

関数データに基づく半教師あり識別・判別問題

九州大学大学院数理学府 川野 秀一

九州大学大学院数理学研究院 小西 貞則

1 はじめに

識別・判別問題において、どの群に属するか既に分かっているデータ（ラベルありデータ）と群の所属が未知のデータ（ラベルなしデータ）が混在している状況の下、ラベルありデータのみに基づいて判別モデルを構成するのではなく、ラベルなしデータを効果的に用いることにより判別モデルを構成し、予測の精度を向上させる手法は、半教師あり学習法と呼ばれている (Chapelle *et al.*, 2006). 近年この学習法は大きな注目を集めており、学習データの取得が高コストである分野、特にバイオインフォマティクスやテキスト分類などの分野で応用されつつある (例えば, Nigam *et al.*, 2000; Chapelle and Zien, 2005; Dean *et al.*, 2006; Pan *et al.*, 2006 参照).

一方、離散時点で経時的に測定・観測されたデータに関数化処理を施し、得られる関数データ集合を解析対象とする手法は関数データ解析と呼ばれ、諸科学のさまざまな分野でその有用性・重要性が報告されている (Ramsay and Silverman, 2002; 2005). しかし、群の所属が既知の関数データ（ラベルあり関数データ）と未知の関数データ（ラベルなし関数データ）の両方の情報を用いた判別モデルに関する研究、つまり、関数データ集合に対する半教師あり学習法の研究についてはいまだ十分に確立していない.

本報告では、各個体の観測データに対して関数化処理を施し、関数データ集合を対象とした半教師ありロジスティック識別・判別法を提案する. モデルに含まれるパラメータの推定には、モデルの安定化を図るために正則化法を適用し、推定アルゴリズムに EM アルゴリズムを用いる. また、モデルに含まれるチューニングパラメータの値を客観的に選択するために、ベイズアプローチに基づくモデル評価基準を導出する. さらに、実データへの適用を通して提案手法の有効性を検証する.

2 関数データ

いま、 n 個の個体に対して特性 $X(\in \mathbf{R})$ について変数 $T(\in \mathbf{R})$ をともない観測された離散的なデータ集合

$$\{(t_{\alpha j}, x_{\alpha j}); j = 1, \dots, N_{\alpha}, t_{\alpha j} \in \mathcal{T}(\subset \mathbf{R})\}, \quad (\alpha = 1, \dots, n), \quad (1)$$

が与えられたとする. ただし、 $x_{\alpha j}$ は α 番目の個体の時点 $t_{\alpha j}$ での観測データとする. この離散データに対して非線形回帰モデルを当てはめることにより、関数データ集合 $\{x_{\alpha}(t); t \in \mathcal{T}\}$ ($\alpha = 1, \dots, n$) を求める.

観測データ集合 $\{(t_{\alpha j}, x_{\alpha j}); j = 1, \dots, N_{\alpha}\}$ は、次のような回帰モデルに従うと仮定する.

$$x_{\alpha j} = u_{\alpha}(t_{\alpha j}) + \varepsilon_{\alpha j}, \quad j = 1, \dots, N_{\alpha}. \quad (2)$$

