Nonparametric Functional Data Analysis
Chapters 8
Functional Nonparametric Supervised Classification

Zhirong Yang

17 April 2007

Introduction

- Problem:
  - $Y_{\text{test}} \leftarrow \text{predict}(\boldsymbol{\chi}_{\text{test}}; \text{Model})$
  - $\text{Model} \leftarrow \text{train}(\{\boldsymbol{\chi}_i, Y_i\}_{i=1,\ldots,n})$
- Method

$$p_g(\chi) = P(Y = g | \boldsymbol{\chi} = \chi), \;\; g \in \bar{G} = \{1, \ldots, G\}$$

$$\hat{y}(\chi) = \arg\max_{g \in \bar{G}} \hat{p}_g(\chi)$$

$$\hat{p}_g(\chi) = \hat{p}_{g,h}(\chi) = \frac{\sum_{i=1}^{n} 1_{[Y_i = g]} K(h^{-1} d(\chi, \chi_i))}{\sum_{i=1}^{n} K(h^{-1} d(\chi, \chi_i))}$$

Method (cont.)

- Learning step

  for $h \in \mathcal{H}$

  for $i = 1, \ldots, n$

  for $g = 1, \ldots, G$

  $$\hat{p}_{g,h}(\chi_i) \leftarrow \frac{\sum_{\{i' : y_{i'} = g\}} K(h^{-1} d(\chi, \chi_i))}{\sum_{i=1}^{n} K(h^{-1} d(\chi, \chi_i))}$$

  enddo

  enddo

  enddo

  $$h_{\text{Loss}} \leftarrow \arg \inf_{h \in \mathcal{H}} \text{Loss}(h)$$

- Predicting

  $$\hat{y}(\chi) \leftarrow \arg\max_g \{\hat{p}_{g,h_{\text{Loss}}}(\chi)\}$$

Computational Issues

- kNN Estimator

$$\hat{p}_{g,k}(\mathbf{x}) = \frac{\sum_{\{i:y_i=g\}} K(h_k^{-1}\mathbf{d}(\mathbf{x},\mathbf{x}_i))}{\sum_{i=1}^{n} K(h_k^{-1}\mathbf{d}(\mathbf{x},\mathbf{x}_i))}$$

$$\text{card}\{i : \mathbf{d}(\mathbf{x}, \mathbf{x}_i) < h_k\} = k$$

- Selecting k

$$p_g^{\text{LCV}}(\mathbf{x}) = \frac{\sum_{\{i:y_i=g\}} K(h_{k_{\text{LCV}}}^{-1}\mathbf{d}(\mathbf{x},\mathbf{x}_i))}{\sum_{i=1}^{n} K(h_{k_{\text{LCV}}}^{-1}\mathbf{d}(\mathbf{x},\mathbf{x}_i))}$$

$$k_{\text{LCV}}(x_{i_0}) = \arg\min_k \text{LCV}(k, i_0)$$

$$\text{LCV}(k, i_0) = \sum_{g=1}^{G} \left(1_{[y_{i_0}=g]} - p_{g,k}^{(-i_0)}(\mathbf{x}_{i_0})\right)^2$$

$$p_{g,k}^{(-i_0)}(\mathbf{x}_{i_0}) = \frac{\sum_{i:y_i=g,i\neq i_0}^{n} K\left(\mathbf{d}(\mathbf{x}_i,\mathbf{x}_{i_0})/h_{k\left(\mathbf{x}_{i_0}\right)}\right)}{\sum_{i=1,i\neq i_0}^{n} K\left(\mathbf{d}(\mathbf{x}_i,\mathbf{x}_{i_0})/h_{k\left(\mathbf{x}_{i_0}\right)}\right)}$$

Performance Evaluation

for $i \in \{1, 2, \ldots, n\}$

$$y_i^{\mathrm{LCV}} \leftarrow \arg\max_{g \in \bar{G}} p_g^{\mathrm{LCV}}(\mathbf{x}_i)$$

enddo

$$\text{Misclassification} \leftarrow \frac{1}{n} \sum_{i=1}^{n} 1_{\left[y_0 \neq y_i^{\mathrm{LCV}}\right]}$$

