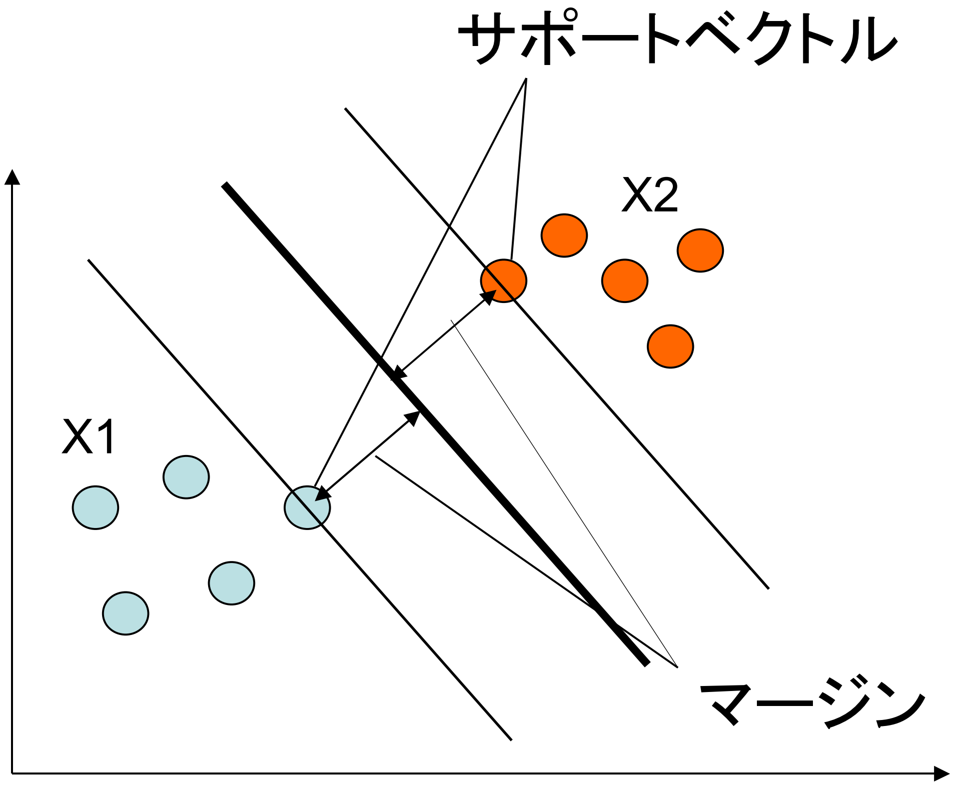
サポートベクターマシンの手書き数字認識への応用

宗久・丹沢研究室
T04KF011
大原　牧

- サポートベクターマシン（以下SVM)は,パターン認識性能の優れた2クラス判別の手法.
- SVM は現在最も強力な学習機械の一つであると認められている.
- 本研究では,SVMについて学習し,手書き数字認識へと応用していく.

- SVMでは,訓練サンプルの中で最も他クラスと近い位置のもの（サポートベクトル）を基準とし,そのユークリッド距離が最も大きくなるような場所に識別境界を定める. つまり,1つのクラスから他のクラスへの距離を最大にするようにする（マージン最大化）.


サポートベクトル
X2
X1
マージン

線形SVM

- 線形SVMとは,与えられた2クラスのデータを線形に分離することが可能である場合のSVM.

- 線形識別関数を,以下のように定義する.

  $f(x)=\mathrm{sign}(g(x))$

  ただし　$g(x)=w^t x+h$

- *sign*(*g*(*x*))は,*g*(*x*)＞0のとき1, *g*(*x*)＜0のとき-1をとる符号関数である.
- xは入力ベクトル,ベクトルwとスカラーhは識別関数を決定するパラメータとなっている.

- 学習データ数をn,学習データをx_i(i=1,2,….n)とする. 教師信号をy_iとし,以下のように定義する.

$$y_i = \begin{cases} 1 & if \quad x_i \in X1 \\ -1 & if \quad x_i \in X2 \end{cases}$$

ただし,X1,X2は学習データの属するクラス.

- y_i=1の学習データに対しては$g(x)$ ≧ 1 を, y_i=-1の学習データに対しては$g(x)$ ≦-1 を満たすと定義する.

- この定義により,マージン領域は $\dfrac{2}{\| w \|}$ で表される.

マージン最大化の基準に沿った識別関数を学習するために解くべき問題は，式（１），（２）の制約付き最適化問題に帰着される.

min $G(w) = \frac{1}{2}\|w\|^2$ （１）

s.t $y_i(w^t x_i + h) - 1 \geqq 0$ （２）

(1),(2)をLagrange 未定乗数法によって問題を書き直すと,以下のような双対問題を得る.

max $$L_D(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1,j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^t x_j$$

s.t $$0 = \sum_{i=1}^{N} \lambda_i y_i \qquad (\lambda i \geqq 0)$$

これを,最急降下法を使って最適解λを求め,パラメータw,hを得ることができる.

非線形SVM

- SVMでは,線形分離不可能な問題に対して,高次元写像空間での線形分離を適用する.

