Supervised Classification for Functional Data by Estimating the Density of Multivariate Depth

Chong Ma – University of South Carolina
David B. Hitchcock – University of South Carolina

March 25, 2016

What is functional data?

- High-frequency measurements (depend on monitoring equipment).
  - hand-writing data.
  - spectral data.
- Smooth but complex processes.
- Repeated functional observations.
- Derivative information maybe useful.

Example for functional data

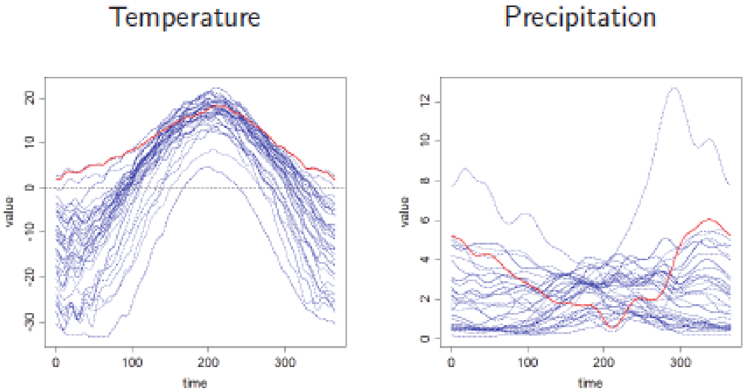

Figure: Average daily temperature and precipitation records in 35 weather stations across Canada (classical and often-used)

Example for functional data

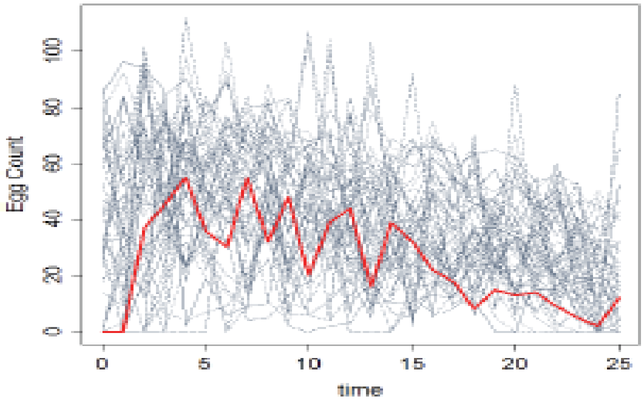

Figure: Records of number of eggs laid by Mediterranean Fruit Fly (Ceratitis capitata) in each of 25 days (courtesy of H.-G. Mller).

Goals in Functional Data Analysis

- Estimation of distribution of functional data.
- Prediction and classification of response variable (scalers, vectors or functions) related to functional data.
- Relationship between derivatives of functions

Supervised classification of functional data

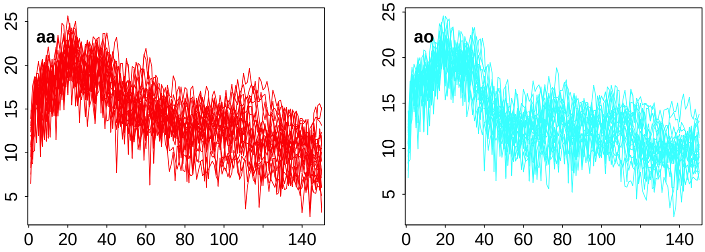

Figure: 20 records of two phonemes in log-periodogram at 150 frequency points.

Supervised Classification Format

- $\{\mathbf{x}_i(t), y_i\}$, $i = 1, \ldots, n$ and $y_i \in \mathcal{G} = \{1, \ldots, G\}$.
- Given a training set of functional data with known group memberships, predict the group membership for a new functional observation.
- Toolbox
  - Multivariate approaches: LDA, PDA, PCA, . . .
  - Functional approaches: k-nearest neighbors, depth-based, . . .

Depth Notion

- Mahalanobis depth (Mahalanobis, 1935; Liu and Singh, 1993)
  $MhD = \{1 + (\mathbf{x} - \mu)'\mathbf{\Sigma}^{-\mathbf{1}}(\mathbf{x} - \mu)\}^{-1}$
- Half-space depth (HD) (Tukey, 1975)
  $HD_F(\mathbf{x}) = \inf_H\{P_F(H) : \text{H is a closed half space in } \mathbb{R}^d, \mathbf{x} \in H\}$
- Simplicial depth (SD) (Liu, 1990)
  $SD(F, \mathbf{x}) = P_F(\mathbf{x} \in S(\mathbf{X_1}, \ldots, \mathbf{X_{d+1}}))$

