値を最大にするものです．(3) 式の制約により，この最大化は $1/\|\boldsymbol{w}\|^2 = 1/\boldsymbol{w}^T\boldsymbol{w}$ の最大化に帰着されます．すなわち，この最適化は

$$\begin{aligned} &\textit{minimize} \quad \boldsymbol{w}^T\boldsymbol{w} \\ &\textit{subject to} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 \end{aligned} \tag{4}$$

というものになります．ここで y_i は $\boldsymbol{x}_i$ が一方の集合に属するとき 1，もう一方の集合に属するとき -1 とします．もしも，この境界が各集合の点を正確に分離するならば，つねに $y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 0$ となります．

このような条件付き最適化は，前にも出てきた Lagrange の未定乗数法で解くことができます．この方法では，下のように評価関数を定義します．

$$L(\boldsymbol{w}, b, \alpha_i) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \sum_i \alpha_i[y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) - 1], \tag{5}$$

ここで $\alpha_i \geq 0$ は未定乗数です．$\boldsymbol{w}$ と b が最適値になるとき，

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{w}} &= \boldsymbol{w} - \sum_i \alpha_i y_i \boldsymbol{x}_i \\ \frac{\partial L}{\partial b} &= -\sum_i \alpha_i y_i \end{aligned} \tag{6}$$

はいずれも 0 となります．そこで，(6) 式の微分を 0 とおくと，

$$\boldsymbol{w} = \sum_i \alpha_i y_i \boldsymbol{x}_i \tag{7}$$

$$\sum_i \alpha_i y_i = 0 \tag{8}$$

が得られます．ここで (5) 式から

$$L(\boldsymbol{w}, b, \alpha_i) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \sum_i \alpha_i y_i \boldsymbol{w}^T\boldsymbol{x}_i - b\sum_i \alpha_i y_i \sum_i \alpha_i. \tag{9}$$

となるので，この式に (7) 式と (8) 式を代入すると

$$\begin{aligned} L(\boldsymbol{w}, b, \alpha_i) =& \frac{1}{2}\left(\sum_i \alpha_i y_i \boldsymbol{x}_i\right)^T \left(\sum_j \alpha_j y_j \boldsymbol{x}_j\right) \\ & - \sum_i \alpha_i y_i \left(\sum_j \alpha_j y_j \boldsymbol{x}_j\right)^T \boldsymbol{x}_i + \sum_i \alpha_i \\ =& -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j + \sum_i \alpha_i \end{aligned} \tag{10}$$

となります．(5) 式の第２項の影響は最小にしなければならず，かつ L は最大にしなければなりません．したがって，この最適化問題は次の２次計画問題に帰着されます．

$$\begin{aligned} &\textit{maximize} \quad -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j + \sum_i \alpha_i \\ &\textit{subject to} \quad \sum_i \alpha_i y_i = 0, \alpha_i \geq 0. \end{aligned} \tag{11}$$

２次計画問題をコンピュータで解くためには，多数市販されているプログラムパッケージが利用できます．

ソフトマージン

上のような議論は，線形分離可能な問題のみに適用されます．２つの集合が線形分離でない場合，２つの集合を完全に分離する超平面は存在しません．

線形分離でない場合に対応する手法のひとつが，ソフトマージン法とよばれるものです．この方法は，(4) 式の制約を次のものにおきかえる，というものです．

$$\textit{subject to} \quad y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b) \geq 1 - \xi_i. \tag{12}$$

ここで，ξ_i はスラック変数とよばれる正の値で，分離誤りを許す程度を表すものです．この置き換えは，図 2 のように，それぞれの集合に属する点が，境界を挟んでそれぞれ反対側のある限定された範囲に存在することを許す，ということを意味しています．この場合には，例えば

$$\textit{minimize} \quad \boldsymbol{w}^T \boldsymbol{w} + C \sum_i \xi_i \tag{13}$$

のような最適化関数が提案されています．上の式の第２項は分離誤りの罰則（ペナルティ）項で，定数 C は，罰則項の影響の度合いを調整しています．

カーネル法

ソフトマージン法は線形分離な枠組みの中でサポートベクタマシンを拡張するものでした．これに対して，以下で説明するカーネル法は，完全に非線形な「曲がった」境界を見つける方法です．

カーネル法の基本的アイデアは，元の空間自身をより次元の高い空間に変形することです．例えば，前回あげた線形非分離問題をみてみましょう（図 3(a)）．この２次元空間を，図 3(b) のような３次元空間に変形すれば，●の点と○の点は線形分離となります．

