

アイリスデータを用いたクラス判別手法の比較

中島 理恵 分部 理恵 (沼田 一道 助教授)

1. はじめに

世の中には、外見だけでは内在的な性質を判別することが困難であるものが多数存在している。例えば山でキノコを見つけたとき、素人が毒の有無を判別するのは難しい。ここで、既に毒の有無がわかっているキノコの外見の特徴から学習して推定する方法が考えられる。これが「クラス判別」と呼ばれる技術である。つまり、ある特徴を持った未知の認識対象に対して、帰属クラスが既知な多数のサンプルを学習しておき、与えられた未知対象の帰属クラスを推定する。

認識対象の特徴とクラスとの確率的な関係がわかっている場合には、ベイズ識別規則によって最適な判別が可能である。しかし実際問題、このような状況は稀であり、特徴とクラスの確率的な関係をサンプルデータから学習する必要がある。データと所属クラス

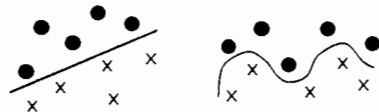


図1 線形分離境界

精度は低くなる。

本研究では、非線形な分離境界をもつサンプルデータについても判別可能な手法に注目する。

2. 研究目的

与えられたサンプルの特徴量(長さ, 重さ, 形など)を多次元空間内の点として配置し、各クラスに属する点集合を出来るだけ誤りなく分離するような境界を求めることで、クラス判別を行う。このとき、この境界を超平面だけに限定してうまくいく場合は稀である。そこで本研究では「ニューラルネット(NN)」「サポートベクターマシン(SVM)」の2つの非線形クラス判別手法に注目する。NNは線形判別素子を多数組み合わせることで合成関数をつくることにより、一方SVMは特徴量データをより高次の空間に写像し、その空間で分離超平面を求めることにより非線形判別を行う。SVMにおいて、最適な超平面を求める問題は2次計画問題に帰着される。このとき、高次空間への写像に伴う計算量の増加に対してカーネルトリックと呼ばれる手段を用いて対処している。

本研究では、アイリスデータを用いて2手法の推定精度を比較し、各々の判別手法の特徴を理解することを研究の目的とする。

3. 本研究で扱う判別分析の基本

学習サンプルとして、 n 個の特徴量データ $\mathbf{x}_i = (x_{i1}, \dots, x_{im})^T$ とその帰属クラス $d_i (+1 \text{ or } -1)$ が N 個与えられたとする ($i = 1 \dots N$)。ここではすべての学習データについて、 $d_i = +1$ のとき $\mathbf{w}^T \mathbf{x}_i - w_0 > 0$ 、 $d_i = -1$ のとき $\mathbf{w}^T \mathbf{x}_i - w_0 < 0$ となるような重み \mathbf{w}, w_0 を「学習」によって求める。

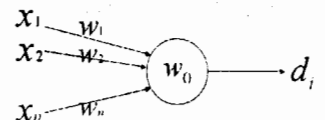


図3 線形判別素子ニューロン

これは2つのクラスを分離する超平面: $\mathbf{w}^T \mathbf{x}_i - w_0 = 0$ を求めることである。この d_i を教師データと呼ぶ。また図3のように表されるニューロンは、1個で上記の操作を行う線形判別素子である。さらに、どちらのクラスに判別されるのかを判定する関数として図4のステップ関数 S を導入すると、 \mathbf{w}, w_0 が求まった後、帰属クラスが未知のサンプル $\mathbf{u} = (u_1, \dots, u_m)^T$ の属するクラスは、 $S(\mathbf{w}^T \mathbf{u} - w_0)$

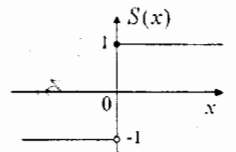


図4 ステップ関数

で推定される。問題は、多数のサンプルから重み \mathbf{w}, w_0 を求める「学習」方法であるが、NNでは教

師信号と出力 s との誤差がサンプル全体で最小となるような重み \mathbf{w}, w_0 を求め、一方 SVM では分離境界付近のサンプルに着目し、誤判別をなくすような定式化を行って最適な重みを求める。ただし NN では計算の都合上、ステップ関数の代わりにシグモイド関数 ($\sigma(t) = (1 + e^{-t})^{-1}$) を用いる。

