

SVMにおけるカーネル行列のSDPによる最適化について

02203090 筑波大学システム情報工学研究科 *反町宏明 SORIMACHI Hiroaki
01703540 筑波大学社会工学系 吉瀬章子 YOSHISE Akiko

1 はじめに

本発表では、パターン認識手法のひとつであるサポートベクターマシン(SVM)で用いられるカーネル行列に対して様々な最適化モデルを示すとともに、テストデータによってその有効性を検証する。

SVMでは、基本的に線形な識別関数によるデータの分類を行うが、より精緻な切り分けを実現するため、データが記述される入力空間 \mathcal{X} から適当な写像 ϕ によって(時に高次の)特徴空間 \mathcal{F} に写し、 \mathcal{F} において最適な識別関数を求める。このとき写像 ϕ は、 \mathcal{X} から陽な写像としては与えられず、 \mathcal{X} 上のデータ x_1, x_2 に対する \mathcal{F} 上の内積 $\langle \phi(x_1), \phi(x_2) \rangle$ が x_1 と x_2 の関数 $K(x_1, x_2)$ で与えられると仮定し、 K で代替される。この関数 K をカーネル関数と呼ぶ。現在多く用いられているカーネル関数には、線形カーネル、多項式カーネル、ガウスクーネルなどがある。

カーネル関数の選択は、SVMにおいて重要な要素であるにもかかわらず、実際には経験的に行われていることが多い。ここではカーネル関数の代表値を与えるカーネル行列が半正定値行列であることに着目した論文[1]をもとに、様々な尺度、制約の下での最適化モデルを提示する。発表では、各カーネルの識別への影響に関する計算機実験結果を報告する予定である。

2 カーネル行列と最適化モデル

この節では、本稿を通じて用いるカーネル行列の基本的な性質と、各種の尺度に基づく最適化モデルについて述べる。既与の整数 N に対して、添え字集合 $S_N = \{1, 2, \dots, N\}$ を定義する。 N 個の \mathcal{X} のデータの内、1あるいは-1の識別値が既知の訓練データが T_r 個 ($T_r \leq N$) あり、データと識別値を組にした集合 $\{(x_i, y_i) \in \mathcal{X} \times \{-1, 1\} : (i \in S_{T_r})\}$ が与えられているとする。

一般に入力空間 \mathcal{X} から特徴空間 \mathcal{F} への写像 ϕ が与えられているとき、カーネル行列 K は、 $K_{ij} = \phi(x_i)\phi(x_j)$ ($i, j \in S_N$) として与えられる。定義より明らかに K は半正定値行列であるが、さらに以下の特徴付けが得られる。

命題 2.1 カーネル関数が入力データ x_i ($i \in S_N$) 上でのみ定義されている関数であるとするれば、行列 K がカーネル行列であるための必要十分条件は K が対称半正定値行列であることである。

以下では、対称行列 K が半正定値行列であることを $K \succeq O$ で表す。また、訓練データに対応するカーネル行列の部分行列を K_{T_r} で表すことにする。

2.1 マージンを尺度とする最適化モデル

特徴空間 \mathcal{F} が \mathbb{R}^m であるとする。このとき訓練データの集合に対するハードマージンモデル (P_H) は、変数 $(w, b) \in \mathbb{R}^{m+1}$ に関する以下の最適化問題で与えられる。

$$(P_H): \min\{w^T w : y_i(w^T \phi(x_i) - b) \geq 1 \ (i \in S_{T_r})\}$$

表記の便宜上、記号 $G_{T_r}(K)_{ij} = y_i y_j K_{i,j}$ ($i \in S_{T_r}$) を導入する。 (P_H) は (w, b) に関する凸計画問題であることから、双対問題を考えることにより、最適解 w_H について以下が成り立つ。

$$w_H^T w_H = \max\{2\alpha^T e - \alpha^T G_{T_r}(K)\alpha : \alpha \geq 0, \alpha^T y = 0\} \quad (1)$$