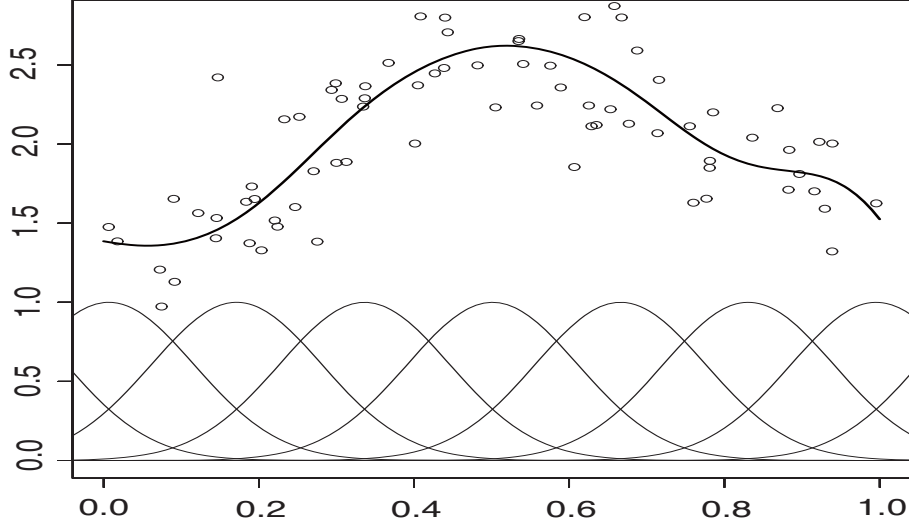


図 1: 基底展開法による離散データの関数化の例。丸印は離散データ，図下部の曲線は基底関数，図上部の曲線は関数データを表す。

ここで， $u_\alpha(\cdot)$ は未知の平均構造， $\varepsilon_{\alpha j}$ は平均 0，分散 $\sigma_{x_\alpha}^2$ の正規分布に従うものとする。未知の平均構造 $u_\alpha(\cdot)$ は，基底関数と呼ばれる m ($< N_\alpha$) 個の既知の関数 $\phi_i(t)$ ($i = 1, \dots, m$) の線形結合により，

$$x_{\alpha j} = \sum_{i=1}^m c_{\alpha i} \phi_i(t_{\alpha j}) + \varepsilon_{\alpha j} = \mathbf{c}_\alpha^T \boldsymbol{\phi}(t_{\alpha j}) + \varepsilon_{\alpha j} \quad (3)$$

と表されると仮定する。ただし， $\mathbf{c}_\alpha = (c_{\alpha 1}, \dots, c_{\alpha m})^T$ は m 次元係数ベクトル， $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_m(t))^T$ は m 次元基底関数ベクトルとする。

基底関数としては，これまでデータの特性・分析目的に応じて，フーリエ級数， B -スプライン (Imoto and Konishi, 2003)，ウェーブレット (Fujii and Konishi, 2006) などさまざまなものが用いられてきた。本報告では，基底関数としてガウス型基底関数を用い，特に次で与えられるガウス型基底関数を用いることとする (Kawano and Konishi, 2007)。

$$\phi_i(t) = \exp \left\{ -\frac{(t - \tau_{i+2})^2}{2h^2} \right\}, \quad h = \frac{\tau_{i+2} - \tau_i}{3}. \quad (4)$$

ここで， τ_i は昇順の数列で， τ_4, τ_{m+1} がそれぞれ全観測時点の始点，終点となるように等間隔に配置された節点とする。したがって，(2) 式と (4) 式よりガウス型基底関数に基づく非線形回帰モデルは次式で表される。

$$f(x_{\alpha j} | t_{\alpha j}; \mathbf{c}_\alpha, \sigma_{x_\alpha}^2) = \frac{1}{\sqrt{2\pi\sigma_{x_\alpha}^2}} \exp \left[-\frac{\{x_{\alpha j} - \mathbf{c}_\alpha^T \boldsymbol{\phi}(t_{\alpha j})\}^2}{2\sigma_{x_\alpha}^2} \right], \quad j = 1, \dots, N_\alpha. \quad (5)$$

係数パラメータ \mathbf{c}_α と誤差分散 $\sigma_{x_\alpha}^2$ は，正則化法に基づいて推定する。すなわち，モデルの対数尤度関数に正則化項を付与した正則化対数尤度関数

$$\ell_{\zeta_\alpha}(\mathbf{c}_\alpha, \sigma_{x_\alpha}^2) = \sum_{j=1}^{N_\alpha} \log f(x_{\alpha j} | t_{\alpha j}; \mathbf{c}_\alpha, \sigma_{x_\alpha}^2) - \frac{N_\alpha \zeta_\alpha}{2} \mathbf{c}_\alpha^T \boldsymbol{\Omega} \mathbf{c}_\alpha$$

$$= -\frac{N_\alpha}{2} \log(2\pi\sigma_{x_\alpha}^2) - \frac{1}{2\sigma_{x_\alpha}^2} (\mathbf{x}_\alpha - \Phi_\alpha \mathbf{c}_\alpha)^T (\mathbf{x}_\alpha - \Phi_\alpha \mathbf{c}_\alpha) - \frac{N_\alpha \zeta_\alpha}{2} \mathbf{c}_\alpha^T \boldsymbol{\Omega} \mathbf{c}_\alpha \quad (6)$$