このようなSVMを,非線形SVMという.

- 非線形SVMの識別関数は次のように定義される.

$$f(x)=\mathrm{sign}(g(x))$$

ただし $g(x)=w^t\Phi(x)+h$

・Φ(x)は,元の特徴空間に存在する学習データのベクトルを，線形分離が可能となるように高次元空間へ変換する写像である.

$$\Phi(x)=(\Phi_1(x),\Phi_2(x),\ldots,\Phi_d(x))^t$$

- Φ(x)を新たなパターンとみなし,今までのxをΦ(x)によって変換する. これにより,変換後の空間において線形SVM を適用したことに相当する.
- 変換後の空間における線形識別境界はxの元の特徴空間では非線形な識別境界をなす.

・計算コストを小さくするために,以下を満たす関数Kを学習と識別に導入する.

$$\Phi(x_1)^t \cdot \Phi(x_2) = K(x_1, x_2)$$

このようなKのことをカーネルと呼ぶ.

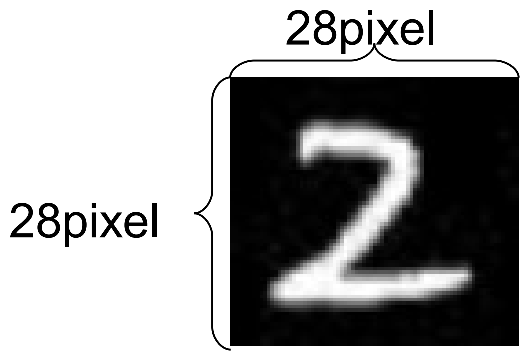
・代表的なカーネル関数に,ガウス型カーネルなどがある.

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

手書き数字認識への応用

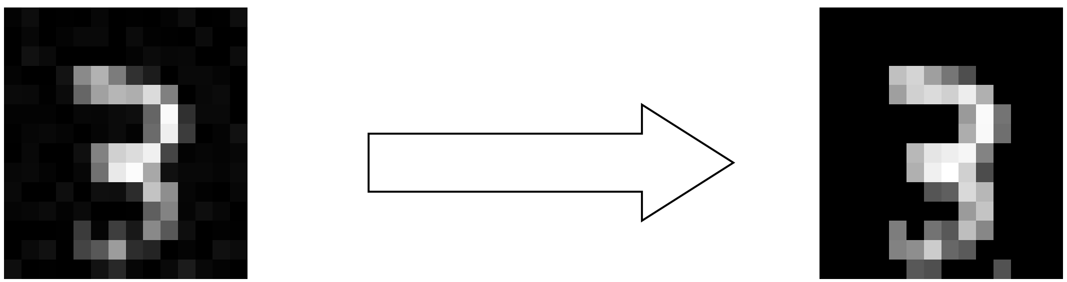
- SVMを利用して,手書き数字の認識について実験する.
- 数字は0~9の10文字で,これをSVMを多クラスの識別が可能なように拡張して識別する.

- 画像データは,MNISTのデータベースを用いる.



- 画像編集ソフトを用いて画像を縮小し,その画素の輝度を特徴次元として抽出し,SVMを適用する.

- 文字のかすれた部分を補正するために,各画素の輝度をxとすると, $255\sqrt{\frac{x}{255}}$ に変換する.



識別方法の比較

- SVMを多クラス識別できるように拡張するためには,1対1の識別を複数組み合わせるか,1対他の識別を行うかのどちらかがある. 今回は1対1の識別を複数組み合わせることにした.
- 1対1の複数の組み合わせによる識別方法として,多数決形式とトーナメント形式の2つを用いて,比較する.

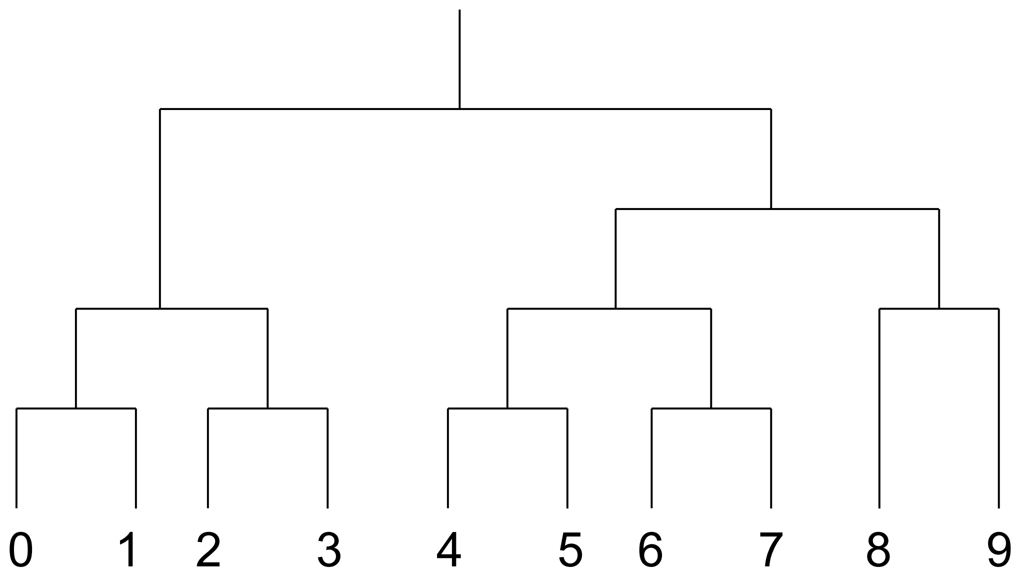
- 多数決形式

クラスの組み合わせの回数分(45回）識別を行い,最も識別された回数の多いものを識別結果とするもの.

