サポートベクトルマシンを使った6パリティ問題の解決

T00K074F 堀内　亮介

1.　初めに

サポートベクトルマシン（Support Vector Machine, 以下 SVM）はパターン認識性能の優れた学習モデルの一つである．SVMは線形しきい素子を用いて 2 クラスのパターン識別を構成している．

本研究ではＳＶＭの認識能力の高さを利用して 6 パリティ問題に対して識別することが出来るかを実験するとともに，ＳＶＭの計算時間の問題に対して最大勾配法と最適勾配法を使用した方法を比較する．

2.　SVM

線形識別関数を次のように定義する

$$f(x) = sign(g(x)) \quad g(x) = w^t x + b \qquad (1)$$

ｘは入力ベクトル,ベクトルｗスカラーｂは識別関数を決定するパラメータである.学習データはｎ個与えられているとし，$x_i (i = 1,2,\ldots,n)$ と表す．これらのデータを 2 つのクラス χ_1, χ_2 に分離する．

学習データ x_i に関する教師信号を y_i とし，次のように定義する

$$y_i = \begin{cases} 1 & if x_i \in \chi_1 \\ -1 & if x_i \in \chi_2 \end{cases} \qquad (2)$$

パラメータｗとｂを求める問題は

制約条件：　$y_i(w^t x_i + b) - 1 \geq 0$　(3)

の下で目的関数：　$L(w) = \frac{1}{2}\|w\|^2$　(4)

を最小とするパラメータを求める問題となる.双対問題に帰着しαを導入すると

制約条件: $0 = \sum_{i=1}^{N} \alpha_i y_i \quad \alpha_i \geq 0, i = 1,\ldots,N$ (5)

の下で目的関数：

$$L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^t x_j \qquad (6)$$

という $\alpha_i (\geq 0) i = 1,\ldots,N$ に関する最適化問題となる．

3.　非線形な場合

非線形な場合は元の特徴ベクトルｘを非線形の写像 $\phi(x)$ によって変換し，$\phi(x)$ を新たなパターンとみなし変換後の空間において線形 SVM を適用する．非線形の場合には，非線形に写像した空間での二つの要素 $\phi(x_1)$ と $\phi(x_2)$ の内積が

$$\phi(x_1)^t \phi(x_2) = K(x_1, x_2) \qquad (7)$$

のように入力特徴 x_1 と x_2 のみから計算できるようにし，最適な非線形写像を構成する．

関数 K の例として

Gauss カーネル：

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \qquad (8)$$

がある．

4.　αを求めるアルゴリズム

- 最大勾配法

ステップ 1

$L(\alpha)$ を最大化するようなαを求めるためεを固定値とし次の更新式で更新していく

$$\tilde{\alpha}_i(t+1) = \alpha(t) + \varepsilon \frac{\partial L(\alpha_i)}{\partial \alpha_i} \qquad (9)$$

ステップ 2

更新したαを制約条件を満たすように正規化をする．

$$\alpha_i(t+1) = \tilde{\alpha}_i(t+1) - \frac{\sum_{i=1}^{N} \tilde{\alpha}_i(t+1) y_i}{y_i * N} \qquad (10)$$

この時 $\alpha_i(t+1) < 0$ ならば $\alpha_i(t+1) = 0$ とする．

ステップ 3

正規化した $\alpha_i(t+1)$ が十分収束したならば更新を終了する．

そうでなければステップ 1 に戻る．

- 最適勾配法

ステップ 1

$L(\alpha) = a\varepsilon^2 + b\varepsilon + c$ という関数があると仮定して，点 $\varepsilon_0, \varepsilon_1, \varepsilon_2$ を設定しこの連立方程式から $L(\alpha) = a\varepsilon^2 + b\varepsilon + c$ のεの極小値 $\varepsilon_{\min}$ を求

めて$\varepsilon_{\min}$を新しいεとし

$$\tilde{\alpha}_i(t+1)=\alpha(t)+\varepsilon\frac{\partial L(\alpha_i)}{\partial\alpha_i}$$ を更新する.