の最大化により推定する。ここで、 $\Phi_\alpha = (\phi(t_{\alpha 1}), \dots, \phi(t_{\alpha N_\alpha}))^T$ 、 $\mathbf{x}_\alpha = (x_{\alpha 1}, \dots, x_{\alpha N_\alpha})^T$ であり、 $\zeta_\alpha (> 0)$ は罰則の度合いを調整する正則化パラメータ、 $\boldsymbol{\Omega}$ は $m \times m$ の既知の非負値定符号行列とする。特に、本報告では D_2 を 2 次の差分行列としたとき、非負値定符号行列として $\boldsymbol{\Omega} = D_2^T D_2$ を用いる。(6) 式の最大化によって得られる正則化推定量 $\hat{\mathbf{c}}_\alpha$ 、 $\hat{\sigma}_{x_\alpha}^2$ は、それぞれ次式で与えられる。

$$\hat{\mathbf{c}}_\alpha = (\Phi_\alpha^T \Phi_\alpha + N_\alpha \zeta_\alpha \hat{\sigma}_{x_\alpha}^2 \boldsymbol{\Omega})^{-1} \Phi_\alpha^T \mathbf{x}_\alpha, \quad \hat{\sigma}_{x_\alpha}^2 = \frac{1}{N_\alpha} (\mathbf{x}_\alpha - \Phi_\alpha \hat{\mathbf{c}}_\alpha)^T (\mathbf{x}_\alpha - \Phi_\alpha \hat{\mathbf{c}}_\alpha). \quad (7)$$

正則化パラメータ ζ_α および基底関数の個数 m の選択方法については、Konishi and Kitagawa (2008) を参照されたい。図 1 は、ガウス型基底関数を用いた非線形回帰モデルによる離散データの関数化の例を表している。ここで、丸印は離散データ、図下部の山形の曲線群が基底関数、図上部の曲線が得られた関数データである。

したがって、 $u_\alpha(t)$ の推定量 $\hat{u}_\alpha(t) = \hat{\mathbf{c}}_\alpha^T \boldsymbol{\phi}(t)$ を関数データ $x_\alpha(t)$ として、離散データ集合 $\{(t_{\alpha j}, x_{\alpha j}); j = 1, \dots, N_\alpha, \alpha = 1, \dots, n\}$ に替えて、関数データ集合

$$\{x_\alpha(t); t \in \mathcal{T}, \alpha = 1, \dots, n\} \quad (8)$$

をデータとして用いる。

3 半教師あり関数データ判別

3.1 関数ロジスティックモデル

いま、 n_1 組のラベルあり関数データ $\{(x_\alpha(t), g_\alpha); \alpha = 1, \dots, n_1\}$ と $(n - n_1)$ 組のラベルなし関数データ $\{x_\alpha(t); \alpha = n_1 + 1, \dots, n\}$ に基づいて L 群半教師あり関数ロジスティック判別方式を構成する。群の所属が既知の関数データ $x_\alpha(t)$ が観測されたとき、それが第 k 群からのものである事後確率を

$$\Pr(g_\alpha = k | x_\alpha), \quad k = 1, \dots, L,$$

とすると、関数データ集合に基づくロジスティックモデルは

$$\log \left\{ \frac{\Pr(g_\alpha = k | x_\alpha)}{\Pr(g_\alpha = L | x_\alpha)} \right\} = \beta_{kf} + \int x_\alpha(t) \beta_k(t) dt = \boldsymbol{\beta}_k^T \mathbf{z}_\alpha, \quad k = 1, \dots, L - 1, \quad (9)$$

として与えられる (Araki *et al.*, 2009)。ただし、 $\beta_k(t)$ は (3) 式の基底関数を用いることにより、

$$\beta_k(t) = \sum_{j=1}^m \beta_{kj} \phi_j(t) \quad (10)$$

と基底展開できるものと仮定し、 $\boldsymbol{\beta}_k = (\beta_{kf}, \beta_{k1}, \dots, \beta_{km})^T$, $\mathbf{z}_\alpha = (1, \mathbf{w}_\alpha^T J)^T$ とする。また、 J は基底関数交差積行列 $J = \int \phi(t)\phi^T(t)dt$ を表し、各 (i, j) 成分は以下で与えられる。

$$J_{ij} = \sqrt{\pi h^2} \exp\left\{-\frac{(\tau_{i+2} - \tau_{j+2})^2}{4h^2}\right\}. \quad (11)$$

