

# Classification via principal differential analysis

Eunseong Jang<sup>a</sup>, Yaeji Lim<sup>1, a</sup>

<sup>a</sup>Department of Applied Statistics, Chung-Ang University, Korea

---

## Abstract

We propose principal differential analysis based classification methods. Computations of squared multiple correlation function (RSQ) and principal differential analysis (PDA) scores are reviewed; in addition, we combine principal differential analysis results with the logistic regression for binary classification. In the numerical study, we compare the principal differential analysis based classification methods with functional principal component analysis based classification. Various scenarios are considered in a simulation study, and principal differential analysis based classification methods classify the functional data well. Gene expression data is considered for real data analysis. We observe that the PDA score based method also performs well.

**Keywords:** functional data analysis, classification, principal differential analysis, functional principal component analysis

---

## 1. Introduction

Functional data analysis (FDA) is a popular statistical method that deals with the analysis and theory of data in the form of functions, images, shapes, and general objects. Basic ideas for FDA were first proposed by Grenander (1950); statistical features and analysis of FDA are established by Ramsay (1982) and Ramsay and Dalzell (1991).

Principal differential analysis (PDA) is a popular technique that estimates a differential operator from a functional data. It was first proposed by Ramsay (1996) to find low dimensional approximations to functional data. PDA finds linear differential operator  $L$ ,

$$L = \beta_0 + \cdots + \beta_{m-1}D^{m-1} + D^m,$$

where  $D^m = d^m/dt^m$  is  $m^{\text{th}}$  derivative, such that  $Lx = 0$  for functional observation  $x$ . The dimension reduction is achieved by projecting the curve onto the null space of the differential operator. PDA has been applied to various fields with related theories still being expanded. For example, Jin *et al.* (2013) extended PDA to allow for coefficients in the linear differential operator to smoothly depend upon a single continuous covariate. Staniswalis *et al.* (2017) proposed local PDA and analyzed the evoked brain potential curves of children.

Functional principal component analysis (FPCA) is another approach for identifying a low-dimensional nonparametric basis for describing functional data; in addition, PDA is often compared with the FPCA (Jin *et al.*, 2013; Dalla *et al.*, 2014). FPCA finds eigenfunctions, which analogue to eigenvectors in multivariate principal component analysis (MPCA), and then calculates functional principal component scores (FPC scores) to explain the variance of observed functional data.

---

<sup>1</sup> Corresponding author: Department of Applied Statistics, Chung-Ang University, 47, Heukseok-ro, Dongjak-Gu, Seoul 06974, Korea. E-mail: yaeji.lim@gmail.com

PDA and FPCA are similar methods in view of low dimensional approximation, and both methods can be used for classification problem. Functional data classification can be performed by functional regression models that feature class labels as responses and the observed functional data as predictors, and most functional data classification methods apply a dimension reduction technique using a truncated expansion in a pre-specified function basis. Müller (2005) proposed functional binary regression models based on FPCA and applied the method to the primary biliary cirrhosis patients data, in order to classify early time courses of bilirubin. Nie *et al.* (2018) also considered FPC scores in logistic regression, then classified alcoholic and control groups using electroencephalography (EEG) dataset.

For the PDA based classification methods, Reimer and Rudzicz (2010) proposed a PDA classifier based on RSQ, and Sattar and Rudzicz (2016) extended the approach by using a support vector machine. Dalla *et al.* (2014) compared quadratic discriminant analysis with PDA components and that of the FPC scores.

In this paper, we consider two PDA based classification methods, RSQ based method and PDA score based method. Unlike other studies that used RSQ and PDA scores for the classification (Reimer and Rudzicz, 2010; Dalla *et al.*, 2014), we use them as covariates in logistic regression model and examine their role as predictors.

The paper is organized as follows. In Section 2, we briefly review the PDA and the classification methods based on PDA are introduced. Simulation results are presented in Section 3, and real data analysis is in Section 4. Finally, in Section 5, some concluding remarks are reported.

## 2. Methodology

### 2.1. Principal differential analysis

Here, we briefly review the PDA. PDA assumes that given functional data  $x(t)$  follows a linear differential operator  $L$  for  $t \in \mathbb{R}$ ;

$$Lx(t) = \beta_0(t)x(t) + \cdots + \beta_{m-1}(t)D^{m-1}x(t) + D^m x(t) = f(t),$$

where  $\beta_j(t)$  is called as weight function or coefficient, and  $f(t)$  is forcing function, which is set to 0 for homogeneous case. Throughout the paper, we set  $f(t) = 0$ .

Given  $N$  observed functional data,  $x_i(t)$ ,  $i = 1, \dots, N$ , coefficients  $\{\beta_j(t)\}_{j=0}^{m-1}$  can be estimated by least squares approach defined as,

$$\text{SSE}_L = \sum_{i=1}^N \int [Lx_i(t)]^2 dt = \sum_{i=1}^N \|Lx_i\|^2. \quad (2.1)$$

In matrix form, (2.1) can be represented as

$$\text{SSE}_L = [\mathbf{w}(t) - \mathbf{Z}(t)\boldsymbol{\beta}(t)]' [\mathbf{w}(t) - \mathbf{Z}(t)\boldsymbol{\beta}(t)],$$

where  $\mathbf{w}(t) = (w_1(t), \dots, w_N(t))'$ , with  $w_i(t) = D^m x_i(t)$ ,  $\boldsymbol{\beta}(t) = (\beta_0(t), \dots, \beta_{m-1}(t))'$ , and  $\mathbf{Z}(t)$  is  $N \times m$  matrix whose  $i^{\text{th}}$  row is  $\mathbf{z}_i(t) = \{-x_i(t), \dots, -D^{m-1}x_i(t)\}$ , for  $i = 1, \dots, N$ . Then, the least square solution can be obtained as

$$\hat{\boldsymbol{\beta}}(t) = [\mathbf{Z}(t)' \mathbf{Z}(t)]^{-1} \mathbf{Z}(t)' \mathbf{w}(t). \quad (2.2)$$

The order  $m$  is often determined by theoretical backgrounds of specific domain. If there is no fundamental theory specifying  $m$  for given data, the order can be selected by cross-validation classification error rate.

A more detailed description of PDA can be found in Ramsay and Silverman (2005).

## 2.2. Smoothing

Random error may be included in observation in analyzing real data; however, it makes the curvature of observed data complex or too variable. Such variation can cause low performance in FDA including PDA. To solve the problem, smoothing techniques are usually applied to reduce the roughness of the observed functional data (Rice, 2004). Popular smoothing methods are basis-expansion, smoothing penalties, and kernel methods.

