高次元小標本における正準相関に基づくパターン認識

玉谷　充[1]，　Inge Koch[2]，　内藤　貫太[3]

[1] 島根大学大学院総合理工学研究科　[2] アデレード大学　[3] 島根大学総合理工学部

1　はじめに

判別分析の研究の起源は Fisher の群内と群間の変動の比を最大にする手法から始まり，近年では，高次元小標本における判別手法の構築およびその理論評価が盛んに研究が行われている．先行研究として Bickel and Levina [1] は，高次元小標本において用いることができる判別手法である Naive Bayes の性能について考察を与えた．さらに，Fan and Fan [2] は高次元小標本における判別の手法として FAIR を提案した．

本講演では，正準相関に基づく新たな判別関数を構築し，その関数の理論的評価およびシミュレーションによる Naive Bayes との比較検証を行った．

2　設定

次元を d とし，データ数を n とし，2 クラスの判別問題について考える．このときに得られるクラス $k(=0,1)$ の n_k 個の観測値ベクトルを $\boldsymbol{X}_{ki} = (X_{ki1}, \ldots X_{kid}) \quad (i = 1, \ldots, n_k)$ とし，$\boldsymbol{X}_{ki} = \boldsymbol{\mu}_k + \boldsymbol{\varepsilon}_{ki} \quad (k = 0, 1, i = 1, \ldots, n_k)$ を仮定する．ただし，$\boldsymbol{\mu}_k \in \mathbb{R}^d$ は平均ベクトルで $\boldsymbol{\varepsilon}_{ki} \sim N_d(\boldsymbol{0}, \Sigma)$ は誤差ベクトルである．

ここでは，高次元小標本の設定を与えていることから $n < d$ または $n = o(d)$ とする．

3　パターン認識手法

3.1　Fisher の線形判別関数と Naive Bayes

等分散性を仮定している場合，Fisher の線形判別関数

$$\delta_{\mathrm{F}}(\boldsymbol{x}) = (\boldsymbol{x} - (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0)/2)^T \, \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

がベイズルールとなり最適である事が知られている．しかしながら，$n < d$ の状況では分散共分散行列 Σ に対する不偏推定量 $\widehat{\Sigma}$ を求めたとしても，逆行列が存在しないので，標本に基づく判別関数 $\widehat{\delta}_{\mathrm{F}}$ を構築する事ができない．そこで，δ_F における Σ の代わりに $D = \mathrm{diag}\Sigma$ を用いた Naive Bayes によって判別関数を構築する．この場合，分散共分散行列 Σ に対する不偏推定量 $\widehat{\Sigma}$ を構築したとしても，$\widehat{D}$ は対角成分が非零である限り逆行列は存在する．

3.2　正準相関に基づく判別分析

我々の提案は，正準相関に基づいて判別関数を構築することである．Koch and Naito [3] では多変量回帰の枠組みで連続応答変数の予測手法を提案しているが，ここではその手法を判別の枠組みにおいて用いることを考える．すなわち，通常の正準相関分析では，$\boldsymbol{X}$，$\boldsymbol{Y}$ を共に中心化された p，q 変量確率変数ベクトルを用いるが，ここでは $\boldsymbol{Y}$ を中心化せずに以下の量を与える：

$$C = \Sigma^{-1/2} E[(\boldsymbol{X} - \mu_X)\boldsymbol{Y}^T] E[\boldsymbol{Y}\boldsymbol{Y}^T]^{-1/2}.$$

ここで，$\boldsymbol{Y}$ は $q=2$ とし，2 値を取る離散型確率変数ベクトルとする：

$$\boldsymbol{Y}=\begin{bmatrix} Y_1 \\ Y_0 \end{bmatrix},\quad Y_1=\begin{cases} 1 & \text{Probability}=\pi, \\ 0 & \text{Probability}=1-\pi. \end{cases},\quad Y_0=1-Y_1.$$

