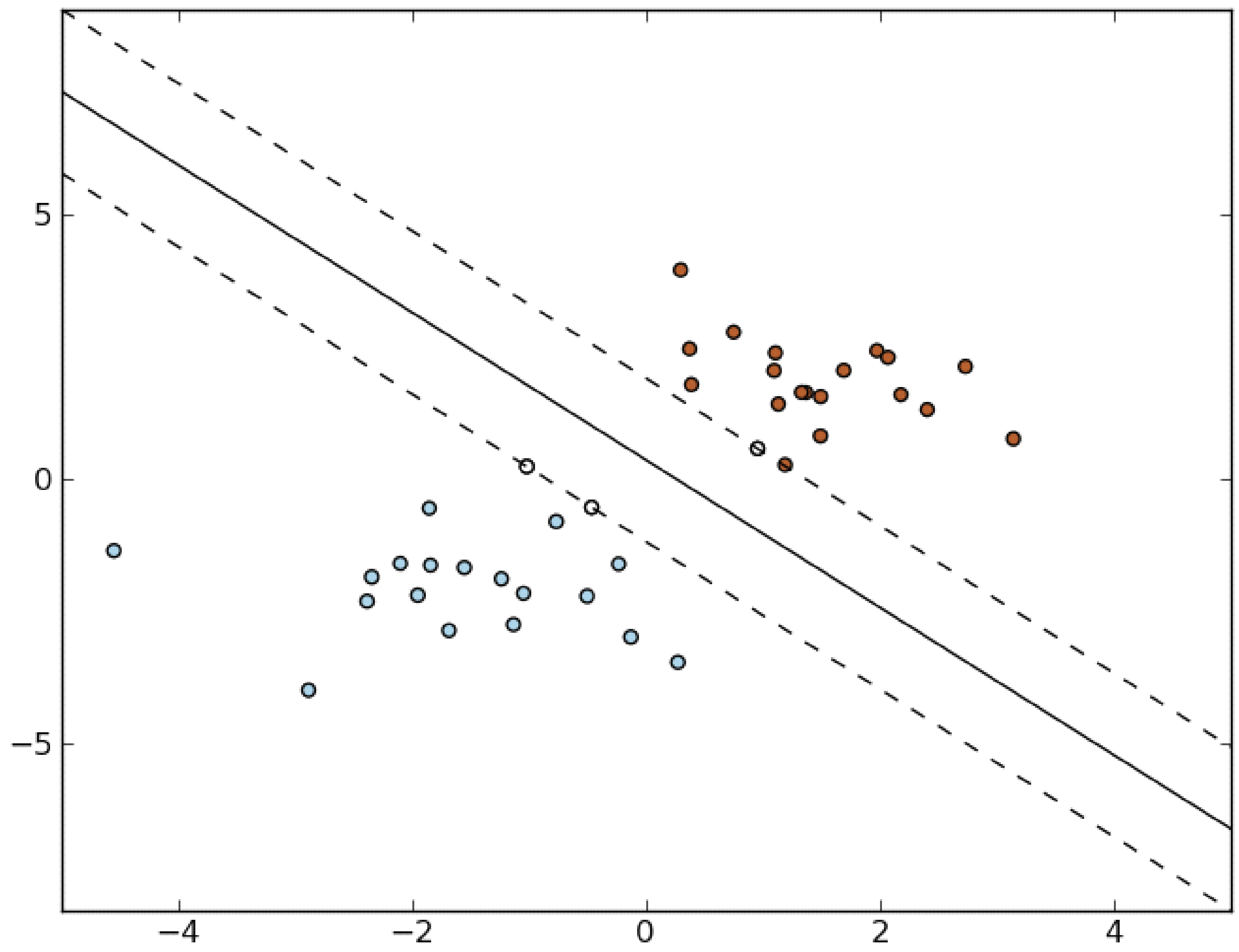
カーネル法入門
-正定値カーネルによるデータ解析-

3.カーネル法の実際
3.4 サポートベクターマシンの基本事項

08T4082A　西野太樹

マージン最大化基準による線形識別

- サポートベクターマシン(**SVM**)
  - 2値の線形識別問題において、識別超平面から最も近い点(サポートベクトル)までの距離を最大にする基準(マージン最大化基準)で超平面を選択する手法