In this paper, we represent functional data  $x_i(t)$ ,  $i = 1, \dots, N$ , as linear combination of  $K$  basis functions,  $\{\phi_k\}$ , which are smooth, continuous, and differentiable functions;

$$x_i(t) = \sum_{k=1}^K c_{ik} \phi_k(t) + \epsilon_i(t) = \mathbf{c}_i' \boldsymbol{\phi} + \epsilon_i(t), \quad i = 1, \dots, N, \quad (2.3)$$

where  $\mathbf{c}_i = (c_{i1}, \dots, c_{iK})'$  with constant  $c_{ik}$ 's,  $\boldsymbol{\phi} = (\phi_1(t), \dots, \phi_K(t))'$ , and  $\epsilon_i(t)$  is random error. Fourier basis, polynomial basis, and B-spline basis are popular basis functions for  $\boldsymbol{\phi}$  (Kosarev and Pantos, 1983; Wand, 2000).

Once the type of basis is determined, coefficient  $\mathbf{c}_i$  is estimated by minimizing ordinary least squares as follow (Ramsay and Silverman, 2005),

$$\text{SMSSE} = \sum_{i=1}^N \sum_{j=1}^T \left[ x_i(t_j) - \sum_{k=1}^K c_{ik} \phi_k(t_j) \right]^2,$$

where  $t_j$ , for  $j = 1, \dots, T$ , is equally spaced time point.

A high degree of smoothing may be obtained by reducing the order of the basis expansion; however, a finer control of the smoothing degree is achieved by using a roughness penalty approach as,

$$\text{PENSSE} = \sum_{i=1}^N \sum_{j=1}^T \left[ x_i(t_j) - \sum_{k=1}^K c_{ik} \phi_k(t_j) \right]^2 + \lambda \int [D^2 x_i(t)]^2 dt, \quad (2.4)$$

where  $\lambda$  is a smoothing parameter controls roughness of the data. If we set large  $\lambda$ , the curvature of data becomes simpler. The optimal value of the smoothing parameter  $\lambda$  can be estimated using proper criterion such as generalized cross-validation (GCV) (Craven and Wahba, 1979).

Throughout this paper, we use Fourier basis and B-spline basis and  $\lambda$  is selected by GCV.

## 2.3. PDA classification

Given functional curve  $x_i(t)$ ,  $i = 1, \dots, N$ , and its corresponding label variable  $y_i \in \{0, 1\}$ ,  $i = 1, \dots, N$ , we consider two different classification methods based on PDA. One is RSQ based method, and the other is PDA scores based method.

### 2.3.1. RSQ based classification

RSQ is one of the popular tool to measure the goodness-of-fit in PDA (Ramsay, 1996) and defined by point-wise squared multiple correlation function;

$$\text{RSQ} := \frac{\sum_{i=1}^N \|D^m x_i\|^2 - \sum_{i=1}^N \|Lx_i\|^2}{\sum_{i=1}^N \|D^m x_i\|^2} = \frac{\text{SSE}_0 - \text{SSE}_L}{\text{SSE}_0},$$

where  $SSE_0 = \sum_{i=1}^N \|D^m x_i\|^2$  is a logical baseline against  $SSE_L$  in (2.1). Then, RSQ has value from 0 to 1, and value equal to 1 implies perfect fit.

In this paper, we use RSQ as the covariate in the logistic regression. To do that, we compute individual RSQ value for each curve,  $RSQ_{i,c}$ , for  $i = 1, \dots, N$ , using the curves in class  $c$ . We consider functional data in class  $c \in C$ ,  $\{x_1^c, \dots, x_{N_c}^c\}$ , where  $N_c$  is the number of curves in class  $c$ . The curves are smoothed by (2.4), and then linear differential operator,  $L^c$ , is estimated by PDA for each class  $c$ . Then  $RSQ_{i,c}$ , for  $i = 1, \dots, N$ , is defined as

$$RSQ_{i,c} := \frac{\|D^m x_i\|^2 - \|L^c x_i\|^2}{\|D^m x_i\|^2}. \quad (2.5)$$

Now,  $RSQ_{i,c}$  with  $c \in \{1, 2\}$  are used as predictors for the logistic regression as follows;

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 RSQ_{i,1} + \alpha_2 RSQ_{i,2} + \epsilon_i, \quad i = 1, \dots, N,$$

where  $p_i = \Pr(c_i = 2)$ . For the binary classification, we use a classification threshold 0.5.

### 2.3.2. PDA scores based classification

Similar to the scores defined in the FPCA, we can define PDA scores for functional data. The  $k^{th}$  PDA scores of  $x_i(t)$  are defined as

$$s_{ik} := \sum_{j=1}^T (x_i(t_j) - \mu(t_j)) \psi_k(t_j), \quad i = 1, \dots, N, \quad k = 1, \dots, m,$$

where  $\mu(t)$  is the mean curve, and the eigenfunctions,  $\{\psi_k(t)\}_{k=1}^m$ , are estimated by resolving differential equation (Dalla *et al.*, 2014). Once the coefficients  $\hat{\beta}(t)$  are obtained through (2.2), the eigenfunctions are estimated as

$$\hat{\psi}_1(t) := e^{\hat{\lambda}_1 t}, \dots, \hat{\psi}_K(t) := e^{\hat{\lambda}_m t},$$

where  $\{\hat{\lambda}_k\}_{k=1}^m$  are complex roots of the characteristic polynomial

$$\lambda^m + \hat{\beta}_{m-1} \lambda^{m-1} + \dots + \hat{\beta}_1 \lambda + \hat{\beta}_0 = 0.$$

Then, PDA scores are also used as predictors for the binary classification methods. For the logistic regression, the formula can be written as

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 \hat{s}_{i1} + \dots + \alpha_m \hat{s}_{im} + \epsilon_i, \quad i = 1, \dots, N,$$

where  $p_i = \Pr(c_i = 2)$ . For the binary classification, we use a classification threshold 0.5.

One may use other conventional classification method, such as support vector machine (SVM) or random forest, rather than the logistic regression.

## 3. Simulation study

### 3.1. Simulation data

For each group  $c \in \{1, 2\}$ , we generate 100 functional data,  $x_i^c(t_j)$ , for  $i = 1, \dots, 100$ , over the interval  $t_j \in [-5, 5]$ . Time points are equally spaced as  $t_j = -5 + j/20$ , for  $j = 1, \dots, 200$ .

Based on the following five scenarios, we generate total  $N = 200$  curves. The simulation settings are motivated from Dalla *et al.* (2014).