一方、(9) 式を変形すると

$$\Pr(g_\alpha = k|x_\alpha) = \frac{\exp\{\boldsymbol{\beta}_k^T \mathbf{z}_\alpha\}}{1 + \sum_{j=1}^{L-1} \exp\{\boldsymbol{\beta}_j^T \mathbf{z}_\alpha\}}, \quad k = 1, \dots, L-1, \quad (12)$$

$$\Pr(g_\alpha = L|x_\alpha) = 1 - \sum_{k=1}^{L-1} \Pr(g_\alpha = k|x_\alpha) = \frac{1}{1 + \sum_{j=1}^{L-1} \exp\{\boldsymbol{\beta}_j^T \mathbf{z}_\alpha\}}, \quad (13)$$

が得られる。確率 $\Pr(g_\alpha = k|x_\alpha)$ はパラメータ $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_{L-1}^T)^T$ に依存しているため、 $\Pr(g_\alpha = k|x_\alpha) = \pi_k(x_\alpha; \boldsymbol{\beta})$ と記述するものとする。

次に、群のラベルを表す $(L-1)$ 次元ラベル変数 $\mathbf{Y} = (Y_1, \dots, Y_{L-1})^T$ を

$$\mathbf{y}_\alpha = (y_1^{(\alpha)}, \dots, y_{L-1}^{(\alpha)})^T = \begin{cases} (0, \dots, 0, \overset{(k)}{1}, 0, \dots, 0)^T & \text{if } g_\alpha = k, \quad (k = 1, \dots, L-1), \\ (0, \dots, 0)^T & \text{if } g_\alpha = L, \end{cases}$$

と定義する。このとき、ラベル変数ベクトル \mathbf{y}_α は多項分布

$$f(\mathbf{y}_\alpha|x_\alpha; \boldsymbol{\beta}) = \prod_{k=1}^{L-1} \pi_k(x_\alpha; \boldsymbol{\beta})^{y_k^{(\alpha)}} \{\pi_L(x_\alpha; \boldsymbol{\beta})\}^{1 - \sum_{j=1}^{L-1} y_j^{(\alpha)}} \quad (14)$$

に従う。また、 $(n - n_1)$ 個のラベルなし関数データ $\{x_\alpha(t); \alpha = n_1 + 1, \dots, n\}$ に対して、どの群から得られたかを表す観測変数

$$\mathbf{t}_\alpha = (t_1^{(\alpha)}, \dots, t_{L-1}^{(\alpha)})^T = \begin{cases} (0, \dots, 0, \overset{(k)}{1}, 0, \dots, 0)^T & \text{if } x_\alpha(t) \text{ が第 } k \text{ 群に属する,} \\ (0, \dots, 0)^T & \text{if } x_\alpha(t) \text{ が第 } L \text{ 群に属する,} \end{cases}$$

を別に得たとする。このとき、 \mathbf{t}_α は (14) 式と同じ確率 $\pi_k(x_\alpha; \boldsymbol{\beta})$ を持つ多項分布に従うと仮定することができる。したがって、 n_1 個のラベルあり関数データと $(n - n_1)$ 個のラベルなし関数データに基づいて、対数尤度関数を

$$\begin{aligned} \ell(\boldsymbol{\beta}) = & \sum_{\alpha=1}^{n_1} \left[\sum_{k=1}^{L-1} y_k^{(\alpha)} \boldsymbol{\beta}_k^T \mathbf{z}_\alpha - \log \left(1 + \sum_{l=1}^{L-1} \exp\{\boldsymbol{\beta}_l^T \mathbf{z}_\alpha\} \right) \right] \\ & + \sum_{\alpha=n_1+1}^n \left[\sum_{k=1}^{L-1} t_k^{(\alpha)} \boldsymbol{\beta}_k^T \mathbf{z}_\alpha - \log \left(1 + \sum_{l=1}^{L-1} \exp\{\boldsymbol{\beta}_l^T \mathbf{z}_\alpha\} \right) \right] \end{aligned} \quad (15)$$

として構成することができる。

また、将来のデータ $x_\alpha(t)$ は事後確率 $\Pr(g_\alpha = k|x_\alpha)$ ($k = 1, \dots, L-1$) が最大となる群に判別される。

