ロジスティック回帰モデルによる確率判別

茨木志織　(指導教員：吉田裕亮)

1　はじめに

一般に, 判別分析とは, 2 変量の場合では散布図上に異なる群 (母集団) からの観測値を同時に打点し, 散布図上で異なる群を分ける領域を求める分析手法のひとつである.

本研究では, その判別分析において群を 0.5 対 0.5 に分けるのではなく, 発生確率に基づいて群を分ける手法として, ロジスティック回帰モデルを用いた手法を提案する.

一般にロジスティック回帰モデルの係数推定は非線形な最適化問題であり, 計算が複雑である. そこで, 本研究ではロジスティック回帰モデルと構造的に類似性の見られるニューラルネット・モデルを利用する. さらにロジスティック回帰モデルを用いて推定される確率的判別直線が, マハラノビス距離による 2 次判別曲線よりもより妥当な確率判別を与えることもみる.
まず, ロジスティック回帰についてみてみよう.

2　ロジスティック回帰モデル

n 次元の変量 $X_1, X_2, \cdots, X_n$ の観測値 $x_1, x_2, \cdots, x_n$ が与えられたとき, ある現象の発生する確率を

$$\mathrm{Prob}\{\text{生起}\,|x_1, x_2, \cdots, x_n\} = \frac{1}{1+\exp(-Z)}$$

で表すモデルをロジスティック回帰モデルという. ただし,

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

で β_i は各変量 X_i の重みである. このモデルは医療統計の分野でよく用いられるモデルのひとつである. また, 次のニューラルネット・モデルと構造的類似が見られる.

3　ニューラルネット・モデル

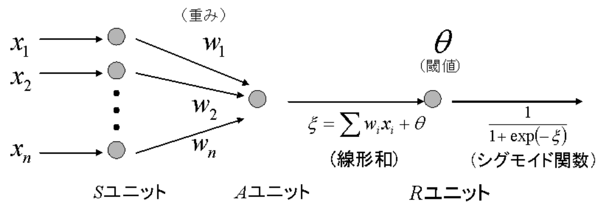

図 1: 単純パーセプトロンモデル

図 1 は S ユニット (Sensory unit), A ユニット (Associative unit), R ユニット (Response unit) の 3 層からなるパーセプトロンと呼ばれる, パターンを分類するために提案されたネットワークモデルのひとつである. 分類対象のパターンを入力すると S ユニットが反応し, 次に A ユニットは S ユニットからの入力を受けて信号を出力する. そして R ユニットは A ユニットからの入力により応答し, 入力パターンを分類する信号を出力する. S ユニットから A ユニットへの結合の強さ (重み w) は次のように学習されていく.

1. 結合係数 w_i と R ユニットのオフセット θ を初期化し, パターン x_i をセットする.
2. A ユニットで x_i と重み w_i の線形和
$$\xi = w_1x_1 + w_2x_2 + \cdots + w_nx_n + \theta$$
を構成し, R ユニットのシグモイド関数に送り,
$$\eta = \frac{1}{1+\exp(-\xi)}$$
を出力値とする.
3. この出力 η を教師信号 T を使って結合係数を
$$w_i(t+1) = w_i(t) + C(T-\eta)x_i$$
と更新していく. 右辺の $w_i(t)$ が時刻 t の結合係数を, 左辺の $w_i(t+1)$ が時刻の $t+1$ の結合係数を表している. ここで C は学習速度を制御するパラメータである.
この学習をある基準に収束するまで繰り返すことにより, 重み w_i を推定する.
この学習は入力データが与えられるたびに重みを更新するので, オンライン学習とも呼ばれている.

4　マハラノビス距離

多変量データの相関に基づき, 算出される距離としてマハラノビス距離がある. 2 次元データで各クラス $(j = 0, 1)$ の平均が $\vec{\mu_j} = {}^t(\mu_{0j}, \mu_{1j})$ で, 分散共分散行列が Σ_j とする. このとき $\vec{x} = (x_0, x_1)$ の各 j 群のマハラノビス距離 D_j は,

$$D_j = \sqrt{{}^t(\vec{x}-\vec{\mu_j})\,\Sigma_j^{-1}\,(\vec{x}-\vec{\mu_j})}$$

で与えられる.

5　提案手法 (例)

本研究ではロジスティック回帰モデルを用いて確率判別を以下のように行うことを提案する.

1. ロジスティック回帰モデルに用いられる係数 β_i をニューラルネット・モデルの重み w_i として求める. ただしニューラルネット・モデルでは教師信号を群 1 のとき 1, 群 0 のとき 0 とする. 求められた係数をロジスティック回帰モデルに適用する.
2. 求められたロジスティック回帰モデルに発生確率を与えることにより, 判別直線を求める.