- Simulation I

$$\begin{aligned} x_i^1(t) &= a_{i1} \cos(\pi t) + a_{i2} \sin(\pi t) + b\epsilon_i(t), \quad i = 1, \dots, 100, \\ x_i^2(t) &= a_{i1} \cos(3\pi t) + a_{i2} \sin(3\pi t) + b\epsilon_i(t), \quad i = 1, \dots, 100, \end{aligned}$$

where the random coefficient  $a_{ik}$ , for  $k = 1, 2$  is generated from  $N(0, 1)$  or  $N(1, 1)$ , and constant weight  $b$  is set to 0.1 or 1.  $\epsilon_i$ 's generated from  $N(\mathbf{0}, \Sigma)$ , with three covariance structures, simple  $\Sigma_{\text{ind}}$ , compound symmetry  $\Sigma_{\text{CS}}$  and the first-order auto-regressive  $\Sigma_{\text{AR}(1)}$  (Kincaid, 2005);

$$\Sigma_{\text{ind}} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{pmatrix}, \quad \Sigma_{\text{CS}} = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & \ddots & & \vdots \\ \rho & \rho & \cdots & \cdots & 1 \end{pmatrix}, \quad \Sigma_{\text{AR}(1)} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{99} \\ \rho & 1 & \rho & \cdots & \rho^{98} \\ \vdots & & \ddots & & \vdots \\ \rho^{99} & \rho^{98} & \cdots & \cdots & 1 \end{pmatrix},$$

with  $\rho = 0.1, 0.4$ , and  $0.7$ .

From the Simulation II to V, all coefficients are generated in the same way as Simulation I, unless specifically stated.

- Simulation II

$$\begin{aligned} x_i^1(t) &= a_{i1} e^{0.25t} \cos(\pi t) + a_{i2} e^{0.25t} \sin(\pi t) + b\epsilon_i(t), \quad i = 1, \dots, 100, \\ x_i^2(t) &= a_{i1} e^{0.5t} \cos(3\pi t) + a_{i2} e^{0.5t} \sin(3\pi t) + b\epsilon_i(t), \quad i = 1, \dots, 100. \end{aligned}$$

- Simulation III

$$\begin{aligned} x_i^1(t) &= a_{i1} (0.01t + 0.4t^2 - 0.003t^3 + 0.0006t^4) + b\epsilon_i(t), \quad i = 1, \dots, 100, \\ x_i^2(t) &= 3a_{i1} \cos(3\pi t) + 3a_{i2} \sin(3\pi t) + b\epsilon_i(t), \quad i = 1, \dots, 100. \end{aligned}$$

- Simulation IV

$$\begin{aligned} x_i^1(t) &= a_{i1} \sin(\pi t) + b\epsilon_i(t), \quad i = 1, \dots, 100, \\ x_i^2(t) &= a_{i1} \cos(\pi t) + b\epsilon_i(t), \quad i = 1, \dots, 100, \end{aligned}$$

where  $a_{i1} \sim N(\mu, 1)$  with  $\mu = 1, 2$ .

- Simulation V

$$\begin{aligned} x_i^1(t) &= a_{i1} \{\cos(\pi t) + 0.1e^t\} + b\epsilon_i(t), \quad i = 1, \dots, 100, \\ x_i^2(t) &= a_{i2} \{\cos(\pi t) + 0.1e^t\} + b\epsilon_i(t), \quad i = 1, \dots, 100, \end{aligned}$$

where  $a_{i1} \sim N(\mu, 1)$  and  $a_{i2} \sim N(-\mu, 1)$  with  $\mu = 1, 2$ .

The generated functional curves from each scenario are presented in Figure 1 and Figure 2. For each data set, PDA classification methods using RSQ and PDA scores are performed. The results are compared with the conventional FPCA classification (Ramsay and Silverman, 2005; Leng and Müller, 2006). We consider one to five PCs in FPCA and the best performed results are presented in the paper.

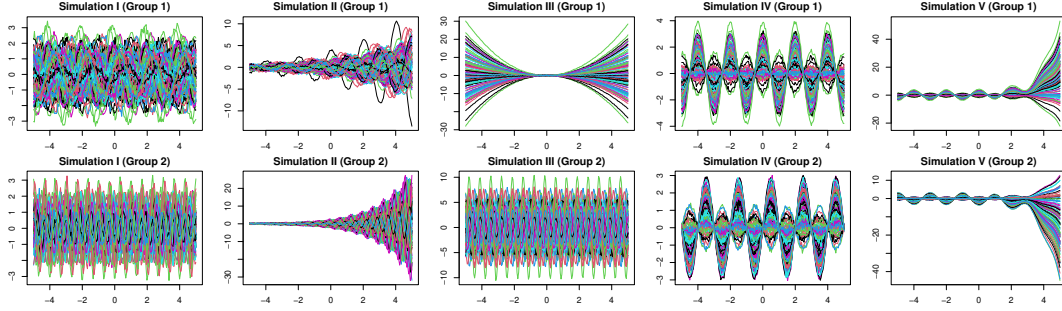


Figure 1: (Left to right) Data curves from simulation I to V when  $a_{ik} \sim N(0, 1)$ ,  $k = 1, 2$  for simulation I to III,  $a_{ik} \sim N(1, 1)$  for simulation IV,  $a_{i1} \sim N(1, 1)$ ,  $a_{i2} \sim N(-1, 1)$  for simulation V, and  $b = 0.1$ .

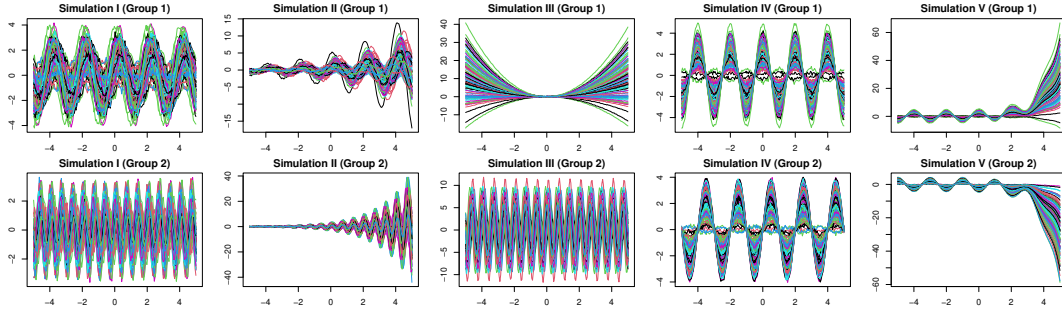


Figure 2: (Left to right) Data curves from simulation I to V when  $a_{ik} \sim N(1, 1)$ ,  $k = 1, 2$  for simulation I to III,  $a_{ik} \sim N(2, 1)$  for simulation IV,  $a_{i1} \sim N(2, 1)$ ,  $a_{i2} \sim N(-2, 1)$  for simulation V, and  $b = 0.1$ .

### 3.2. Results

We divide the data set into training and test data, where the training data is 80% of whole data. B-spline and Fourier basis are considered for smoothing. The number of basis functions is determined by grid search algorithm. For the PDA classification methods, the optimal order of differential equation,  $m$ , is obtained by 8-fold cross validation.