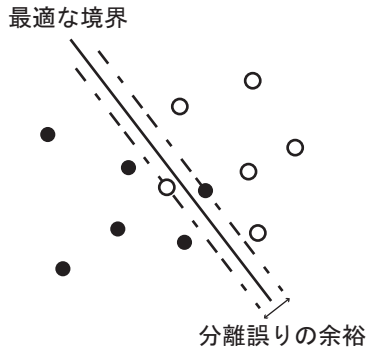

図 2: ソフトマージン

高次元空間への変換を Φ で表すことにします．変換された空間では，距離が定義されていないと，２つの集合を「最適に」分離する超平面を見つけることができません．また変換後の空間での距離は，元の空間での距離と関係がある（つまり，元の空間で「遠く離れた」２点は，変換後の空間でも「遠く離れている」）ほうが好都合です．このような条件を満たす変換を定義するために，点 $\boldsymbol{x}$ と点 $\boldsymbol{x}'$ の組に対してカーネル関数 $K(\boldsymbol{x}, \boldsymbol{x}')$ を考えます．カーネル関数は，

$$K(\boldsymbol{x}, \boldsymbol{x}') = \Phi(\boldsymbol{x})^T \Phi(\boldsymbol{x}'). \tag{14}$$

という関係を満たすものです．この式は，カーネル関数が，変換 Φ で変換された高次元空間で測られた距離に相当するものであることを意味しています．この結果，マージンを通常の距離のかわりにカーネル関数を用いて測り，このマージンをもとに最適化を行なって分離超平面を求めると，もとの空間では，この境界は「曲がった」境界になります．変換後の境界は

$$\boldsymbol{w}^T \Phi(\boldsymbol{x}) + b = 0. \tag{15}$$

で表されます．(7) 式を，$\boldsymbol{x}$ を $\Phi(\boldsymbol{x})$ におきかえたうえで (15) 式に代入すると，

$$\sum_i \alpha_i y_i \Phi(\boldsymbol{x}_i^T) \Phi(\boldsymbol{x}) + b = \sum_i \alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b = 0 \tag{16}$$

となります．(11) 式の最適化関数も，変換後の形は $\boldsymbol{x}_i^T \boldsymbol{x}_j$ を $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ におきかえることで得られます．このことは，境界を求めるのに必要な計算はすべて $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を用いて可能で，変換後の空間や変換 Φ が実際にどういうものかは，知る必要はない，ということを示しています．

正定値の二次形式となっているカーネル関数 K は，(14) 式の条件を満たすことが知られています．このようなカーネル関数の例には，つぎのようなものがあります．

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^T \boldsymbol{x}' + 1)^p \quad \text{（多項式カーネル）} \tag{17}$$

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{\sigma^2}\right) \quad \text{（ガウシアンカーネル）} \tag{18}$$

経験リスクと期待リスク

経験リスクとは，分離したい集合内の既知の点のうち，誤った分類をされる点の割合をさします．「最適」な境界を求めるのに，本当に最小にしなければならないのはこの経験リスクではありません．本当

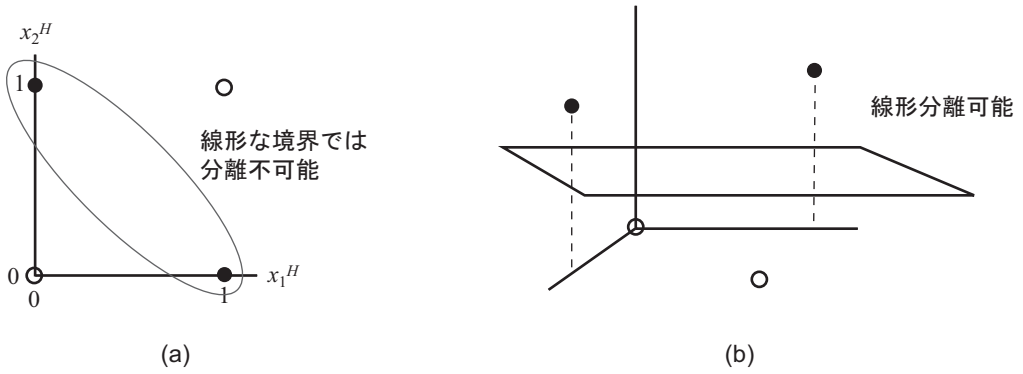

図 3: 高次元空間への変換

に最小化しなければならないのは，各集合の（既知のものも未知のものも含めた）全ての点のうち，誤った分類をされる点の割合です．こちらの割合のほうは期待リスクといいます．

線形分離可能な問題の場合，既知の点を線形分離可能なわけですから，経験リスクをゼロにする分離超平面が存在する，ということになります．「マージンが最大になる境界を見つける」というサポートベクタマシンの考え方は，経験リスクをゼロにする分離超平面のなかから，期待リスクを最小にするものを選ぶ，ということに相当します．