3.2 モデルの推定

モデルのパラメータ $\beta = (\beta_1^T, \dots, \beta_{L-1}^T)^T$ は、対数尤度関数に正則化項を付与した次の正則化対数尤度関数の最大化に基づいて推定する。

$$\ell_\lambda(\beta) = \ell(\beta) - \frac{n_1 \lambda}{2} \sum_{k=1}^{L-1} \beta_k^T K \beta_k. \quad (16)$$

ここで、 $\lambda (> 0)$ は正則化パラメータであり、モデルの複雑さを調整する働きを持つとともに、モデルの安定化に寄与する機能も持つ。 K は $(m+1) \times (m+1)$ 型の行列であり、具体的には、

$$K = \begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & K^* \end{pmatrix} \quad (17)$$

で与えられる。ここで、 $\mathbf{0}$ は m 次元 0 ベクトル、 K^* は $m \times m$ 型の既知の非負値定符号行列である。 K^* としては、単位行列 I_m 、2 次差分行列 D_2 による $D_2^T D_2$ などさまざまなものが用いられるが (詳細については、例えば、Konishi and Kitagawa, 2008 参照)、本報告では単位行列 I_m を用いる。

しかし、(16) 式の正則化対数尤度関数は未知の量 $t_k^{(\alpha)}$ を含んでいるため、この正則化対数尤度関数の最大化は一般に困難である。そこで、変数 $t_k^{(\alpha)}$ を潜在変数と捉え、EM アルゴリズム (Dempster *et al.*, 1977) を適用することを考える。

EM アルゴリズムは不完全な状態で観測されたデータについて、最尤法 (正則化法) に基づいた推測を行うための推定アルゴリズムであり、統計学のみならず工学・医学・経済学などのさまざまな分野で用いられている (詳細については、例えば、McLachlan and Peel, 2000; 小西他, 2008 参照)。EM アルゴリズムを今回の半教師あり関数ロジスティックモデルに適用することにより、(16) 式の最大値を与える推定値 $\hat{\beta}$ は以下のアルゴリズムに従い求めることができる。

[推定アルゴリズム]

Step1 ラベルあり関数データ $\{(x_\alpha(t), g_\alpha); \alpha = 1, \dots, n_1\}$ のみに基づいて、パラメータ β を推定する。また、この関数の最大化には数値的最適化法の 1 つである Newton-Raphson 法を用いる (詳細については、例えば、Green and Silverman, 1994 参照)。

Step2 得られた推定値 $\hat{\beta}$ に基づいて L 群判別方式 $\hat{\Pr}(g_\alpha = k | x_\alpha) = \pi_k(x_\alpha; \hat{\beta})$ を構成する。

Step3 Step2 で得た判別方式を用いることにより、ラベルなし関数データ $x_\alpha(t)$ ($\alpha = n_1 + 1, \dots, n$) に対して事後確率 $\hat{\Pr}(g_\alpha = k | x_\alpha)$ ($k = 1, \dots, L$) を求める。この事後確率に従い、 t_α を各 α に対して、次のように推定する。

$$\hat{t}_\alpha = (\hat{t}_1^{(\alpha)}, \dots, \hat{t}_{L-1}^{(\alpha)})^T = (\hat{\Pr}(g_\alpha = 1 | x_\alpha), \dots, \hat{\Pr}(g_\alpha = L - 1 | x_\alpha))^T. \quad (18)$$

Step4 (16) 式に含まれる $t_k^{(\alpha)}$ を $\hat{t}_k^{(\alpha)}$ に置き換えた関数の最大化に基づいて推定値 $\hat{\beta}$ を得る。また、この関数の最大化には Newton-Raphson 法を用いる。

Step5 収束条件が満たされるまで Step2 から Step4 を繰り返す.

ここで, 上記のアルゴリズムの収束条件は, j 番目, $(j+1)$ 番目の更新における正則化対数尤度関数をそれぞれ $\ell_\lambda^{(j)}$, $\ell_\lambda^{(j+1)}$ と定めるとき, $|\ell_\lambda^{(j+1)} - \ell_\lambda^{(j)}| < \varepsilon$ とする (ε は十分小さい正の数とする).

(16) 式の最大化に基づいて得られる推定値 $\hat{\beta}$ を (14) 式に代入することにより, 統計モデル $f(\mathbf{y}_\alpha | x_\alpha; \hat{\beta})$ が得られる. しかし, この統計モデルは正則化パラメータ λ の値に依存しているため, このパラメータの値を適切に選択する必要がある. 次節では, このパラメータの値を適切に選択するために, 統計モデル $f(\mathbf{y}_\alpha | x_\alpha; \hat{\beta})$ を評価する基準をベイズ理論の観点から導出する.