The classification results are validated by classification error rate (CER), calculated as

$$\text{CER} = 1 - \text{accuracy} = 1 - \frac{\text{number of classified correct}}{\text{number of classified total}}.$$

To provide the lowest achievable error rate, the estimated Bayes error rate for each data generating scheme is also presented. It is hard to obtain Bayes error directly (Fukunaga, 2013), we estimate the error based on the nearest neighbor (NN) classifiers (Cover and Hart, 1967; Tumer and Ghosh, 2003). Consider  $E_{\text{bayes}} = 1 - \sum_{i=1}^2 \int_{C_i} P(c_i) p(x|c_i) dx$ , where  $C_i$  is the region where class  $i$  has the highest posterior,  $P(c_i)$  is the priori class probability of class  $i$ , and  $p(x|c_i)$  is the class likelihood. Then, for a two-class problem with sufficiently large data, we have

$$\frac{1}{2} (1 - \sqrt{1 - 2E_{\text{NN}}}) \leq E_{\text{bayes}} \leq E_{\text{NN}},$$

where  $E_{\text{NN}} = \lim_{n \rightarrow \infty} P(\theta \neq \theta'_n | x, x'_n)$ . Here,  $\theta$  denotes the true class of  $x$  and  $\theta'_n$  denotes the class of  $x_n$ , where  $x_n$  is the nearest neighbor of  $x$ . For more details, see Cover and Hart (1967).

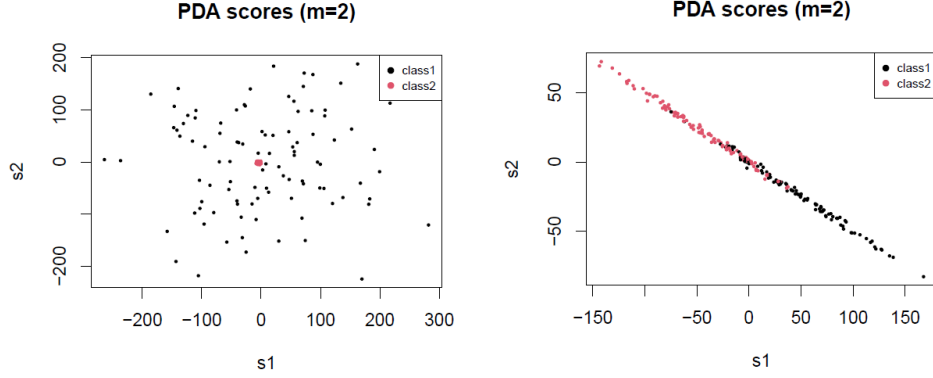


Figure 3: (Left) the scatter plots of the PDA scores,  $(s_{i1}, s_{i2})$  for simulation I and (Right) simulation V.

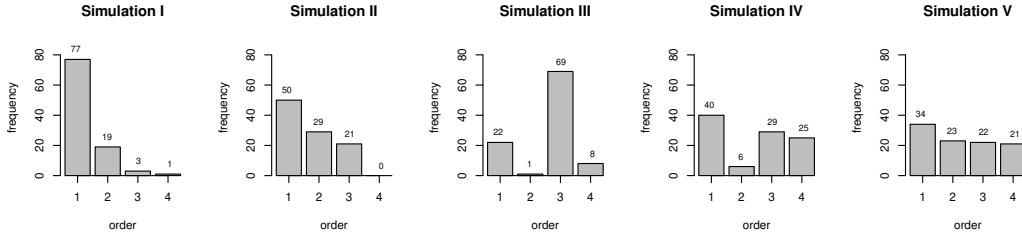


Figure 4: The histograms of selected order  $m$  in 100 repetitions. For the Simulation I–III, we set  $a_{ik} \sim N(0, 1)$ , and for the Simulation IV and V, we set  $\mu = 1$  in the generation of  $a_{ik}$ . Also,  $b = 0.1$  and independent covariance structure assumed.

For the implementation of PDA and FPCA, R package “fda.usc” and “fda” are used for computation. We repeat the process 100 times with different training-test set division. The averaged error rates and their standard error are presented in Tables 1–5. Marked in boldface are the lowest error rate for each case.

Note that the results from Fourier basis are not different from that of B-spine, results from Fourier basis are omitted. Also, the classification based on SVM gives similar results to logistic regression, therefore we omit the results from SVM. The whole results are available in [https://github.com/yaejili/PDA\\_clustering](https://github.com/yaejili/PDA_clustering).

From the tables, PDA score based method shows similar performance as FPCA classification. However, RSQ based method shows a different performance. The simulation I to III, RSQ based method shows the superiority over the two score based methods, while in simulation IV and V, RSQ based method works poorly. From the (2.5), we can expect that the RSQ cannot distinguish two curves that have the same structure, but differ in amplitude only, as in the simulation V. Also, if the two curves have similar norm as in the simulation IV, the RSQ based method fails, even worse than random-guessing.

Note that the range of RSQ values when  $\Sigma = \Sigma_{\text{ind}}$  and  $b = 0.1$  is  $-8.88 \times 10^{-4} \sim 4.09 \times 10^{-4}$  for class 1 and  $-8.89 \times 10^{-4} \sim 4.09 \times 10^{-4}$  for class 2 in the simulation V ( $\mu = 1$ ), and  $-1.01 \times 10^{-8} \sim 2.22 \times 10^{-8}$  for class 1 and  $-1.23 \times 10^{-7} \sim 5.54 \times 10^{-8}$  for class 2 in the simulation IV ( $a_{ij} = N(1, 1)$ ). On the other hand, the range of RSQ values is  $-1.78 \times 10^{-4} \sim 1.83 \times 10^{-6}$  for class 1 and  $-1.24 \times 10^{-3} \sim 3.71 \times 10^{-6}$  for class 2 in the simulation I ( $a_{ij} = N(0, 1)$ ).

Table 1: Averaged error rate and its standard error in parenthesis for B-spline smoothed data of the simulation I