5
0
−5
−4
−2
0
2
4

マージンの計算（１）

- 例： $f(x)=\mathrm{sgn}(w^T x+b)$

(w, b)に λ >0 を掛けても識別関数は変化しないので

$$\begin{cases} \min(w^T X_i + b) = 1, & i : Y_i = +1 \\ \max(w^T X_i + b) = -1, & i : Y_i = -1 \end{cases}$$

と仮定する.

上のmin,maxを達成するデータ点（サポートベクトル）をそれぞれ X_*^+, X_*^-

と置くと, $w^T(X_*^+ - X_*^-) = 2$ であるので, それぞれのクラスのサポート

ベクトル間の距離は $\frac{2}{\|w\|}$ で与えられる.

マージンの計算（2）

したがって，マージン最大化を達成するパラメータは

$$\max \frac{1}{\|w\|} \quad subject\ \ to \quad \begin{cases} w^T X_i + b \geq 1, & i : Y_i = +1, \\ w^T X_i + b \leq -1, & i : Y_i = -1 \end{cases}$$

の解となる．最小化問題に変換し、条件をまとめると

$$\min_{w,b} \|w\|^2 \quad subject\ \ to \quad Y_i(w^T X_i + b) \geq 1 \quad (\forall_i = 1, \ldots, N)$$

の解を求めることになる．

正則化（1）

- 正則化
  - ノルム付きベクトル空間Vの部分集合A上の最適化問題

$$\min_{w \in A} \Omega(w)$$

    を考えたとき，その解一意的でない場合に，これを一意にする操作を正則化という.

  - 最適化の解が理論的に一意であっても，数値的に安定しない場合に用いられる.

正則化（2）

- **Tikhonov正則化**
  - 先の最適化問題に正則化項を加える

$$\min_{w \in A} \Omega(w) + \lambda \|w\|^2$$

  - λ は正則化係数と呼び，Ωと正則化項のバランスを決める
  - 比較的に弱い仮定のもと、一意で安定

ソフトマージン（１）

- SVMにおける線形識別可能性の仮定 $Y_i(w^T X_i + b) \geq 1$

  を緩和し，

  $$Y_i(w^T X_i + b) \geq 1 - \xi_i \quad （\xi_i \geq 0） \quad （i = 1, \ldots, N）$$

  と置き換える．ξ_i が大きくならないよう目的関数も

  $$\min_{w, b, \xi_i} \|w\|^2 + C \sum_{i=1}^{N} \xi_i \quad subject \quad to \quad \begin{cases} Y_i(w^T X_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} (\forall_i)$$

  と変更する．係数Cは，マージン最大化の項と完全な線形識別からの緩和を表す項との釣り合いを決める．

ソフトマージン（２）

- 変更した目的関数

$$\min_{w,b,\xi_i} \|w\|^2 + C\sum_{i=1}^{N} \xi_i \quad subject \quad to \quad \begin{cases} Y_i(w^T X_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} (\forall_i)$$

の解による識別関数を<u>ソフトマージン**SVM**</u>と呼ぶ.

- また，変更前の目的関数

$$\min_{w,b} \|w\|^2 \quad subject \quad to \quad Y_i(w^T X_i + b) \quad (\forall_i = 1, \ldots, N)$$

の解による識別関数を<u>ハードマージン**SVM**</u>と呼ぶ.

