

サポートベクターマシンによる観測データのクラス分けに関する基礎的検討

武蔵工業大学 学生会員 嶽仁志

武蔵工業大学 正会員 丸山 收

1.はじめに

サポートベクターマシン(SVM)と呼ばれるパターン認識手法が注目されている。これまで、いくつもの高度なパターン認識力を持つ手法が提唱されてきたが、その中でもサポートベクターマシンは非常に高い識別能力を持っていることで知られている。サポートベクターマシンが他のパターン認識手法と比べて特に優れているところは、マージン最大化という考え方で未学習のデータに対して高い識別能力を発揮できるところにある。本研究では、数値シミュレーションによりサポートベクターマシンの基礎的な検討を行う。

2.サポートベクターマシン

2-1 サポートベクターマシンとは

サポートベクターマシンとは、1990年代に Vapnik¹⁾らによって提案されたもので、カーネルトリックと組み合わせることで非線形の識別関数にも対応できるようになった学習モデルのことである。パターン認識を行うには、認識対象からデータを用いて、それを線形しきい素子によって判定し2つのパターンに識別する。また、パターン認識の確率を少しでも高くするために、過去のデータとその識別結果との関係を事前に学習させなければならない。

2-2 線形しきい素子

線形しきい素子とはサポートベクターマシンの基礎となるニューロンを単純化したモデルである。

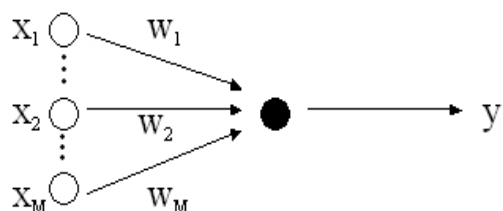


図1 線形しきい素子

図1のように特徴ベクトル $x = (x_1, \dots, x_M)$ に、それぞれ対応する結合荷重 $w = (w_1, \dots, w_M)$ を掛け、それ

らの和がしきい値(●)より大きい小さいかで判定し、認識対象を識別する。このとき M はデータの個数である。

2-3 識別関数と訓練サンプル

図1を式に示したものを識別関数といい、

$$y = \text{sign}(w^T \cdot x - h) = \text{sign}(u) \quad (1)$$

で表される。(1)式 $\text{sign}(u)$ は、 $u > 0$ のときに1を、 $u \leq 0$ のときに-1を出力する。このとき、-1,1の出力値を2つのクラス C_1, C_2 にそれぞれ振り分ける。

訓練サンプルとは過去のデータとその識別結果のことである。ここで、過去のデータを N 個の特徴ベクトル $x_1, \dots, x_N$ に、それに対する識別結果を

$t_1, \dots, t_N$ とする。また、識別結果 $t_1, \dots, t_N$ を C_1, C_2 の2つのクラスに振り分けておく。

2-4 ガウスカーネル

非線形の識別関数に対応するには、カーネルトリックを用いる必要がある。カーネルトリックとは対象物が高次元で写像しても、実際には高次元での計算は避けて、カーネルの計算のみで最適な識別関数を構成することであり、ガウスカーネルなどが使用されている。カーネル法においてガウスカーネルを用いると無限次元になり、全ての特徴ベクトルが線形独立となる。

3.データ解析

3-1 内容

本解析では、 X_1, X_2 の2入力、1出力 y のシステムを考える。ここで、 X_1, X_2 は0~100の1様な乱数として2500個用意し、 X_1 はX軸に X_2 はY軸に値する。まず、乱数2500個中1000個を訓練サンプルとして学習させ、残りの1500個で C_1, C_2 に識別するようにした。この解析結果を、正解の識別結果・サポートベクターマシンによる解析結果・識別境界を

キーワード サポートベクターマシン、パターン認識、数値シミュレーション

連絡先 〒158-8557 東京都世田谷区玉堤 1-28-1

変えた時のサポートベクターマシンによる解析結果の計3枚の図をそれぞれ case1～case4 の4パターン作り比較した。ここで言う case1 は $X_1 = X_2$ 式の境界線を引き $X_1 > X_2$ なら+1 を $X_1 < X_2$ なら-1 を出力し、case2 は中心(50,50)、半径 20 の円があり円の内部にデータがあるならば+1 を外部にデータがあるならば-1 を出力する。case3 は中心(0,0)、半径 20 の1/4 円と中心(30,60)、半径 20 の円と中心(70,20)、半径 20 の円があり case2 と同様に円の内部にデータがあるならば+1 を外部にデータがあるならば-1 を出力する。case4 は中心(50,50)、半径 20 の円があり円の内部にデータがあるならば+1 を外部にデータがあるならば-1 を出力する。しかし case4 では入力データ(X_1, X_2)に 0～10 の乱数 α を加える。このとき、乱数が増えられたことでクラスが変わったとしても所属するクラスは変わらないこととする。