	$b$	$a_{ij}$	Bayes error	Method	CER
$\Sigma_{\text{ind}}$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.45 (0.11)
				FPCA	0.53 (0.17)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.06 (0.03)
				FPCA	0.08 (0.04)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.07$	PDA(RSQ)	<b>0.02 (0.02)</b>
				PDA(scores)	0.45 (0.08)
				FPCA	0.52 (0.14)
$\Sigma_{\text{CS}}, \rho = 0.1$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.44 (0.12)
				FPCA	0.53 (0.17)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.06 (0.03)
				FPCA	0.08 (0.04)
	1	$N(0, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.11$	PDA(RSQ)	<b>0.04 (0.03)</b>
				PDA(scores)	0.55 (0.07)
				FPCA	0.55 (0.08)
$\Sigma_{\text{CS}}, \rho = 0.4$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.44 (0.10)
				FPCA	0.53 (0.17)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.06 (0.03)
				FPCA	0.08 (0.03)
	1	$N(0, 1)$	$0.02 \leq E_{\text{bayes}} \leq 0.12$	PDA(RSQ)	<b>0.05 (0.04)</b>
				PDA(scores)	0.53 (0.07)
				FPCA	0.56 (0.05)
$\Sigma_{\text{AR}(1)}, \rho = 0.1$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.02)</b>
				PDA(scores)	0.45 (0.11)
				FPCA	0.53 (0.17)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.06 (0.03)
				FPCA	0.08 (0.04)
	1	$N(0, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.09$	PDA(RSQ)	<b>0.02 (0.02)</b>
				PDA(scores)	0.48 (0.09)
				FPCA	0.54 (0.14)
$\Sigma_{\text{AR}(1)}, \rho = 0.4$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.46 (0.12)
				FPCA	0.53 (0.18)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.06 (0.03)
				FPCA	0.08 (0.04)
	1	$N(0, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.09$	PDA(RSQ)	<b>0.08 (0.03)</b>
				PDA(scores)	0.48 (0.07)
				FPCA	0.51 (0.11)
$\Sigma_{\text{AR}(1)}, \rho = 0.7$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.02)</b>
				PDA(scores)	0.45 (0.12)
				FPCA	0.53 (0.17)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.01 (0.02)</b>
				PDA(scores)	0.06 (0.03)
				FPCA	0.08 (0.04)
	1	$N(0, 1)$	$0.07 \leq E_{\text{bayes}} \leq 0.14$	PDA(RSQ)	<b>0.16 (0.05)</b>
				PDA(scores)	0.50 (0.07)
				FPCA	0.54 (0.08)
		$N(1, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.07$	PDA(RSQ)	<b>0.06 (0.03)</b>
				PDA(scores)	0.16 (0.05)
				FPCA	0.09 (0.04)



Table 2: Averaged error rate and its standard error in parenthesis for B-spline smoothed data of the simulation II

	$b$	$a_{ij}$	Bayes Error	Method	CER
$\Sigma_{\text{ind}}$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.44 (0.08)
				FPCA	0.42 (0.18)
	0.1	$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.10 (0.05)
				FPCA	0.08 (0.03)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.45 (0.07)
				FPCA	0.43 (0.18)
$\Sigma_{\text{CS}}, \rho = 0.1$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.44 (0.08)
				FPCA	0.42 (0.18)
	0.1	$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.10 (0.05)
				FPCA	0.08 (0.03)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.49 (0.08)
				FPCA	0.43 (0.18)
$\Sigma_{\text{CS}}, \rho = 0.4$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.02)</b>
				PDA(scores)	0.46 (0.07)
				FPCA	0.43 (0.18)
	0.1	$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.10 (0.04)
				FPCA	0.08 (0.03)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.03$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.14 (0.05)
				FPCA	0.08 (0.03)
$\Sigma_{\text{AR}(1)}, \rho = 0.1$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.44 (0.08)
				FPCA	0.42 (0.17)
	0.1	$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.10 (0.04)
				FPCA	0.08 (0.03)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.43 (0.07)
				FPCA	0.42 (0.17)
$\Sigma_{\text{AR}(1)}, \rho = 0.4$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.45 (0.08)
				FPCA	0.42 (0.17)
	0.1	$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.10 (0.04)
				FPCA	0.08 (0.03)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.02 (0.02)</b>
				PDA(scores)	0.47 (0.06)
				FPCA	0.44 (0.17)
$\Sigma_{\text{AR}(1)}, \rho = 0.7$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.45 (0.08)
				FPCA	0.42 (0.17)
	0.1	$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.10 (0.05)
				FPCA	0.08 (0.03)
	1	$N(0, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.03$	PDA(RSQ)	<b>0.03 (0.02)</b>
				PDA(scores)	0.50 (0.08)
				FPCA	0.43 (0.16)
$\Sigma_{\text{AR}(1)}, \rho = 0.7$	1	$N(0, 1)$	$0.02 \leq E_{\text{bayes}} \leq 0.05$	PDA(RSQ)	<b>0.02 (0.02)</b>
				PDA(scores)	0.18 (0.05)
				FPCA	0.08 (0.03)
	1	$N(1, 1)$	$0.02 \leq E_{\text{bayes}} \leq 0.05$	PDA(RSQ)	<b>0.02 (0.02)</b>

Table 3: Averaged error rate and its standard error in parenthesis for B-spline smoothed data of the simulation III

	$b$	$a_{ij}$	Bayes Error	Method	CER
$\Sigma_{\text{ind}}$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.47 (0.15)
				FPCA	0.51 (0.19)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.07 (0.04)
				FPCA	0.07 (0.04)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.04$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.47 (0.13)
				FPCA	0.52 (0.08)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.08 (0.05)
				FPCA	0.07 (0.04)
$\Sigma_{\text{CS}}, \rho = 0.1$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.48 (0.15)
				FPCA	0.51 (0.19)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.02)</b>
				PDA(scores)	0.07 (0.04)
				FPCA	0.06 (0.04)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.04$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.46 (0.09)
				FPCA	0.52 (0.17)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.10 (0.05)
				FPCA	0.06 (0.04)
$\Sigma_{\text{CS}}, \rho = 0.4$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.02)</b>
				PDA(scores)	0.47 (0.15)
				FPCA	0.51 (0.19)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.00 (0.01)</b>
				PDA(scores)	0.07 (0.04)
				FPCA	0.07 (0.04)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.05$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.50 (0.08)
				FPCA	0.52 (0.15)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.08 (0.04)
				FPCA	0.07 (0.04)
$\Sigma_{\text{AR}(1)}, \rho = 0.1$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.48 (0.15)
				FPCA	0.50 (0.07)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.07 (0.04)
				FPCA	0.07 (0.04)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.04$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.47 (0.12)
				FPCA	0.51 (0.19)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.00 (0.00)</b>
				PDA(scores)	0.10 (0.05)
				FPCA	0.06 (0.04)
$\Sigma_{\text{AR}(1)}, \rho = 0.4$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.02)</b>
				PDA(scores)	0.47 (0.15)
				FPCA	0.49 (0.07)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.02 (0.02)</b>
				PDA(scores)	0.06 (0.04)
				FPCA	0.07 (0.04)
	1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.06$	PDA(RSQ)	<b>0.02 (0.02)</b>
				PDA(scores)	0.48 (0.11)
				FPCA	0.50 (0.07)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.02 (0.02)</b>
				PDA(scores)	0.12 (0.05)
				FPCA	0.07 (0.04)
$\Sigma_{\text{AR}(1)}, \rho = 0.7$	0.1	$N(0, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.47 (0.15)
				FPCA	0.51 (0.19)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.01 (0.01)</b>
				PDA(scores)	0.06 (0.04)
				FPCA	0.07 (0.04)
	1	$N(0, 1)$	$0.02 \leq E_{\text{bayes}} \leq 0.07$	PDA(RSQ)	<b>0.04 (0.03)</b>
				PDA(scores)	0.49 (0.12)
				FPCA	0.52 (0.17)
		$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	<b>0.03 (0.02)</b>
				PDA(scores)	0.13 (0.04)
				FPCA	0.06 (0.04)

Table 4: Averaged error rate and its standard error in parenthesis for B-spline smoothed data of the simulation IV