ソフトマージンと正則化

- ソフトマージンSVMの最適化を正則化問題とすると

$$\min_{w,b,\xi_i} \sum_{i=1}^{N} \xi_i + \lambda \|w\|^2 \quad subject\ \ to\ \ \xi_i \geq (1 - Y_i(w^T X_i + b))_+ \ (\forall_i)$$

と書きなおすことができる．さらに，条件より最小となるのは $\xi_i = (1 - Y_i(w^T X_i + b))_+$ の場合であるから

$$\min_{w,b} \sum_{i=1}^{N} (1 - Y_i(w^T X_i + b))_+ + \lambda \|w\|^2$$

という制約条件のない最適化問題に書き直せる．

SVMのカーネル化（１）

$$f(x) = \mathrm{sgn}(\langle h, \Phi(x) \rangle_{\mathrm{H}} + b) = \mathrm{sgn}(h(x) + b)$$

- 上記の線形識別関数から，ソフトマージンSVMの目的関数

$$\min_{h,b,\xi_i} \|h\|_{\mathrm{H}}^2 + C\sum_{i=1}^{N} \xi_i \quad subject \quad to \quad \begin{cases} Y_i(\langle h, \Phi(X_i) \rangle + b) \geq 1 - \xi_i, \\ \xi_i \geq 0 \end{cases}$$

を得る．また正則化問題として

$$\min_{h,b} \sum_{i=1}^{N} (1 - Y_i(\langle h, \Phi(X_i) \rangle + b))_+ + \lambda \|h\|_{\mathrm{H}}^2$$

と表すこともできる．

SVMのカーネル化（２）

- カーネル法と同様に先の関数の最小値は

$$h = \sum_{i=1}^{N} c_i \Phi(X_i)$$

の形の元により達成される．このとき

$$\|h\|^2 = \sum_{i,j=1}^{N} c_i c_j k(X_i, X_j),$$

$$\langle h, \Phi(X_i) \rangle_{\mathrm{H}} = \sum_{j=1}^{N} c_j k(X_i, X_j)$$

であるから

SVMのカーネル化（３）

$$\min_{c_i, b, \xi_i} \sum_{i,j=1}^{N} c_i c_j k(Xi, Xj) + C \sum_{i=1}^{N} \xi_i$$

$$subject\ \ to\ \ \begin{cases} Y_i(\sum_{j=1}^{N} k(X_i, X_j)c_j + b) \geq 1 - \xi_i, \\ \xi_i \geq 0 \end{cases} (\forall_i)$$

という最小化問題によってパラメータを決めればよい．

SVMの応用例

- （例）手書き数字認識への応用
  - LeCun et al.(2001)
  - データは「０」～「９」までの手書き数字のデータベースMNISTより取得
  - 60000画像の訓練データでパラメータ推定
  - 10000画像のテストデータで実験

表 3.1　手書き数字データベース MNIST への応用
(LeCun et al. (2001) より抜粋)

Method	Test error (%)
k-NN Euclid	5.00
40PCA + quadratic	3.30
RBF network (1000)	3.60
LeNet-4	1.10
LeNet-5	0.95
SVM poly4	1.10
RS-SVM poly5	1.00

SVMの性質（1）

- カーネル法
  - 変換したデータの線形結合 $\sum_{i=1}^{N} c_i k(\cdot, X_i)$ の形で解を表現
  - 最適化すべき目的関数はグラム行列 $k(X_i, X_j)$ で表記

- 凸最適化
  - 識別関数を求めるための最適化問題は凸最適化の一つである2次計画であり，局所解の問題がなく容易

SVMの性質（2）

- 高次元の特徴空間
  – SVMはカーネルトリックを用いて，高次元の特徴空間を用いながらも計算効率を損ねない

- スパース表現
  – 最終的な識別関数の表現にデータすべてを用いず，サポートベクタルと呼ばれる部分集合だけど用いる

SVMの性質（3）

- 正則化
  - SVMは, ヒンジ損失関数を用いた正則化法と考えることができる. これはスプライン平滑化などと関連する.