ステップ 2, ステップ 3

最急降下法と同じ.

5. 実験・結果

(1) 2 パリティ問題の実験

実験には Gauss カーネルを使用しσの値はσ=1 と設定する.

結果

表 1：2 パリティ問題の実験結果

SV	α	w	b
00	6.45856	-6.45856	0
01	6.45856	6.45856	0
10	6.45856	6.45856	0
11	6.45856	-6.45856	0

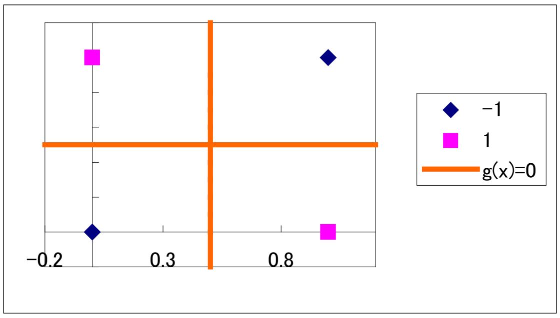

図 1：SVM で線形分離した図

(2) 6 パリティ問題に対しての実験

実験には Gauss カーネルを使用しσの値はσ=1 と設定する.

最大勾配法と最適勾配法を使いαが収束するまでの時間と計算回数を比較する.

結果

表 2：6 パリティ問題の実験結果

SV	α	w	b
0,0,0,0,0,0	5.12982	-5.12982	0
0,0,0,0,0,1	5.12982	5.12982	0
0,0,0,0,1,0	5.12982	5.12982	0
⋮	⋮	⋮	⋮
1,1,1,1,0,1	5.12982	5.12982	0
1,1,1,1,1,0	5.12982	5.12982	0
1,1,1,1,1,1	5.12982	-5.12982	0

(3) 最大勾配法と最適勾配法の比較

最大勾配法のεの値は 0.001, 0.01, 0.1

最適勾配法は 3 点の値を（0.001, 0.005, 0.01）,（0.01, 0.05, 0.1）,（0.05, 0.1, 0.5）として実験した

表 3：実行時間と収束するまでの回数の表

最大勾配法		最適勾配法	
ε＝0.001		ε＝0.001、0.005、0.01	
実行時間	25m31.546s	実行時間	38.390s
計算回数	1239235	計算回数	4436
ε＝0.01		ε＝0.01、0.05、0.1	
実行時間	3m47.671s	実行時間	42.171s
計算回数	185972	計算回数	4892
ε＝0.1		ε＝0.05、0.1、0.5	
実行時間	30.390s	実行時間	46.968s
計算回数	24799	計算回数	5450

6. 考察

- SVM が 6 パリティ問題のような高次元のパリティ問題に対しての認識性能を持っていることがわかった.
- 最適勾配法を使ったとき最大勾配法を使ったときよりも収束時間が小さくなりよりよい結果を出すことが出来た.
- 最大勾配法はεの値を変化させることによってαの値が収束するまでの時間が変わってしまうが, 学習データごとに最適なεを見つけるのは困難である.
- 最適勾配法は$\varepsilon_0,\varepsilon_1,\varepsilon_2$の点を変化させても最大勾配法と比べると余り収束時間が変化しなかったので, さまざまな学習データに対して安定した収束時間を期待できると考えられる.

参考文献

[1] T.Hastie, R.Tibshirani, J.Friedman: TheElements of Statistical Learning-Data Mining, Inference, and Prediction, Springer-Verlag, pp371-379 ,(2001).

[2] 前田英作: ″痛快！ サポートベクトルマシン ～古くて新しいパターン認識手法～″, 情報処理学会誌, Vol.42, No.7, (2001).

[3] 栗田多喜夫: サポートベクターマシン入門,(online),<http://www.neurosci.aist.go.jp/~kurita/lecture/svm/>（2004-9)