	$b$	$a_{ij}$	Bayes Error	Method	CER
$\Sigma_{\text{ind}}$	0.1	$N(1, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	0.51 (0.08)
				PDA(scores)	<b>0.16 (0.05)</b>
				FPCA	<b>0.16 (0.05)</b>
		$N(2, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.21 (0.11)
				PDA(scores)	0.03 (0.02)
				FPCA	<b>0.02 (0.02)</b>
	1	$N(1, 1)$	$0.03 \leq E_{\text{bayes}} \leq 0.06$	PDA(RSQ)	0.56 (0.06)
				PDA(scores)	<b>0.16 (0.05)</b>
				FPCA	<b>0.16 (0.05)</b>
$\Sigma_{\text{CS}}, \rho = 0.1$	0.1	$N(1, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	0.33 (0.18)
				PDA(scores)	0.16 (0.05)
				FPCA	<b>0.15 (0.05)</b>
		$N(2, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.04 (0.04)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
	1	$N(1, 1)$	$0.04 \leq E_{\text{bayes}} \leq 0.08$	PDA(RSQ)	0.54 (0.07)
				PDA(scores)	0.16 (0.05)
				FPCA	<b>0.15 (0.05)</b>
$\Sigma_{\text{CS}}, \rho = 0.4$	0.1	$N(1, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	0.42 (0.16)
				PDA(scores)	0.16 (0.05)
				FPCA	<b>0.15 (0.05)</b>
		$N(2, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.13 (0.08)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
	1	$N(1, 1)$	$0.04 \leq E_{\text{bayes}} \leq 0.07$	PDA(RSQ)	0.53 (0.09)
				PDA(scores)	0.19 (0.06)
				FPCA	<b>0.17 (0.05)</b>
$\Sigma_{\text{AR}(1)}, \rho = 0.1$	0.1	$N(1, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.55 (0.07)
				PDA(scores)	0.03 (0.02)
				FPCA	<b>0.02 (0.02)</b>
		$N(2, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.02$	PDA(RSQ)	0.26 (0.19)
				PDA(scores)	0.16 (0.05)
				FPCA	<b>0.15 (0.05)</b>
	1	$N(1, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.04 (0.03)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
$\Sigma_{\text{AR}(1)}, \rho = 0.4$	0.1	$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.52 (0.06)
				PDA(scores)	0.15 (0.05)
				FPCA	<b>0.14 (0.05)</b>
		$N(2, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.55 (0.06)
				PDA(scores)	0.16 (0.05)
				FPCA	<b>0.15 (0.05)</b>
	1	$N(1, 1)$	$0.06 \leq E_{\text{bayes}} \leq 0.11$	PDA(RSQ)	0.49 (0.07)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
$\Sigma_{\text{AR}(1)}, \rho = 0.7$	0.1	$N(1, 1)$	$0.00 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.42 (0.13)
				PDA(scores)	<b>0.15 (0.05)</b>
				FPCA	<b>0.15 (0.05)</b>
		$N(2, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	<b>0.02 (0.02)</b>
				PDA(scores)	0.03 (0.02)
				FPCA	<b>0.02 (0.02)</b>
	1	$N(1, 1)$	$0.07 \leq E_{\text{bayes}} \leq 0.13$	PDA(RSQ)	0.47 (0.08)
				PDA(scores)	0.15 (0.05)
				FPCA	<b>0.13 (0.05)</b>
		$N(2, 1)$	$0.01 \leq E_{\text{bayes}} \leq 0.03$	PDA(RSQ)	0.53 (0.07)
				PDA(scores)	0.03 (0.02)
				FPCA	<b>0.02 (0.02)</b>

Table 5: Averaged error rate and its standard error in parenthesis for B-spline smoothed data of the simulation V

	$b$	$\mu$	Bayes Error	Method	CER
$\Sigma_{\text{ind}}$	0.1	1	$0.10 \leq E_{\text{bayes}} \leq 0.17$	PDA(RSQ)	0.54 (0.07)
				PDA(scores)	0.18 (0.06)
				FPCA	<b>0.17 (0.06)</b>
	2		$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.54 (0.07)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
	1	1	$0.10 \leq E_{\text{bayes}} \leq 0.18$	PDA(RSQ)	0.53 (0.07)
				PDA(scores)	0.19 (0.06)
				FPCA	<b>0.16 (0.06)</b>
$\Sigma_{\text{CS}}, \rho = 0.1$	0.1	1	$0.11 \leq E_{\text{bayes}} \leq 0.19$	PDA(RSQ)	0.55 (0.05)
				PDA(scores)	<b>0.18 (0.05)</b>
				FPCA	0.18 (0.06)
	2		$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.50 (0.09)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
	1	1	$0.10 \leq E_{\text{bayes}} \leq 0.17$	PDA(RSQ)	0.52 (0.07)
				PDA(scores)	0.18 (0.06)
				FPCA	<b>0.17 (0.06)</b>
$\Sigma_{\text{CS}}, \rho = 0.4$	0.1	1	$0.10 \leq E_{\text{bayes}} \leq 0.17$	PDA(RSQ)	0.54 (0.07)
				PDA(scores)	<b>0.18 (0.06)</b>
				FPCA	<b>0.18 (0.06)</b>
	2		$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.56 (0.06)
				PDA(scores)	0.02 (0.02)
				FPCA	<b>0.02 (0.02)</b>
	1	1	$0.11 \leq E_{\text{bayes}} \leq 0.20$	PDA(RSQ)	0.50 (0.08)
				PDA(scores)	0.19 (0.06)
				FPCA	<b>0.17 (0.06)</b>
$\Sigma_{\text{AR}(1)}, \rho = 0.1$	0.1	1	$0.13 \leq E_{\text{bayes}} \leq 0.23$	PDA(RSQ)	0.51 (0.07)
				PDA(scores)	0.19 (0.06)
				FPCA	<b>0.18 (0.05)</b>
	2		$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.55 (0.06)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
	1	1	$0.11 \leq E_{\text{bayes}} \leq 0.19$	PDA(RSQ)	0.50 (0.07)
				PDA(scores)	0.19 (0.06)
				FPCA	<b>0.18 (0.06)</b>
$\Sigma_{\text{AR}(1)}, \rho = 0.4$	0.1	1	$0.11 \leq E_{\text{bayes}} \leq 0.19$	PDA(RSQ)	0.56 (0.06)
				PDA(scores)	0.18 (0.06)
				FPCA	<b>0.18 (0.05)</b>
	2		$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.45 (0.07)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
	1	1	$0.11 \leq E_{\text{bayes}} \leq 0.20$	PDA(RSQ)	0.53 (0.07)
				PDA(scores)	0.19 (0.06)
				FPCA	<b>0.18 (0.06)</b>
$\Sigma_{\text{AR}(1)}, \rho = 0.7$	0.1	1	$0.10 \leq E_{\text{bayes}} \leq 0.18$	PDA(RSQ)	0.55 (0.06)
				PDA(scores)	<b>0.16 (0.06)</b>
				FPCA	0.17 (0.06)
	2		$0.01 \leq E_{\text{bayes}} \leq 0.01$	PDA(RSQ)	0.52 (0.08)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>
	1	1	$0.11 \leq E_{\text{bayes}} \leq 0.20$	PDA(RSQ)	0.50 (0.08)
				PDA(scores)	0.18 (0.06)
				FPCA	<b>0.17 (0.05)</b>
$\Sigma_{\text{AR}(1)}, \rho = 0.7$	0.1	1	$0.11 \leq E_{\text{bayes}} \leq 0.20$	PDA(RSQ)	0.48 (0.08)
				PDA(scores)	<b>0.02 (0.02)</b>
				FPCA	<b>0.02 (0.02)</b>