4 モデル評価基準

Schwarz (1978) は, ベイズアプローチに基づいてモデル評価基準 BIC を提案した. BIC は, 最尤法によって推定されたモデルを評価するための基準であったが, Konishi *et al.* (2004) はパラメータ β の事前分布を (退化した) 正規分布と考えることで, 正則化法により推定されたモデルの事後確率のラプラス近似から得られるモデル評価基準を導出した.

$$\begin{aligned} \text{GBIC} = & -2 \sum_{\alpha=1}^{n_1} \log f(\mathbf{y}_\alpha | x_\alpha; \hat{\beta}) + n_1 \lambda \sum_{k=1}^{L-1} \hat{\beta}_k^T K \hat{\beta}_k + \log |R(\hat{\beta})| \\ & - (L-1) \log |K|_+ - (L-1)(m+1-d) \log \lambda - (L-1)d \log \left(\frac{2\pi}{n_1} \right). \end{aligned} \quad (19)$$

ここで, $|K|_+$ は行列 K の 0 でない固有値の積であり, d は行列 K の階数, $R(\hat{\beta})$ は次で与えられる $(L-1)(m+1) \times (L-1)(m+1)$ 型行列である.

$$R(\hat{\beta}) = -\frac{1}{n_1} (H \odot G)^T (H \odot G) + \frac{1}{n_1} I + \lambda J. \quad (20)$$

ここで, \odot は行列の成分毎の積を表す. G は $G = (Z, \dots, Z)$ となる $n_1 \times (L-1)(m+1)$ 型行列であり ($Z = (z_1, \dots, z_{n_1})^T$), $H = (\boldsymbol{\pi}_1 \mathbf{1}_{m+1}^T, \dots, \boldsymbol{\pi}_{L-1} \mathbf{1}_{m+1}^T)$ ($\boldsymbol{\pi}_k = (\Pr(g_1 = k | \mathbf{x}_1), \dots, \Pr(g_{n_1} = k | \mathbf{x}_{n_1}))$), $I = \text{diag}\{Z^T \text{diag}(\boldsymbol{\pi}_1) Z, \dots, Z^T \text{diag}(\boldsymbol{\pi}_{L-1}) Z\}$, $J = \text{diag}\{K, \dots, K\}$ である.

モデル評価基準 GBIC の値を最小にする正則化パラメータ λ の値を選択し, 対応するモデルを最適なモデルとする.

5 適用例

Spellman *et al.* (1998) による遺伝子発現プロファイルデータへの適用を通して提案手法の有効性を検証する. このデータは, 77 時点における 6178 個の酵母菌の遺伝子発現量を測定したものであり, 今回は 24 時点における細胞周期が同定された欠損値を除く 632