最大識別回数が等しい場合は,識別関数の出力の合計が最も大きいものを識別結果とする.

・トーナメント形式

1対1の識別をトーナメントで9回行い,最後まで残ったクラスを識別結果とするもの.



- 多数決形式とトーナメント形式でそれぞれ識別を行った.

特徴次元　14*14(196）次元

各数字の学習データ数　300

各数字のテストデータ数　100

識別形式	識別率(%)	1文字あたりの認識時間（s）
多数決	94.5	3.3
トーナメント	94.6	0.9

学習データ数の比較

- 学習データ数によって識別率にどのような影響があるか調べるため,各数字の学習データ数をそれぞれ100,300,500,750とした時の識別率を求めた.

特徴次元　14*14(196）次元

各数字のテストデータ数　100

識別形式　トーナメント形式

学習データ数	識別率(%)	学習時間(m）
100	92.9	8
300	94.6	18
500	95.3	71
750	95.5	241

特徴次元数の比較

- 画像を7*7pixel,14*14pixelに縮小したものと,元の画像(28*28pixel)の３つから,それぞれの画素の輝度を特徴次元として抽出し,SVMを適用して比較した.

各数字の学習データ数　750

各数字のテストデータ数　100

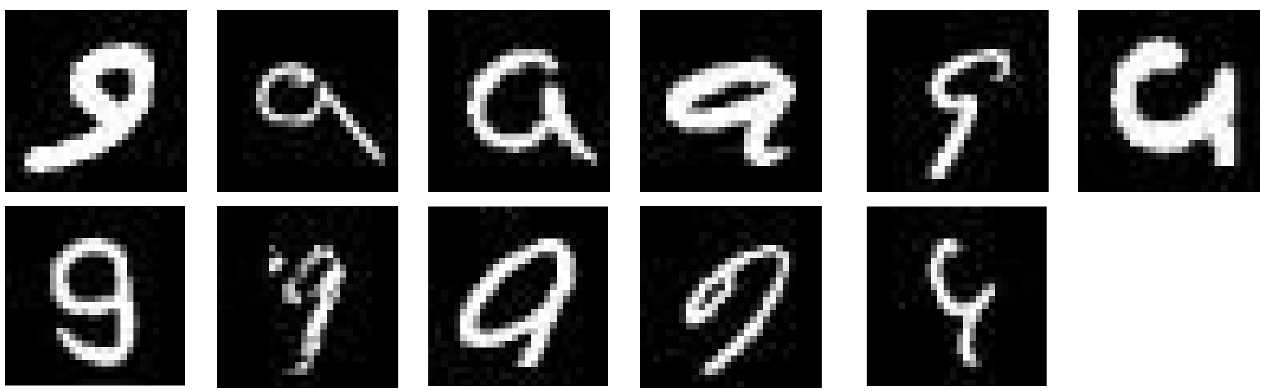
識別形式　トーナメント形式

特徴次元	識別率	学習時間（m)	1文字あたりの識別時間(s)
7*7(49）	90.1%	115	0.3
14*14(196）	95.5%	241	1.0
28*28(784)	95.5%	601	3.8

794次元での各数字の識別率

クラス	識別率(%)	クラス	識別率(%)
0	98	5	95
1	100	6	97
2	97	7	95
3	95	8	95
4	94	9	89

- 最も識別率の低かった9のクラスで,識別に失敗した画像は以下のものである.



まとめ

- 多数決形式とトーナメント形式では,識別率に差はなかったが,識別時間ではトーナメント形式が多数決形式の $\frac{1}{4}$ 程度となった.
- 学習データ数の増加が進むにつれ,識別率の上昇は緩やかになったのに対し,学習時間は飛躍的に増えていった.

- 次元の増加による識別率の向上は見られたが,196次元と784次元では,識別率に変化はなかった.
- 数字によって識別率の高いものと悪いものが存在した.
- 今後の課題としては,異なる特徴抽出法の導入や文字の傾きの補正による,識別率の向上

などが考えられる.

参考文献(1)

[1]Nello Cristianini,John Shawe-Taylor
“サポートベクターマシン入門” 2005
P129-P163

[2] 青葉雅人:Support Vector Machine ってなに？(online)
<http://www.neuro.sfc.keio.ac.jp/~masato/study/SVM/index.htm>

参考文献(2)

[3]篠原正一:サポートベクターマシンの花品種識別への応用(2007)