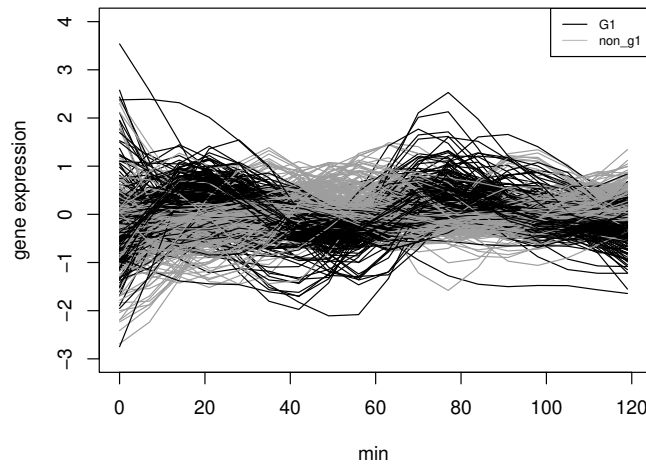


Figure 5: Temporal gene expression data of yeast cell cycle. Black lines indicate G1 phase, and gray lines indicate non-G1 phases smoothed by B-spline method.

Table 6: Averaged classification error rate and its standard error in parenthesis for the gene expression data

Smoothing method	Method	CER
Without smooth	PDA(RSQ)	0.42 (0.04)
	PDA(scores)	0.11 (0.03)
	FPCA (num of PCs=1)	0.21 (0.03)
	FPCA (num of PCs=2)	0.18 (0.02)
	FPCA (num of PCs=3)	<b>0.10 (0.02)</b>
	FPCA (num of PCs=4)	<b>0.10 (0.02)</b>
	FPCA (num of PCs=5)	0.11 (0.02)
B-spline basis smooth	PDA(RSQ)	0.36 (0.04)
	PDA(scores)	<b>0.10 (0.03)</b>
	FPCA (num of PCs=1)	0.24 (0.03)
	FPCA (num of PCs=2)	0.21 (0.03)
	FPCA (num of PCs=3)	0.11 (0.02)
	FPCA (num of PCs=4)	0.11 (0.02)
	FPCA (num of PCs=5)	0.11 (0.02)
Fourier basis smooth	PDA(RSQ)	0.36 (0.04)
	PDA(scores)	0.11 (0.02)
	FPCA (num of PCs=1)	0.17 (0.03)
	FPCA (num of PCs=2)	0.15 (0.03)
	FPCA (num of PCs=3)	<b>0.10 (0.02)</b>
	FPCA (num of PCs=4)	<b>0.10 (0.02)</b>
	FPCA (num of PCs=5)	<b>0.10 (0.02)</b>

However, RSQ based method improves the classification performance compare to the two score based methods in simulations I to III. The results from the score based methods are worse than the random-guessing. Figure 3 presents the scatter plots of the PDA scores,  $(s_{i1}, s_{i2})$ , for simulations I and V. We can clearly observe that the PDA scores are well-separated in simulation V when compared to simulation I.

For the selected order  $m$  in PDA, we plot the histogram of  $m$  for each simulation (Figure 4).

#### 4. Real data analysis

Leng and Müller (2006) classified yeast cell's genes with respect to its relation to cell cycle type using temporal gene expression data (<http://genome-www.stanford.edu/cellcycle/data/rawdata/>). Each gene expression is measured in intervals of 7 minutes from 0 to 119 minutes, and we obtain 18 time points