(1)と命題 2.1 より、任意の対称半正定値行列 K に対して、 K に対応するカーネル関数 ϕ が存在し、問題 (P_H) の最適値が定まることがわかる。この性質をもとに、よりよい最適値 $w_H^T w_H$ を与える行列 K の導出を試みる。

式 (1) の右辺を、 (α, K) に対する最適化問題であることみなすと、Schur-Complement の補題を用いることにより以下の (ν, λ, K) 上の最適化問題 (SDP_H) として表わすことができる [1]。

$$(SDP_H): \min\{t : \text{trace}(K) = 1, M_H(\nu, \lambda, K) \succeq 0\}$$

ここで制約 $\text{trace}(K) = 1$ は自明な解 $K = O$ を避けるために加えた制約であり、また $M_H(\nu, \lambda, K)$ は以下で与えられる対称行列である。

$$M_H(\nu, \lambda, K) = \begin{pmatrix} K & O & O & O \\ O & G_{T_r}(K) & e + \nu + \lambda y & O \\ O & (e + \nu + \lambda y)^T & t & O \\ O & O & O & \text{diag}(\nu) \end{pmatrix}.$$

e は要素がすべて 1 のベクトルを意味する。 $M_H(\nu, \lambda, K)$ を各変数に関する行列の線形結合で表すことにより後述の (2) の形式に変形でき、問題 (SDP_H) が半正定値計画問題 (SDP) であることがわかる。

同様に訓練データに対する l_1 ノルムのソフトマージンモデル (P_{S1})

$$(P_{S1}): \min\{w^T w + c \sum_{i \in S_{T_r}} \xi_i : y_i(w^T \phi(x_i) - b) \geq 1 - \xi_i, \xi_i \geq 0 \ (i \in S_{T_r})\},$$

および訓練データに対する l_2 ノルムのソフトマージンモデル (P_{S2})

$$(P_{S2}): \min\{w^T w + c \sum_{i \in S_{T_r}} \xi_i^2 : y_i(w^T \phi(x_i) - b) \geq 1 - \xi_i \ (i \in S_{T_r})\}$$

(ともに c は正のパラメータ) に対しても上述とほぼ同じ議論が適用でき、それぞれの問題に対する半正定値計画問題 SDP_{S1} , SDP_{S2} を得ることができる。

2.2 アラインメントを尺度とする最適化モデル

Tr 個の入力データに対する識別値 y が既知の場合、目標となる(訓練データに対応する)カーネル行列は $K_{Tr} = yy^T$ である。このことから、行列 K_{Tr} と yy^T との仰角のコサインを与えるアラインメント

$$A(K) = \frac{\langle K_{Tr}, yy^T \rangle}{\sqrt{\langle yy^T, yy^T \rangle} \sqrt{\langle K_{Tr}, K_{Tr} \rangle}}$$

をカーネル行列 K の尺度と考えることができる。 $A(K)$ の最大化を直接にモデル化しようとすると、 K に関する2次式 $\langle K, K \rangle$ を含む制約をおく必要があるが、2.1節と同様に Schur-Complement の補題を用いることにより、以下のような行列の組 (A, K) に関するSDPとして定式化できる[1].

$$(SDP_{AI}) : \max\{\langle K_{Tr}, yy^T \rangle : M_{AI}(A, K) \succeq 0\}$$

ここで $M_{AI}(A, K)$ は以下で与えられる対称行列である。

$$M_{AI}(A, K) = \begin{pmatrix} A & K_{Tr}^T & O & O \\ K_{Tr} & I & O & O \\ O & O & 1 - \text{trace}(A) & O \\ O & O & O & K \end{pmatrix}$$