3.1. ニューラルネット (NN)

NN では、分離境界が非線形であるような判別を可能にするため、多数のニューロンを例えば図 5 のように 3 層に組み合わせる。 n を特徴量数、 m を第 2 層数、 l を第 3 層(出力)数とすると、第 2 層での出力は重み w_{ij} ($i = 0 \dots n, j = 1 \dots m$) を用いて $f_j(x) = y_j = \sigma(w_{j0} + w_{j1}x_1 + \dots + w_{jn}x_n - w_{j0})$ とかける。また第 3 層での出力は、重み v_{jk} ($j = 0 \dots m, k = 1 \dots l$) を用いて、

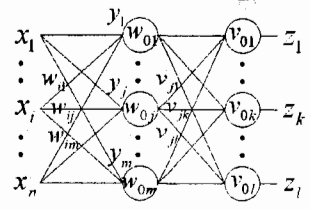


図 5 NN の構造

$g_k(y) = z_k = \sigma(v_{k1}y_1 + \dots + v_{km}y_m - v_{k0})$ とかける。従って NN の出力 $\mathbf{h} = (z_1, \dots, z_l)$ は $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ という合成関数で表される。

重み w_{ij}, v_{jk} は、全学習データに対する教師データ \mathbf{d} と出力 \mathbf{h} の差の総和(誤差) E ($E = \sum_{x \in S} \|h(x) - d(x)\|^2$) を最小にするように決める。具体的には適当な重み w_{ij}, v_{jk} から出発して、最急降下法により最小(極小)点へ向かわせる。勾配ベクトル $\nabla E(v) = \partial E / \partial v_{jk}$ は、「①出力 h と教師データ d の差」「② $\sigma(h)$ の微分」「③ $f_j(x)$ 」の 3 つの積の和となる。 $\nabla E(w) = \partial E / \partial w_{ij}$ も同様である。 α を非負の学習係数とすると、漸化式 $v_{jk}^{new} = v_{jk}^{old} - \alpha \cdot \partial E / \partial v_{jk}^{old}$, $w_{ij}^{new} = w_{ij}^{old} - \alpha \cdot \partial E / \partial w_{ij}^{old}$ により重みの更新をする。それによって新たな出力 \mathbf{h} を得て同様のことを行い、これを重みが変わらなくなるまで繰り返す。

3.2. サポートベクターマシン (SVM)

SVM はニューロンの重み(分離超平面を決定する \mathbf{w}, w_0) を 2 次計画問題に帰着させて決定する。図 6 のようなサンプル(特徴数: 2, クラス ●: +1, ×: -1) が与えられたとき、●と×を分離できる超平面は無数に存在するが、両クラスを出来るだけ誤りなく判別できるように分離超平面の向き (\mathbf{w} で決まる) を決定する。すなわち、法線方向 \mathbf{w} の超平面をその方向に平行移動するとき、正しく分離

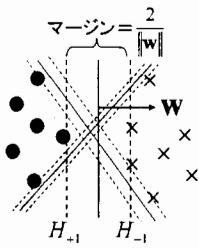


図 6 線形 SVM

できる限界位置 (●に接する超平面を H_{+1} , ×に接する超平面を H_{-1} とする) を考え、 H_{+1} と H_{-1} の間の距離(マージン)が最大になるように \mathbf{w} を選ぶ。このとき、 H_{+1} (H_{-1}) を $\mathbf{w}^T \mathbf{x} - w_0 = 1$ ($\mathbf{w}^T \mathbf{x} - w_0 = -1$) とすることが出来るので、クラス ● のサンプルについては $\mathbf{w}^T \mathbf{x} - w_0 \geq 1$, クラス × のサンプルについては $\mathbf{w}^T \mathbf{x} - w_0 \leq -1$ となるようにする。また、 H_{+1} と H_{-1} の距離は $2/\|\mathbf{w}\|$ で与えられるので、 \mathbf{w}, w_0 を決定する問題は、 $d_i \cdot (\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1$ の制約下で、 $\|\mathbf{w}\|^2 / 2$ を最小化する 2 次計画問題となる。分離超平面 $\mathbf{w}^T \mathbf{x} - w_0 = 0$ は H_{+1} と H_{-1} のちょうど中間に位置する。

