カーネル法入門
―正定値カーネルによるデータ解析―

4.2 サポートベクターマシンの拡張

茨城大学工学部情報工学科
08T4038G　篠塚 晃一

4.2.1 多クラス識別問題へのサポートベクターマシンの拡張

- 多クラスの識別
  - クラスラベルを $Y \in \{1, 2, \ldots, L\}$ によって表し、$(X_1, Y_1), \ldots, (X_N, Y_N)$を$(X, Y)$ のサンプルとするとき、 識別関数 $h: X \to \{1, 2, \ldots, L\}$を構成する。

4.2.1 多クラス識別問題へのサポートベクターマシンの拡張

- SVMを多クラスの識別問題に適用する方法は2通りの方針がある。
  - マージン最大化基準を多クラスに拡張することにより、SVMの多クラス版を構成する。
  - 多クラス識別問題を複数の2クラス識別問題に分割し、それぞれの識別問題を通常のSVMによって解き、最終的に結果をまとめる。

4.2.2 多クラスサポートベクターマシン

- SVMや類似のマージン最大化識別器を多クラスに拡張する試みはさまざまなものがある。
- ここでは、ユークリッド空間 R^m のデータに対する線形識別器についてのCrammer and Singer(2001)の方法を簡単に紹介する。

Crammer and Singer(2001)の方法

- L個の線形関数 $w_l^T x$ を用意し、次の式によりLクラスの識別関数を構成する。

$$h(x) = \arg\max_{l=1,\ldots,L} w_l^T x$$

- 線形関数 $w_l^T x$ に対し、データ点 (X_i, Y_i) のマージンを以下の式により定義する。

$$\mathrm{Margin_i} = w_{Y_i}^T X_i - \max_{l \neq Y_i} w_l^T X_i$$

Crammer and Singer(2001)の方法

- すべてのデータが識別可能な場合は、以下の条件によりスケールを固定する。

$$w_{Y_i}^T X_i - w_l^T X_i \geq 1 \quad (l \neq Y_i, \forall i)$$

$\xi_i \geq 0$ の導入により
ソフトマージンに変換

- 最適化問題(主問題)を得る。

$$\min_{W,\xi} \frac{\beta}{2} \| W \|^2 + \sum_{i=1}^{N} \xi_i \quad subject\ \ to \quad w_{Y_i}^T X_i + \delta_{lY_i} - w_l^T X_i \geq 1 - \xi_i \ \ (\forall l, i).$$

Crammer and Singer(2001)の方法

双対問題を以下のように求める。

- Langrange双対関数は以下の式の (W,ξ) に関する最小値として与えられる。

$$L(W,\xi,\eta)=\frac{\beta}{2}\|W\|^2+\sum_{i=1}^{N}\xi_i+\sum_{i=1}^{N}\sum_{l=1}^{L}\eta_{il}\{(w_l-w_{Y_i})^T X_i-\delta_{lY_i}+1-\xi_i\}$$

$$(\eta_{il}\geq 0,\forall l,i)$$

$e_r=(0,\ldots,1,\ldots,0)^T\in R^L$ を用いてL次元ベクトル変数

$\tau=e_{Y_i}-\eta_i(\eta_i=(\eta_{i1},\ldots\eta_{iL}))$ を導入する

- 双対問題

$$\min_{\tau}-\frac{1}{2}\sum_{i,j=1}^{N}(X_i^T X_j)\tau_i^T\tau_j+\beta\sum_{i=1}^{N}\tau_i^T e_{Y_i},\ \text{subject to}\ \tau_i\leq e_{Y_i},\sum_{l=1}^{L}\tau_{il}=1\ (\forall i)$$

Crammer and Singer(2001)の方法

- τ^*をこの最適解とすると、識別関数は以下のようになる。

$$h(x) = \arg\max_{l=1,\dots,L} \sum_{i=1}^{N} \tau_{il}^* (X_i^T x)$$

4.2.3 2クラス識別器の組み合わせ
a. 2クラス識別問題への分割法

- one-vs-rest
  - 「クラス1対その他」,…,「クラスL対その他」のL個の2クラス識別問題に分割する。
- one-vs-one
  - 任意の2つのクラスの組(i,j)に対しクラスi対クラスjの識別器を構成し、合計L(L-1)/2個の2クラス識別問題を用いる。
- ECOC
  - 合計M個の2クラス識別問題に分割するために、L個の各クラスに対してMビットの2値符号を割り当てる。M個の各2クラス識別器は、L個のクラスを+1と-1の符号に従って2つに分割し、その符号をクラスラベルとして識別器を構成する。