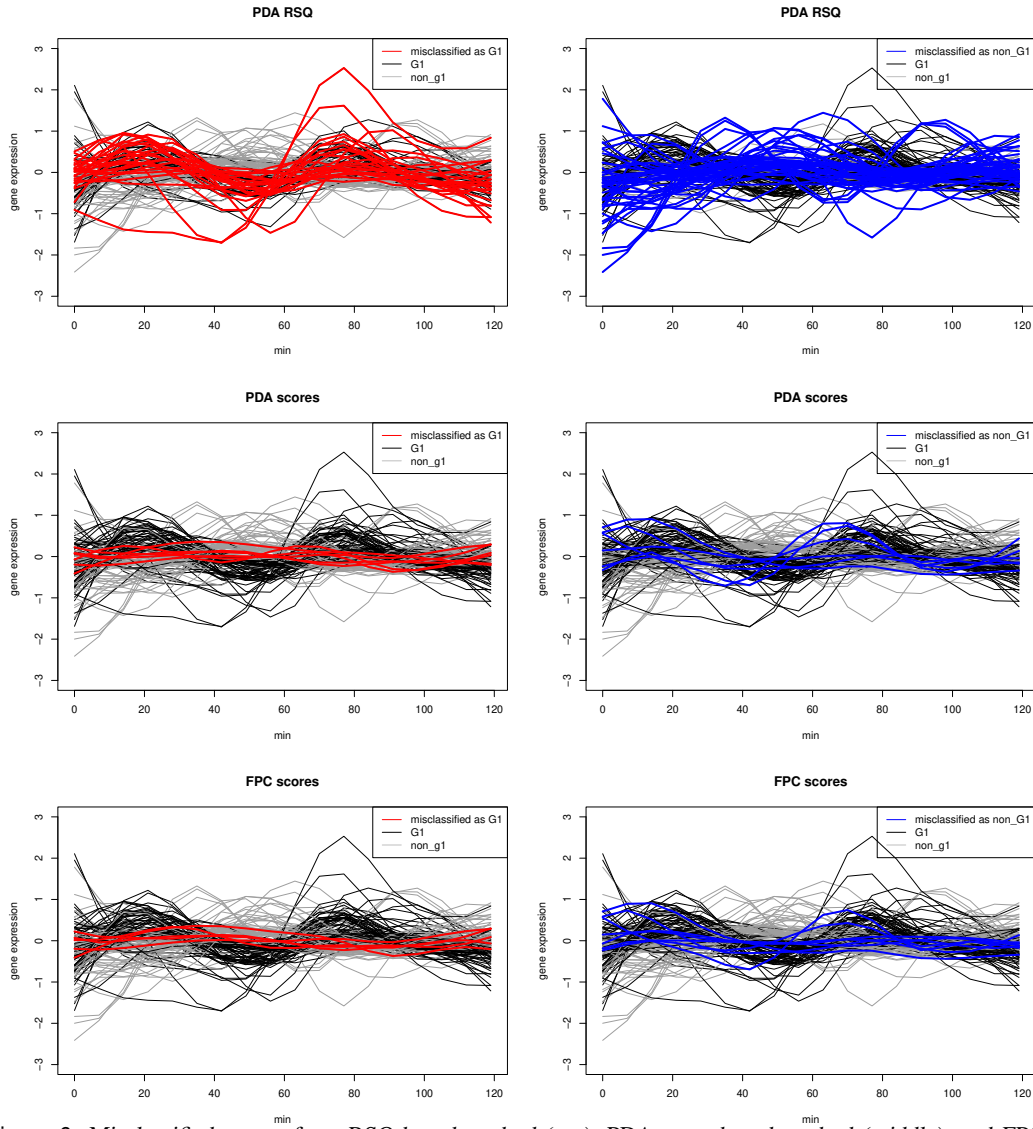


Figure 6: Misclassified curves from RSQ based method (top), PDA score based method (middle), and FPCA score based method (bottom).

for each gene. In this paper, we used 611 gene expression data for classification which the relation to cell cycle phases are known and all data points exist. Each gene expression data is labeled along with its relation to cell cycle phases. Analogous to Leng and Müller (2006), we grouped five classes (G1, S, S/G2, G2/M, and M/G1 phases) into two classes of G1 and non-G1 (S, S/G2, G2/M, and M/G1 phases). Figure 5 presents the 611 gene expression data smoothed by B-spline method. Each trajectory represents the expression of one specific gene.

For validation, we randomly divide the 611 curves into 489 training and 122 test curves, and we repeat this process 100 times with different splits. The classification methods based on PDA and FPCA

are applied on three cases: raw data, B-spline basis smoothed data, and Fourier basis smoothed data. The number of basis and roughness penalty parameter,  $\lambda$ , are determined by GCV. The optimal order of differential equation,  $m$ , in PDA is obtained from a 10-fold CV classification, and the number of PCs in FPCA are set to one to five.

Table 6 summarize the results. Without smoothing technique, FPCA with three PCs works best in classification two groups. Also, FPCA works well smoothed data using Fourier basis. PDA with RSQ works poorly for all cases, while PDA with scores work well for smoothed cases.

We check misclassified genes for each method in one test data set. The curves are smoothed by B-spline method. We can observe that the RSQ based method works poorly in classification two groups compare to the other two methods (Figure 6). For the PDA score based method, seven G1 genes misclassified as non-G1, and six non-G1 genes wrongly classified as G1. FPC score based method provide similar results.

As we have shown in the results of Simulation IV and V in the Section 3, RSQ gives similar values to the curves differ only by a constant or differ only by a phase shift. Figure 5 shows the two groups have similar amplitude and differ by a phase shift. Therefore, RSQ based classification do not work well, while two score based method classify the two groups well.

## 5. Concluding remarks

PDA is originally developed for the estimation of a differential operator from a functional data set. We focus on the role of PDA as a classification methodology. Here, we propose two principal differential analysis based classification methods. RSQ and score based classification methods are explained, and their performances are compared with the functional principal component score based classification method.

In all cases, there was no superior methodology. Each methodology had its pros and cons. PDA score based classification method works similarly to the FPCA score based classification method, while the RSQ based method shows a different tendency.

We expect that PDA based classification can be improved by considering principal differential equations, such as non-homogeneous cases. Also, choosing optimal order  $m$  in PDA is a critical problem in the method and are left for future study.

## References

- Cover T and Hart P (1967). Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, **13**, 21–27.
- Craven P and Wahba G (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation, *Numerische Mathematik*, **31**, 377–403.
- Dalla Rosa M, Sangalli LM, and Vantini S (2014). Principal differential analysis of the Aneurisk65 data set, *Advances in Data Analysis and Classification*, **8**, 287–302.
- Fukunaga K (2013). *Introduction to Statistical Pattern Recognition* (2nd ed), Elsevier, Boston.
- Grenander U (1950). Stochastic processes and statistical inference, *Arkiv Matematik*, **1**, 195–277.
- Jin S, Staniswalis JG, and Mallawaarachchi I (2013). Principal differential analysis with a continuous covariate: low-dimensional approximations for functional data, *Journal of Statistical Computation and Simulation*, **83**, 1964–1980.
- Kincaid C (2005). Guidelines for selecting the covariance structure in mixed model analysis. In *Proceedings of the Thirtieth Annual SAS Users Group International Conference* (Vol. 30, pp. 198–

- 130). SAS Institute Inc Cary NC.
- Kosarev EL and Pantos E (1983). Optimal smoothing of noisy data by fast Fourier transform, *Journal of Physics E: Scientific Instruments*, **16**, 537.
- Leng X and Müller HG (2006). Classification using functional data analysis for temporal gene expression data, *Bioinformatics*, **22**, 68–76.
- Müller HG (2005). Functional modelling and classification of longitudinal data, *Scandinavian Journal of Statistics*, **32**, 223–240.
- Nie Y, Wang L, Liu B, and Cao J (2018). Supervised functional principal component analysis, *Statistics and Computing*, **28**, 713–723.
- Ramsay JO (1982). When the data are functions, *Psychometrika*, **47**, 379–396.
- Ramsay JO (1996). Principal differential analysis: Data reduction by differential operators, *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 495–508.
- Ramsay JO and Dalzell CJ (1991). Some tools for functional data analysis, *Journal of the Royal Statistical Society: Series B (Methodological)*, **53**, 539–561.
- Ramsay JO and Silverman BW (2005). *Functional Data Analysis*, Springer, New York.
- Reimer M and Rudzicz F (2010). Identifying articulatory goals from kinematic data using principal differential analysis. In *Eleventh Annual Conference of the International Speech Communication Association*, Makuhari, Chiba, Japan September 26–30.
- Rice JA (2004). Functional and longitudinal data analysis: perspectives on smoothing, *Statistica Sinica*, **14**, 631–647.
- Sattar F and Rudzicz F (2016). Principal differential analysis for detection of bilabial closure gestures from articulatory data, *Computer Speech & Language*, **36**, 294–306.
- Staniswalis JG, Doodoo C, and Sharma A (2017). Local principal differential analysis: Graphical methods for functional data with covariates, *Communications in Statistics-Simulation and Computation*, **46**, 2346–2359.
- Tumer K and Ghosh J (2003). Bayes Error rate estimation using classifier ensembles, *Smart Engineering System Design*, **5**, 95–109.
- Wand MP (2000). A comparison of regression spline smoothing procedures, *Computational Statistics*, **15**, 443–462.

Received June 24, 2020; Revised September 17, 2020; Accepted January 4, 2021