どのような超平面を選んでも、誤ったクラスに入るサンプルが存在する場合、各サンプルに非負の補正量 ξ_i ($i = 1 \dots n$) を導入し、 $d_i \cdot (\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1$ の制約を緩めて同様に考える。ただし、この場合は、マージンの大きさの逆数と補正量の総和はトレードオフの

関係にあるので、重み C を用いて両者を統合したものを最小化する。この場合の 2 次計画問題は(1) のようになる。

さらに境界が線形でない場合には、特徴量データ \mathbf{x} を高次の空間に写像し ($\mathbf{x} \rightarrow \boldsymbol{\varphi}(\mathbf{x})$)、写像後の空間で分離超平面を求める。この操作により、元のデータについての非線形判別ができることを期待する。この写像を行うと高次元での内積計算 ($\boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j)$) が必要となるが、これは $\boldsymbol{\varphi}(\mathbf{x})$ をうまく選

ぶことにより元空間の次元の計算で済ませることができる。 $\varphi(x_i)^T \varphi(x_j) = K(x_i, x_j)$ を満たす K をカーネル関数と呼ぶ。カーネル関数としては、ガウシアンカーネル $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\delta^2)$ などが知られており、本研究でもこれを用いる。パラメータ C , δ については様々な値を代入して試行する。

4. 実験

本研究ではサンプルとして、アイリスデータを用いる。これは、3種類のアヤメの花（セトサ 50本、パーシクル 50本、バージニカ 50本）と、各々の花についての4つの特徴量（花びら長さ、花びら幅、がく片長さ、がく片幅）から成るデータである。

NNは「学習」および「推定」のプログラムを作成し、一方SVMについては解法ツール「LINGO8」への変換プログラムを作成した後、このツールを用いて2次計画問題を解く。また、線形SVMを用いた場合（表にはSVMと記載）と、カーネル関数を用いた場合（表にはKnlと記載）に分けて以下の実験を行う。

各々の判別手法を用いてアイリスデータについての「学習」を行う。このとき、入出力の組み合わせをさまざまに変えて行い、その結果、線形判別が可能であるかどうか、非線形判別が必要な場合に正しい判別が可能であるかどうかを調べる。

次に、「学習」の結果を用いて、種類が未知のアヤメについての「推定」を行う。総サンプル150個のうち、140個から「学習」し、残り10個の花の種類を各々の判別手法を用いて推定する。これを15回繰り返して行い、その結果から各々の判別手法の推定精度を比較し検討する。

ただし、NNは特徴量（入力）を0以上1以下の値に正規化する。また花の種類（クラス）を決定する条件として、各々の花について得られた出力 $\mathbf{h} = (z_1, z_2, z_3)$ （0以上1以下の値）の中で、最も1に近い値をとる花の種類と判別することにする。

また、SVMは2クラスの判別しか行えないため、3クラスを判別する際には、「セトサ or その他」「パーシクル or その他」「バージニカ or その他」と場合分けをして判別を行う必要がある。

5. 実験結果

1. 「学習」の結果

i) アヤメの種類（出力）が2種類の場合

入力（特徴量）を2次元の2種類（「花びら長さ・花びら幅」「がく片長さ・がく片幅」と4次元の場合で行う。またSVMにおいて、 $\delta=20$ とし、重み C はその都度適当な値を用いる。

- ・「セトサとパーシクル」「セトサとバージニカ」では線形SVMにおいて誤りなく学習できたことにより、線形判別が可能であるといえる。NNにおいても最適な重み w_{ij}, v_{jk} を求めることができた。
- ・「パーシクルとバージニカ」ではNN, SVM共に、学習終了時点でほぼ同じ数の誤りが存在した。またSVMにおいて、線形SVMとカーネル関数を用いた場合を比べると誤りの数に差がなかった。これよりクラスの異なるサンプルが複雑に入り混じったデータではないことがわかる。

ii) アヤメの種類（出力）が3種類の場合

i)と同様、入力を2次元の2種類（「花びら長さ・花びら幅」「がく片長さ・がく片幅」と4次元の場合で行う。ただしSVMは、前章で述べたように場合分けをして判別を行い、その結果をまとめてNNとの比較を行う。また表1は入力が2次元の「花びら長さ・花びら幅」、表2は4次元の場合の学習結果である。