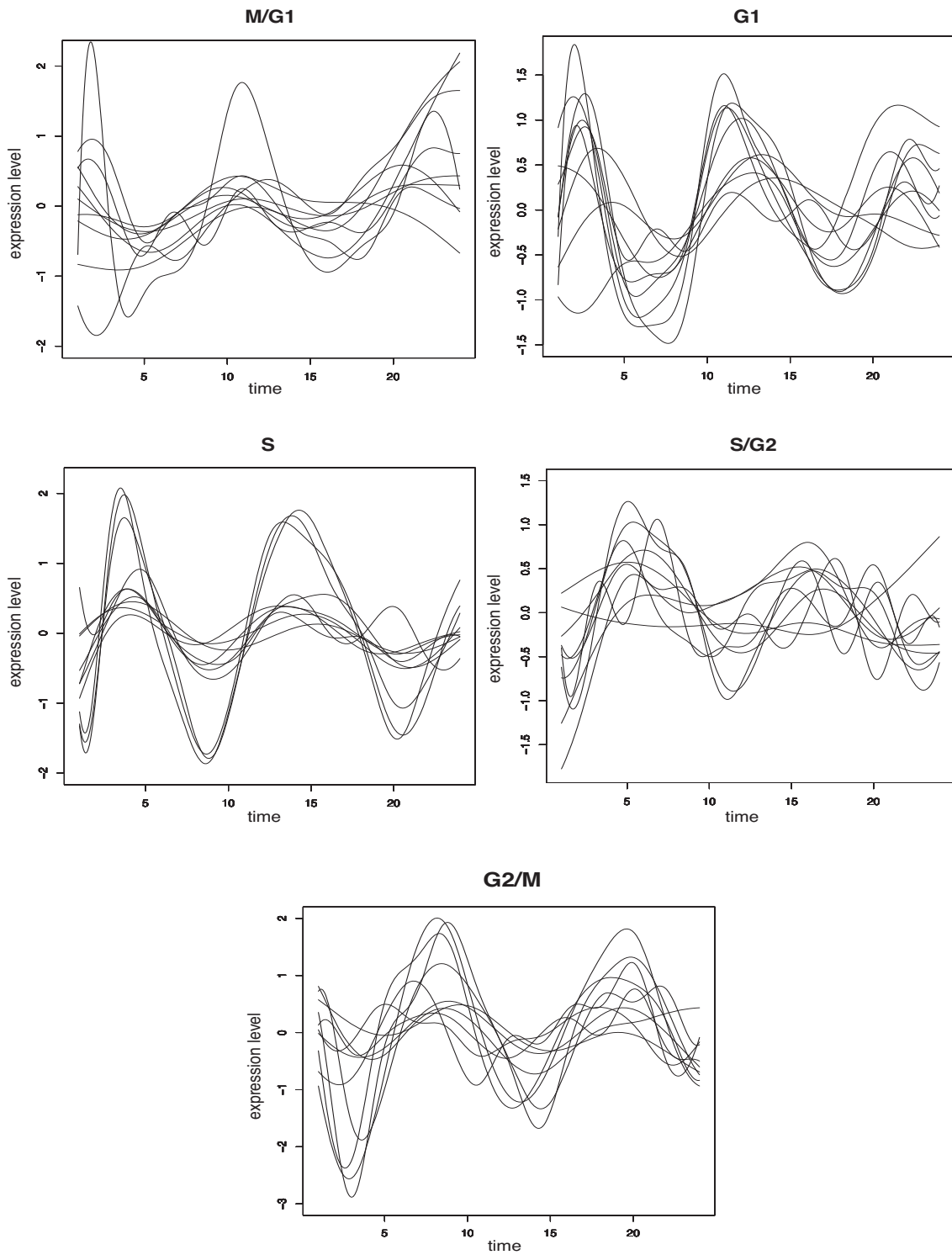


図 2: 各群における関数データ集合の例. 各群にはそれぞれ 10 本の関数データが描かれている.

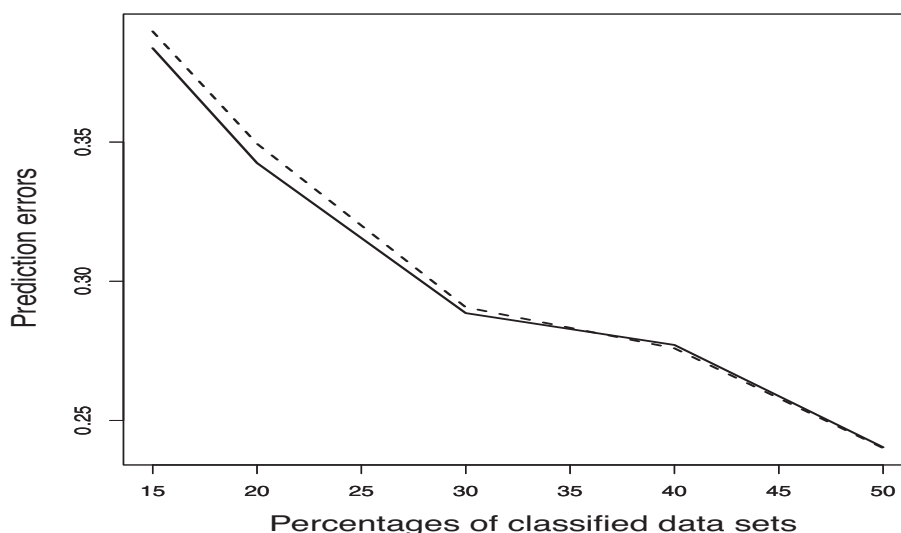


図 3: 10 回の反復に対する予測誤差の平均の推移 (実線：提案手法，破線：教師あり関数ロジスティック判別手法)