Tukey's Half Space Depth

The *halfspace* depth of a point $\mathbf{p}$ with respect to $S \subset \mathbb{R}^d$ is defined as

$$depth_S(\mathbf{p}) = \min_{\mathbf{a} \in \mathbb{R}^d \setminus \{0\}} |\{\mathbf{q} \in S | \langle \mathbf{a}, \mathbf{q} \rangle > \langle \mathbf{a}, \mathbf{p} \rangle \}|$$

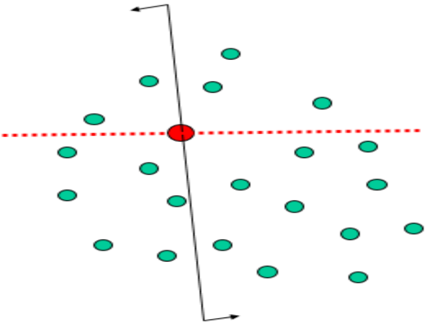

Figure: Tukey half space depth

Simplicial Depth

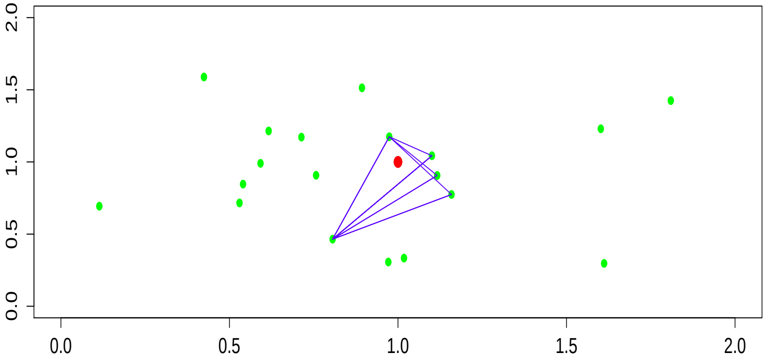

Figure: Simplicial Depth

Modified Band Depth

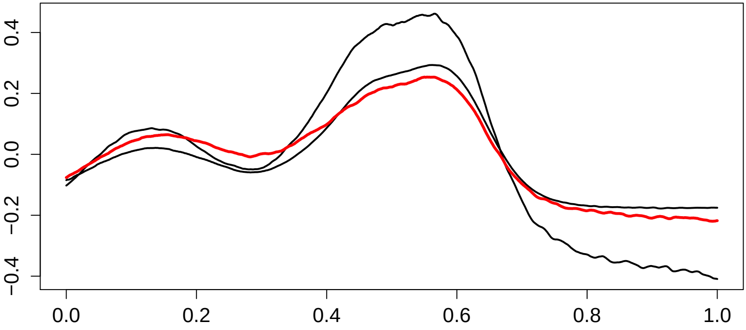

Figure: Modified Band Depth

Modified Band Depth

- Denote a functional observation $\{\mathbf{x}(t), t \in \mathcal{T}\}$ by $\mathbf{x}$, where $\mathcal{T}$ is compact. Without loss of generality, we set $\mathcal{T}$ as $[0, 1]$. Assume $\mathbf{x_1}, \ldots, \mathbf{x_n} \overset{i.i.d}{\sim} F_{\mathbf{X}}$, where $F_{\mathbf{X}}$ is the cumulative distribution of the stochastic process $\mathbf{X}$.
- The population version of modified band depth is

$$MBD_J(\mathbf{x}) = \sum_{j=2}^{J} MBD^{(j)}(\mathbf{x}) = \sum_{j=2}^{J} E[\lambda_r(A(\mathbf{x}; \mathbf{X}_1, \ldots, \mathbf{X}_j))]$$

where

$$A(\mathbf{x}; \mathbf{X}_1, \ldots, \mathbf{X}_j) = \{t \in \mathcal{T} : \min_{i=1,\ldots,j} \mathbf{X}_i(t) \leq \mathbf{x}(t) \leq \max_{i=1,\ldots,j} \mathbf{X}_i(t)\}$$

and $\lambda_r(A(\mathbf{x}; \mathbf{X}_1, \ldots, \mathbf{X}_j)) = \lambda(A(\mathbf{x}; \mathbf{X}_1, \ldots, \mathbf{X}_j))/\lambda(\mathcal{T})$

Modified Band Depth

- The sample version of $MBD^{(j)}(\mathbf{x})$ is

$$MBD_n^{(j)}(\mathbf{x}) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < \cdots < i_j \le n} \lambda_r(\{t \in \mathcal{T} : \min_{i=1,\ldots,j} \mathbf{x}_i(t) \le \mathbf{x}(t) \le \max_{i=1,\ldots,j} \mathbf{x}_i(t)\})$$

- $MBD_n^{(j)}(\mathbf{x})$ is consistent to $MBD^{(j)}(\mathbf{x})$.
- The computation load for $MBD_n^{(j)}(\mathbf{x})$ is $O(n^J)$.