このとき，固有値問題 $CC^T\boldsymbol{p}=\nu^2\boldsymbol{p}$ を考えるとフィッシャーの評価基準に帰着し，中心化されたデータに対して $\boldsymbol{b}=\Sigma^{-1/2}\boldsymbol{p}$ によって射影を施せば良い事が分かる．ゆえに，$\boldsymbol{b}=\Sigma^{-1/2}\boldsymbol{p}$ の推定量を構築したいのだが，$n<d$ だと Fisher の線形判別関数と同様 $\widehat{\Sigma}$ の逆行列が得られない．そこで，X, Y をそれぞれ $d\times n$ のデータ行列とし，行列 C の推定量を以下によって与えることにする：

$$\widehat{C}=\widehat{D}^{-1/2}\left(\frac{1}{n}XY^T\right)\left(\frac{1}{n}YY^T\right)^{-1/2}.$$

4 誤判別確率

判別関数の精度の良し悪しを測るために，判別関数 $\widehat{\delta}$ と適当なパラメータ空間 Γ の元 $\theta=(\boldsymbol{\mu}_1,\boldsymbol{\mu}_0,\Sigma)$ に対して，観測ベクトル $\boldsymbol{X}$ がクラス 1 からのものであるときに，$\widehat{\delta}$ がクラス 0 と判別する確率を以下で定義する：

$$W(\widehat{\delta},\theta)=P(\widehat{\delta}(\boldsymbol{X})\le 0\mid \boldsymbol{X}_{ki}, k=1,0, i=1,\ldots,n_k).$$

また，判別関数 $\widehat{\delta}$ の最悪誤判別確率を $W(\widehat{\delta})=\max_{\theta\in\Gamma}W(\widehat{\delta},\theta)$ で定義する．パラメータ空間 Γ は

$$\Gamma=\left\{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_0,\Sigma)\ \middle|\ (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T D^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)\ge C_d, \lambda_{\max}(R)\le b_0, \min_{1\le j\le d}\sigma_{jj}^2>0\right\}$$

とする．ただし，$\Sigma=(\sigma_{ij})$ で $R=D^{-1/2}\Sigma D^{-1/2}$ は相関行列である．このとき，誤差ベクトルの成分が適当な正則条件を満足するのであれば，パラメータ空間 Γ における Naive Bayes $\widehat{\delta}_{\mathrm{NB}}$ とここで提案する $\widehat{\delta}_{\mathrm{NEW}}$ の最悪誤判別確率は一致することが示される．

5 シミュレーション

シミュレーションでは $\widehat{\delta}_{\mathrm{NB}}$ と $\widehat{\delta}_{\mathrm{NEW}}$ の精度比較を行った．設定は，各クラスの観測値ベクトル $\boldsymbol{X}_k$ $(k=0,1)$ は d 次元正規分布に従っていると仮定し，$\widehat{\delta}_{\mathrm{NB}}$ と $\widehat{\delta}_{\mathrm{NEW}}$ の CV 誤差をデータ数，次元，共分散の数値を変えながら比較を行った．結果として，共分散がある程度大きいときには $\widehat{\delta}_{\mathrm{NB}}$ よりも $\widehat{\delta}_{\mathrm{NEW}}$ の方が CV 誤差が小さく，一方で共分散が小さいときでさえも，データ数が小さくかつ次元が大きければ $\widehat{\delta}_{\mathrm{NEW}}$ の方が CV 誤差が小さいという結果が得られた．

参考文献

[1] Bickel, P. J. and Levina, E. (2004). Some theory for Fisher's linear discriminant function, “ naive Bayes ”, and some alternatives when there are many more variables than observations. *Bernoulli* **10**, 989-1010.

[2] Fan, J. and Fan, Y. (2008). High-dimensional classification using features annealed independence rules. *Ann. Statist.* **36**, 2605-2637.

[3] Koch, I. and Naito, K. (2010). Prediction of multivariate responses with a selected number of principal components. *Computational Statistics* & *Data Analysis.* **54**, 1791-1807.