6　数値実験

本研究で提案した判別直線とマハラノビス距離の比で求められる判別曲線について, 精度を比較する.
以下のような 2 変量の群を用意する.
群 1：$N(\mu_1, \Sigma_1)$ に従う 600 個のデータ

$$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1.03 & 0.36 \\ 0.36 & 1.09 \end{pmatrix}, \rho = 0.34$$

群 0：$N(\mu_0, \Sigma_0)$ に従う 600 個のデータ

$$\mu_0 = \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix}, \Sigma_0 = \begin{pmatrix} 1.03 & -0.36 \\ -0.36 & 1.09 \end{pmatrix}, \rho = -0.34$$

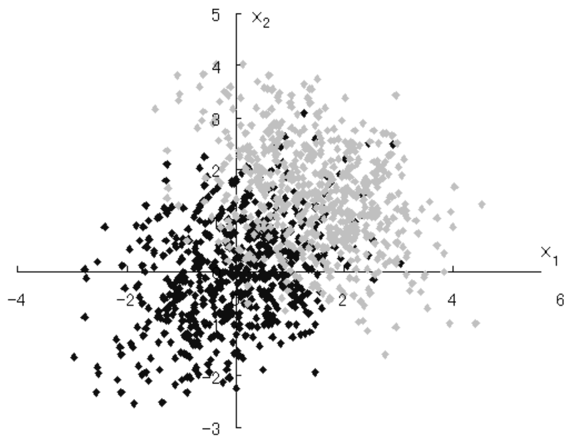

図 2:

図 2 のように 0 群と 1 群 が混在した領域が存在しているものを用意する. 上においてρ は相関係数である. 群 1 に含まれる確率が 0.8 となるような判別直線を先のように求めていく.

6.1　提案手法の場合

6.1.1　重みの推定

群 1 と群 0 を分ける直線は

$$w_1x_1 + w_2x_2 + \theta = 0$$

と表せる. θ は R ユニットのオフセットである. 今回の実験では,

$$w_1 = -13.852, w_2 = -13.142, \theta = 20.967$$

を得る. これより直線は

$$x_2 = -1.05x_1 + 1.60$$

となり, 2 つの群を分ける判別直線が推定される.

次にニューラルネット・モデルで得られた結合係数をロジスティック回帰モデルに適用する. 群 1 に含まれる確率が 0.8 のロジスティック回帰モデルは,

$$0.8 = \frac{1}{1 + \exp(-w_1x_1 - w_2x_2 - \theta)}$$

と表される. この式を満たす x_1 と x_2 の関係を描く.

6.1.2　結果

右の図 3 のような群 1 と群 0 を分ける直線が引ける. また, この直線よりも下側にある群 1 のデータ数を数えたところ, 490 個 (/600 = 0.816) があることがわかる (図 4). よって, この直線より下側に群 1 のおよそ 0.8 が存在する.

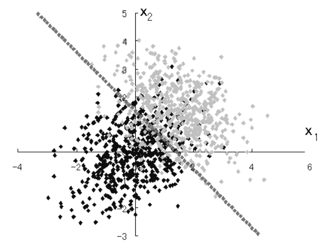

図 3: 群 1, 群 0 と確率判別直線

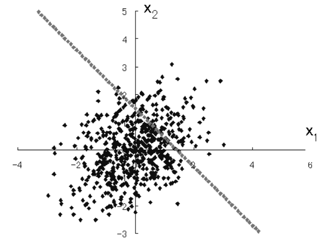

図 4: 群 1 と確率判別直線

6.2　マハラノビス距離の比の場合

d_1：群 1 の平均までのマハラノビス距離
d_0：群 0 の平均までのマハラノビス距離
とし,　　d_1：$d_0 = 8:2$　　　… (※)
となるような x_1 と x_2 の関係を求めることで, 確率判別曲線を描くと以下ようになる.

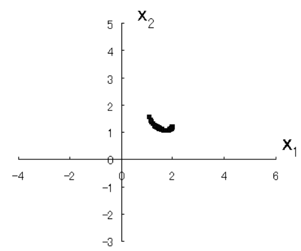

図 5: マハラノビス距離を用いた確率判別曲線

図 5 のような曲線が求められた. (※) を満たすような x_1 と x_2 の範囲は非常に狭いため, 判別曲線を短い区間でしか引くことが出来ず, 分かりにくい.

7　まとめ

今回, 2 次元の場合, ロジスティック回帰モデルを用いることによって確率判別直線を引くことができた. マハラノビス距離の比を用いても確率判別線の 2 次曲線が描けるが, ロジスティック回帰モデルの方が正確で視覚的にもわかり易く, 確率の比で群を分ける手法に適していると思われる. 今後の課題として, バイキング法により確率判別直線を重ねることで, 非線形な確率判別に拡張することも考えられる.

参考文献

1. 竹村彰通, 谷口正信, 統計学の基礎　, 岩波書店 (2003)
2. 中野馨, 入門と実習 ニューロコンピュータ, 技術評論社 (1989)