Example: Speech Recognition

$n = 2000$ pairs $(\mathbf{x}_i, y_i)_{i=1,\ldots,n}$,

$\mathbf{x}_i = (\chi(f_1), \chi(f_2), \ldots, \chi(f_{150}))$ is the discretized log-periodogram.

$$y_i \in \{1, 2, 3, 4, 5\} \text{ with } \begin{cases} 1 \leftrightarrow \text{“sh”} \\ 2 \leftrightarrow \text{“iy”} \\ 3 \leftrightarrow \text{“dcl”} \\ 4 \leftrightarrow \text{“aa”} \\ 5 \leftrightarrow \text{“ao”} \end{cases}$$

Training by 5-fold cross-validation; Each fold has 50 pairs; Testing set has 250 pairs.

Speech Recognition: single run

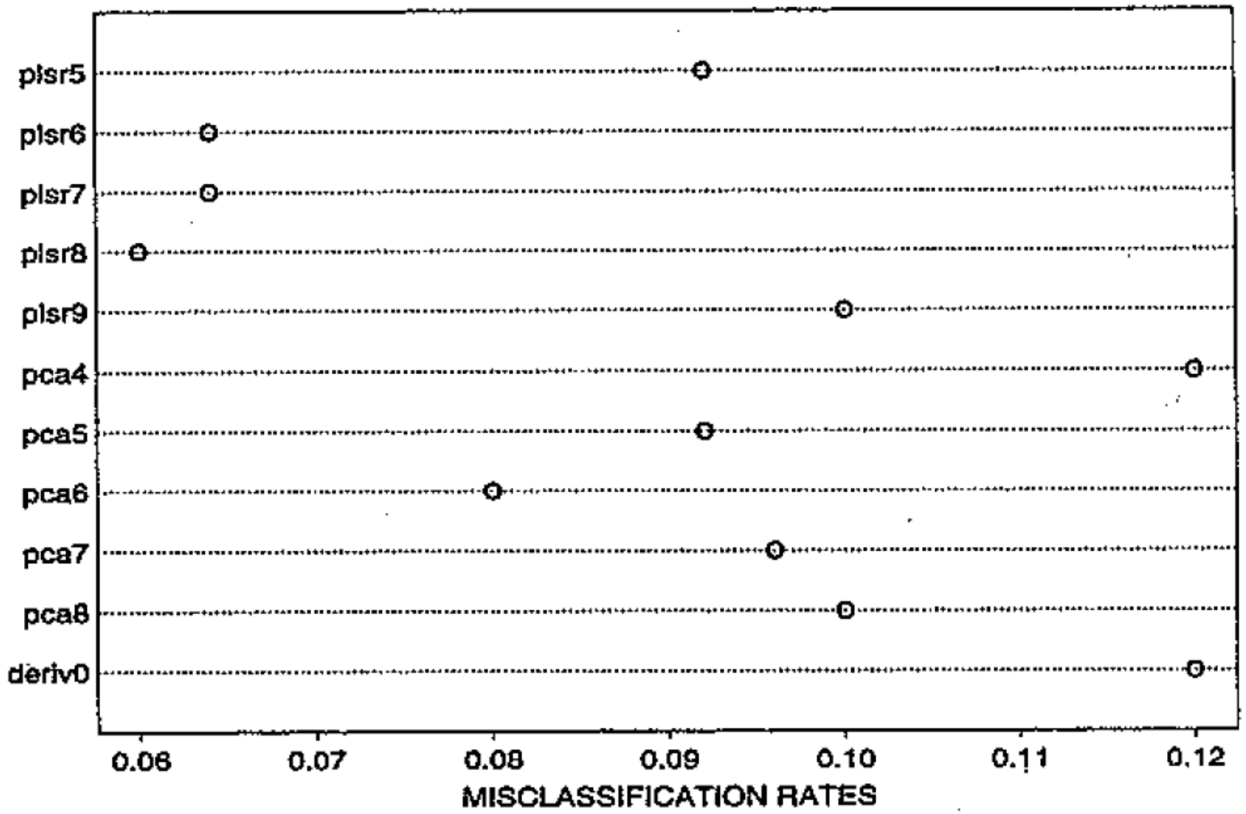

Fig. 8.1. Speech Recognition Data Discrimination: One Run

Speech Recognition: repeat 50 times

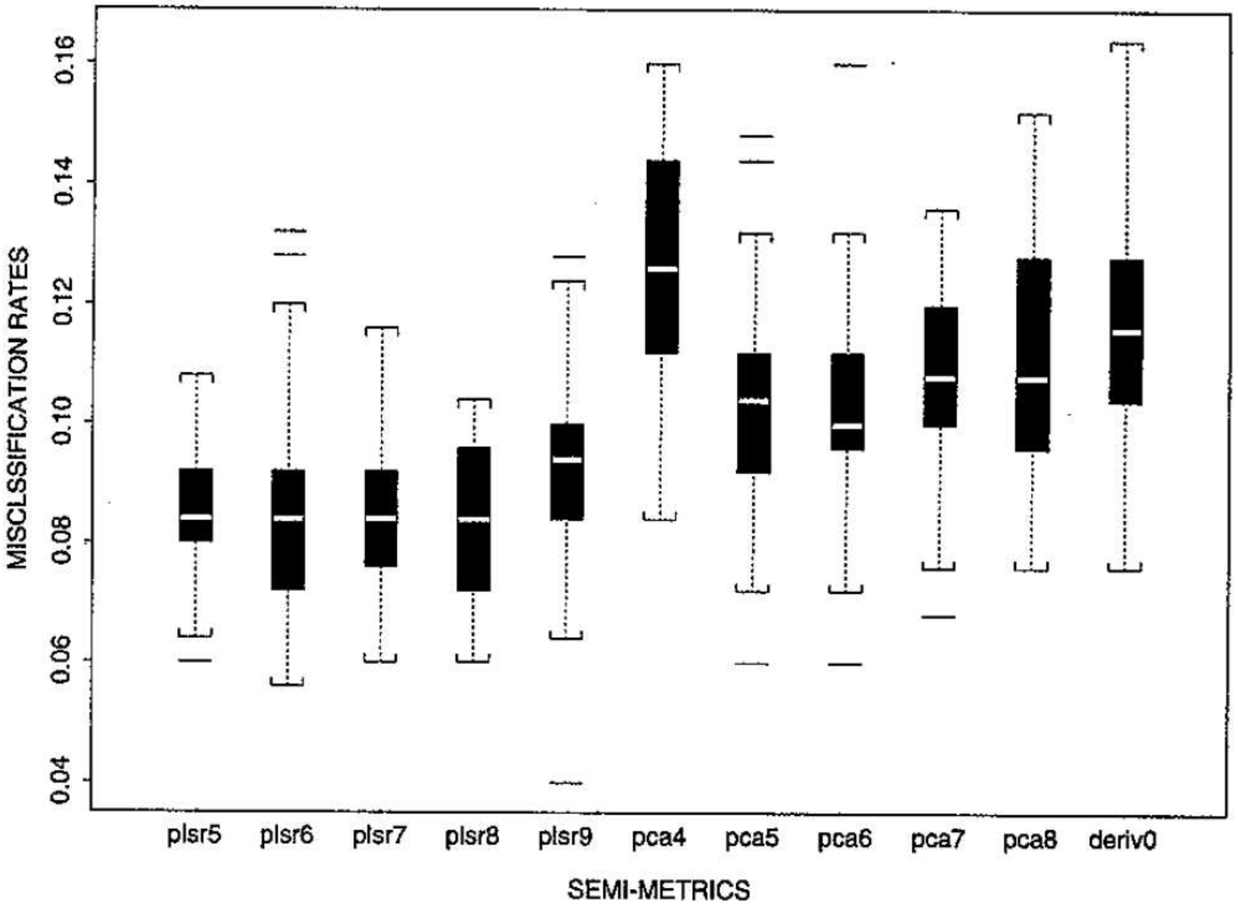

Fig. 8.2. Speech Recognition Data Discrimination: 50 runs