3-2 結果

スペースの都合上、全ての結果を貼ることが出来ないため、ここでは case1, case3 の2パターンの結果を記しておく。

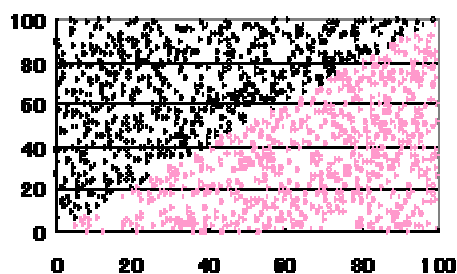


図2 正解の識別結果(case1)

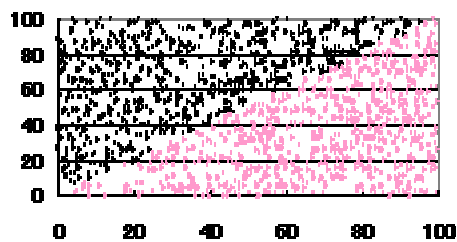


図3 SVMによる解析結果(case1)

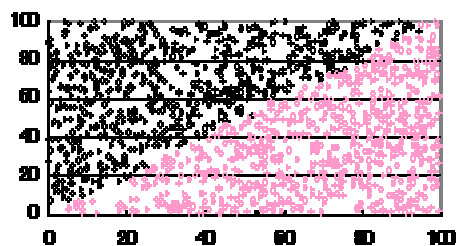


図4 識別境界を変えた時の SVM による解析結果(case1)

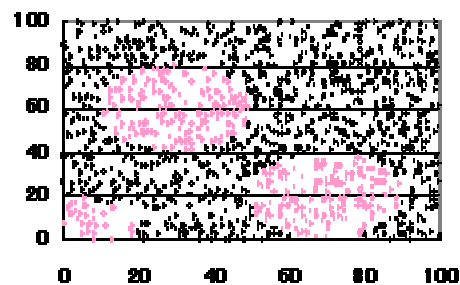


図5 正解の識別結果(case3)

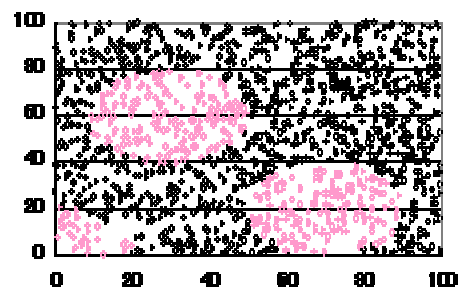


図6 SVMによる解析結果(case3)

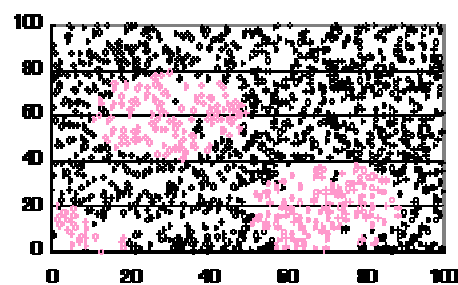


図7 識別境界を変えた時の SVM による解析結果(case3)

4.考察

正解の識別結果とサポートベクターマシンによる解析結果を見比べると多少の誤差はあるものの、かなり高い確率で一致していることが分かる。今回、識別境界を変えた時のサポートベクターマシンによる解析結果に間違いが多かったのは、誤差が出やすいようにパラメータを調整したためで、裏を返せばさらに誤差を減らすことは可能であると考えられる。また、このサポートベクターマシンを用いることで、例えば降雨量や斜面のデータを入力することで斜面崩壊の有無を求めることも可能だと考えられる。つまり、物事を2つのパターンに識別するにはサポートベクターマシンは非常に有効な手段であることが分かる。

5.参考文献

1. 栗田多喜夫：サポートベクターマシン入門
平成14年7月