- ・「セトサ」のサンプルは入力の次元にかかわらず、NN, SVM共に、正しく学

表1 2入力3出力の場合

	正解数 (50)		
	NN	SVM	Knl
セトサ	50	50	50
パーシクル	46	0	48
バージニカ	47	46	46

習が行えた。

- 「パーシクル」のサンプルは学習精度が悪い結果となった。特に、表1において線形SVMを用いた場合はひとつも正解を得られなかった。しかしカーネル関数を用いると正解数が上がったことから、SVMにおいて分離境界が非線形となるとき学習がうまくできたといえる。またNNにおいては正解数がカーネル関数を用いた場合とほぼ同じになったので、NNでの非線形判別ができているとわかる。また、表2においても同様のことがいえる。

表2 4入力3出力の場合

花びら長さ・花びら幅 びくた長さ・びくた幅	正解数 (50)		
	NN	SVM	Knl
セトサ	50	50	50
パーシクル	47	26	49
バージニカ	49	48	48

II. 推定の結果

SVMでは場合分けをし、その結果をまとめる。このときカーネル関数を用い、 $\delta=20$ とする。

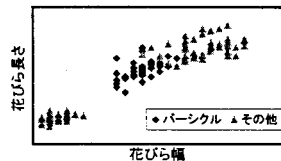
NNでは推定数150個のうち誤判別は5個であった。これは「パーシクル」と「バージニカ」を互いに誤判別していることがわかった。

一方、カーネル関数を用いた推定では重み $C=10$ の場合は誤判別が4個あった。「パーシクル」と「バージニカ」が誤っていることはわかるが、それがどの花の種類に誤判別されているのかを知ることが不可能であった。また、 C の値を0.01~0.05に変えることによって誤判別がゼロとなった。

6. 考察

実験の結果より、以下に考察を述べる。

- 表1の線形SVMの判別精度が悪い原因は散布図(図7)より明らかである。この条件で分離超平面を得ることは不可能であるとわかる。
- SVMにおける「推定」実験の結果から、重み C の値をどのように決め、図7パーシクルとその他の分布るかによって判別への影響が大きいことがわかる。 C の値を小さくすることは多くの補正量を許す方向に働くので、マージンの距離を大きくとることに相当し、判別性能が上がったと考えられる。
- 「学習」実験ではNNとSVMに大きな差は見られなかったが、「推定」実験ではSVMの方が良い結果を得られた。この原因は「推定」を行う際に、NNでは学習結果をそのまま用いなければ成らないのに対し、SVMでは重み C を変えて新たな分離超平面を求めることができるため、クラスが未知のサンプルに対して最適な判別結果を返すことが可能となるからであると考えられる。
- NNの良さは、クラスが2つ以上の場合でも判別が可能などである。このときSVMでは場合分けが必要となり、さらにそれぞれに対応した2次計画問題を解かなくてはならない。



7. まとめ

本研究ではアイリスデータを用いて、2つのクラス判別手法(NNとSVM)の判別性能を比較評価した。NNにおいては、学習の結果より非線形判別ができることが確認できた。しかし推定に関しては限界を感じる結果となった。SVMは、学習の結果にかかわらず良い推定精度が得られることがわかった。しかしカーネル関数には多くの種類がある。そのため本研究で用いた関数以外を用いて、さらに良い判別を行える可能性もあるが、適用する関数の選び方は決まっていなため、試行によって決定しなければならない課題も残っている。

参考文献

- [1] 上坂吉則：ニューロコンピューティングの数学的基礎；近代科学社(1993)
- [2] 富田秀樹：非線形多変量解析；朝倉書店(1996)
- [3] 前田英作：痛快!サポートベクトルマシン；情報処理,42巻7号,pp.676-683(2001)
- [4] Hiroshi Yadohisa：Data Representation System <http://www.sci.kagoshima-u.ac.jp/~drs/ja/data/>