個の遺伝子データを用いた。また，細胞周期は M/G1, G1, S, S/G2, G2/M の 5 群に分類されている。

まず，各遺伝子に対して，得られた 24 時点の離散データに非線形回帰モデルをあてはめることにより，図 2 のような関数データ集合を得た。次に，得られた 632 個の関数データを無作為に 300 個の学習データと 332 個のテストデータに分けた。また，半教師あり学習を行うために，学習データ数の 15%, 20%, 30%, 40%, 50% をラベルあり関数データと無作為に割り当て，各ラベルあり関数データの割合に対して残りをラベルなし関数データと設定した。

適用に際して，得られた関数データに基づいて半教師あり関数ロジスティックモデルを構築し，332 個のテストデータを用いて予測誤差を計算した。また，正則化パラメータ λ の値は 4 節で導出したモデル評価基準 GBIC の最小化に基づいて選択した。比較対象として，Araki *et al.* (2009) により提案された教師あり関数ロジスティック判別モデルを用いた。ここで，上記の教師あり関数ロジスティックモデルは，ラベルあり関数データのみに基づいて推定されることに注意されたい。

図 3 は 10 回の反復により得られた予測誤差の平均の推移を表している。ここで，実線が提案手法，破線が教師あり関数ロジスティック判別手法 (Araki *et al.*, 2009) を示す。図 3 から，提案手法は，学習データ中のラベルあり関数データの割合が少ないときに予測誤差を小さくしているのが見て取れる。これより，少量のラベルあり関数データと大量のラベルなし関数データが得られたという状況においては，提案手法は従来の教師あり判別手法より有効に働くものと考えられる。

参考文献

- [1] Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2009): Functional logistic discrimination via regularized basis expansions. *Communications in Statistics - Theory and Methods (to appear)*.
- [2] Chapelle, O., Schölkopf, B. and Zien, A. (2006): *Semi-Supervised Learning*. Cambridge, MA: MIT Press.
- [3] Chapelle, O. and Zien, A. (2005): Semi-supervised classification by low density separation. *Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics*, 57–64.
- [4] Dean, N., Murphy, T.B. and Downey, G. (2006): Using unlabelled data to update classification rules with applications in food authenticity studies. *Journal of the Royal Statistical Society C*, **55**, 1–14.
- [5] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977): Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, **39**, 1–38.
- [6] Fujii, T. and Konishi, S. (2006): Nonlinear regression modeling via regularized wavelets and smoothing parameter selection. *Journal of Multivariate Analysis*, **97**, 2023–2033.
- [7] Green, P.J. and Silverman, B.W. (1994): *Nonparametric Regression and Generalized Linear Models*. London: Chapman & Hall.
- [8] Imoto, S. and Konishi, S. (2003): Selection of smoothing parameters in B -spline nonparametric regression models using information criteria. *Annals of the Institute of Statistical Mathematics*, **55**, 671–687.
- [9] Kawano, S. and Konishi, S. (2007): Nonlinear regression modeling via regularized Gaussian basis functions. *Bulletin of Informatics and Cybernetics*, **39**, 83–96.
- [10] Konishi, S., Ando, T. and Imoto, S. (2004): Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*, **91**, 27–43.
- [11] Konishi, S. and Kitagawa, G. (2008): *Information Criteria and Statistical Modeling*. New York: Springer.
- [12] 小西貞則, 越智義道, 大森裕浩 (2008): 計算統計学の方法. 朝倉書店.
- [13] McLachlan, G.J. and Peel, D. (2000): *Finite Mixture Models*. New York: Wiley.

- [14] Nigam, K., McCallum, A., Thrun, S. and Mitchell, T. (2000): Text classification from labeled and unlabeled documents using EM. *Machine Learning*, **39**, 103–134.
- [15] Pan, W., Shen, X., Jiang, A. and Hebbel, R.P. (2006): Semi-supervised learning via penalized mixture model with application to microarray sample classification. *Bioinformatics*, **22**, 2388–2395.
- [16] Ramsay, J.O. and Silverman, B.W. (2002): *Applied Functional Data Analysis*. New York: Springer.
- [17] Ramsay, J.O. and Silverman, B.W. (2005): *Functional Data Analysis*. Second Edition. New York: Springer.
- [18] Schwarz, G. (1978): Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- [19] Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Bostein, D., Futcher, B. (1998): Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, **9**, 3273–3297.