2.3 カーネル行列の構成法

前節までは、カーネル行列を一般的な対称正定値行列として定式化を行ったが、ここでは、 K を多項式カーネル行列等の、 p 個の既存のカーネル行列 $K_l (l = 1, 2, \dots, p)$ の線形結合 $K = \sum_{l=1}^p \mu_l K_l$ で与えらる。カーネル行列を特徴付ける正定値性 $K \succeq O$ は、 $\mu_l (l = 1, 2, \dots, p)$ に関する以下の制約 (C1) に置き換えることができる。

$$(C1) : K = \sum_{l=1}^p \mu_l K_l \succeq O.$$

ただし正定値行列性に関する制約は、最適解を求める際に多大な計算コストを必要とする場合がある。そこで比較検討のため、以下のような正定値性に関する十分条件を与える制約 (C2) も同時に取り扱うことにする。

$$(C2) : K = \sum_{l=1}^p \mu_l K_l, \mu_l \geq 0 (l = 1, 2, \dots, p).$$

次節にまとめるように、制約 (C2) に置き換えることにより、計算量が非常に少ない定式化が可能になるモデルもある。

3 各最適化モデルの比較

2.1, 2.2節で述べた最適化モデルに、2.3節で述べた制約を組み合わせ得られる8つの定式化について、最適化モデルとして比較を行う。

すべてのモデルは、半正定値計画問題 (SDP) として定式化される。

(SDP) の標準形を、 $m \times m$ の行列 $F_i (i = 0, 1, \dots, n)$ と n 次元ベクトル c が与えられているとき、ベクトル $x \in \mathbb{R}^n$ に対する最適化問題

$$\min \left\{ c^T x : F_0 + \sum_{i=1}^n F_i x_i \succeq O \right\} \quad (2)$$

として与えるならば、各最適化モデルでのSDPのサイズを示す (m, n) の組は、以下の表としてまとめられる。

特にモデル (SDP_{AI}) については、使用するカーネル $K_i (i = 1, 2, \dots, p)$ が1次独立であれば、下記の表の値が示すように、定式化されたSDPのサイズを大幅に縮小することができる。さらにこのモデルに制約 (C2) を加えた場合は、 $p \times p$ 行列の線形方程式1回解く手間で解が得られることがわかる。

表：各モデルのSDPのサイズ
(括弧内は $K_i (i = 1, 2, \dots, p)$ が1次独立の場合)

モデル/制約	(C1)	(C2)
(SDP_H)	$m = p + Tr + 2$ $n = N + 2Tr + 3$	$m = p + Tr + 2$ $n = p + 2Tr + 3$
(SDP_{S1})	$m = p + 3Tr + 2$ $n = N + 3Tr + 3$	$m = p + 2Tr + 2$ $n = p + 3Tr + 3$
(SDP_{S2})	$m = p + Tr + 2$ $n = N + 2Tr + 3$	$m = p + Tr + 2$ $n = p + 2Tr + 3$
(SDP_{AI})	$m = N + 2Tr + 1$ $n = N + 2Tr + 1$ ($m = p$) ($n = N + p + 1$)	線形方程式 $O(p^3)$

4 各カーネル行列を用いた計算機実験

3節で示した8つのモデルに対して、既存のカーネル行列として多項式、線形、ガウシアン3つを取り上げ、計算機実験を行った。OSはSunOS5.8上を使用し、SDPを解くソフトウェアとしてSeDuMi105R4[2]を、その他のプログラムはMATLAB6.0.0.88.R12を使用した。発表では、SVMのライブラリから抽出した共通なテストデータを用いて、各最適化モデルで得られたカーネル行列に対するチューニング効果について報告する予定である。

References

- [1] G. Lanckriet, N. Cristianini, L.E. Ghaoui and M.I. Jordan. Learning the Kernel Matrix with Semi-Definite Programming. Technical Report UCB/CSD-02-1206, Computer Science Division, University of California, Berkeley, California, 94720. October, 2002.
- [2] J. Sturm. SeDuMi105R4
<http://fewcal.kub.nl/sturm/software/sedumi.html>