4.2.3 2クラス識別器の組み合わせ
a. 2クラス識別問題への分割法

表4.1 : ECOCの符号の例。
4クラスの識別を6個の
2クラス識別問題に
分割している。

class	f_1	f_2	f_3	f_4	f_5	f_6
C_1	-1	-1	-1	1	1	1
C_2	-1	1	1	-1	-1	1
C_3	1	-1	1	-1	1	-1
C_4	1	1	-1	-1	1	1

表4.2 : 4クラス識別問題の符号。
左: one-vs-rest, 右: one-vs-one.

class	f_1	f_2	f_3	f_4
C_1	+1	-1	-1	-1
C_2	-1	+1	-1	-1
C_3	-1	-1	+1	-1
C_4	-1	-1	-1	+1

class	f_1	f_2	f_3	f_4	f_5	f_6
C_1	+1	+1	+1	0	0	0
C_2	-1	0	0	+1	+1	0
C_3	0	-1	0	-1	0	+1
C_4	0	0	-1	0	-1	-1

4.2.3 2クラス識別器の組み合わせ
b. 識別器の組み合わせ法

- Hamming復号法
  - ECOCを用いた場合の、クラスを決定するための代表的な方法。
  - クラスlのECOC符号を $W_l = (W_{l1}, \ldots, W_{lM})$ とするとき以下の式を識別関数に用いる。

$$h(x) = \arg\max_l \sum_{a=1}^{M} W_{la} f_a(x)$$

4.2.4 構造化出力への拡張

- クラス識別問題
  - ラベル間に構造は入っていない。
- 応答変数Yがラベル系列やツリーなどの構造を持つ場合
  - 構造に応じた識別関数の構造が必要
- このような問題に対してもSVMの拡張がなされている。

4.2.4 構造化出力への拡張

- データ$(X_1,Y_1),\ldots,(X_N,Y_N)\in X\times Y$に対し、$X\times Y$上の特徴写像$\Phi: X\times Y \to R^m$を用意し、回帰写像$h: X\to Y$を以下の式により定義する。

$$h(x)=\arg\max_{y\in Y} w^T\Phi(x,y)$$

- マージンを以下の式で定義する。

$$w^T\Phi(X_i,Y_i)-\max_{y\neq Y_i} w^T\Phi(X_i,y)$$

- マージン最大化基準による最適パラメータwは以下の式の解として求められる。

$$\min_{w,\xi}\frac{\beta}{2}\|w\|^2+\sum_{i=1}^{N}\xi_i \ \text{ subject to } \ w^T\Phi(X_i,Y_i)+\delta_{yY_i}-w^T\Phi(X_i,y)\geq 1-\xi_i \ \ (\forall i, y\in Y)$$

4.2.5 その他の方法

- Herbrichet al., 2000
  - ランク付けの問題にマージンの最大化のアイデアを適用
- Keerthiet al., 2001
  - 効率的解法SMOの改良が考えられている
- Chapelle,2007; Joachims,2006; Shalev-Shwartz et al., 2007
  - 双対問題を使わずに主問題のまま解くアプローチ
- Mavroforakis and Theodoridis, 2006
  - SVMのソフトマージンを各クラスの縮小凸結合間の距離として捉え直し、高速な幾何学的最適化アルゴリズムを用いる方法

4.2.5 その他の方法

- Langrangean SVM (Mangasarian and Musicant, 2001)
  - マージン最大化基準をSVMと異なる目的関数で実現する
- Tax and Laskov(2003), LaSVM(Bordes et al.,2005)
  - データがあらかじめすべて与えられるのではなく、ひとつずつ与えられるたびに推定量を更新していくオンライン学習をSVMに適用する方法
- Graf et al., 2005; Zanni et al., 2006;
  - 並列計算によって計算時間を削減する試み