Adjusted Modified Band Depth

- We propose an adjusted modified band depth for $\mathbf{x}$ ($AMBD(\mathbf{x})$) in order to make it more sense statistically. And we use bootstrapping method to estimate the population version of $AMBD(\mathbf{x})$.
  - $AMBD_J(\mathbf{x}) = MBD^J(\mathbf{x}) = E[\lambda_r(A(\mathbf{x}; \mathbf{X}_1, \ldots, \mathbf{X}_J))]$
  - $AMBD_n^J(\mathbf{x}) = \frac{1}{nb} \sum_{i=1}^{nb} \lambda_r(A(\mathbf{x}; \mathbf{x}_{i1}^*, \ldots, \mathbf{x}_{iJ}^*))$
- $\{\mathbf{x}_{i1}^*, \ldots, \mathbf{x}_{iJ}^*\}$ is a bootstrapping sample of size J from the sample of n functional data, $i = 1, \ldots, nb$. And we also have $AMBD_n^J(\mathbf{x}) \stackrel{a.s.}{=} AMBD^J(\mathbf{x})$ (see e.g. Cuevas, Febrero and Fraiman 2005).

Multivariate Depth

- The multivariate depth is formated for supervised classification problem. $\{\mathbf{x}_i(t), y_i\}$, $i = 1, \ldots, n$ and $y_i \in \mathcal{G} = \{1, \ldots, G\}$. Given a new functional observation $\mathbf{x}_0 \sim F_{\mathbf{X}_0}$, assume $F_{\mathbf{X}_0} \in \{F_{\mathbf{X}_1}, \ldots, F_{\mathbf{X}_G}\}$, the multivariate depth for $\mathbf{x}_0$ is defined by

$$MD(x_0) = (AMBD_{F_{\mathbf{X}_1}}(\mathbf{x}_0), \ldots, AMBD_{F_{\mathbf{X}_G}}(\mathbf{x}_0))'$$

  where
  $AMBD_{F_{\mathbf{X}_g}}(\mathbf{x}_0) = E[\lambda_r(A(\mathbf{x}_0; \mathbf{X}_1, \ldots, \mathbf{X}_J \overset{i.i.d}{\sim} F_{\mathbf{X}_g}))|\mathbf{X}_0 = \mathbf{x}_0]$
- $MD(\mathbf{x}_0)$ is a G-dimension "observation" from $MD(\mathbf{X}_0)$ which is a random vector related to $F_{\mathbf{X}_0}$.

Multivariate Depth

- We assume $MD(\mathbf{X}_0)$ follows a multivariate normal distribution, i.e., $MD(\mathbf{X}_0) \sim N_G(\mu_0, \mathbf{\Sigma}_0)$. It implies that $MD(\mathbf{X}_g) \sim N_G(\mu_g, \mathbf{\Sigma}_g)$ under the assumption of $\mathbf{X}_0 \sim F_{\mathbf{X}_0} \in \{F_{\mathbf{X}_1}, \ldots, F_{\mathbf{X}_G}\}$.
- Given G groups of observed functional data, $\{\mathbf{x}_1^g, \ldots, \mathbf{x}_{n_g}^g\}$,$g = 1, \ldots, G$, we are able to calculate the maximum likelihood estimators for μ_g and $\mathbf{\Sigma}_g$.

$$\hat{\mu}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} MD(\mathbf{x}_i^g) \tag{1}$$

$$\hat{\mathbf{\Sigma}}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} (MB(\mathbf{x}_i^g) - \hat{\mu}_g)(MB(\mathbf{x}_i^g) - \hat{\mu}_g)' \tag{2}$$

Forensic Casework

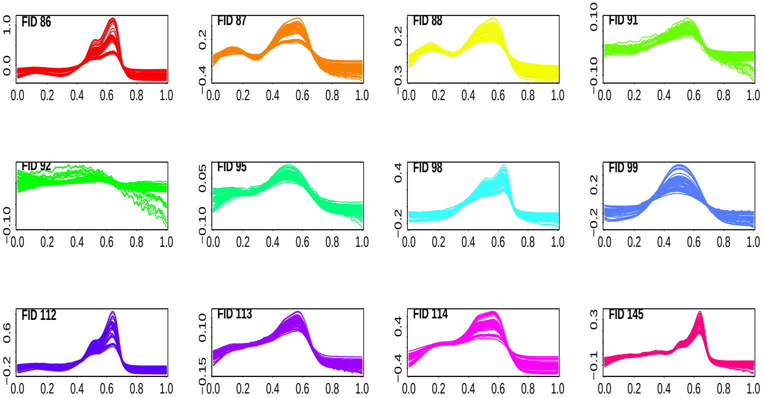

Figure: 12 blue acrylic textile fiber absorbance spectral data

Forensic Casework

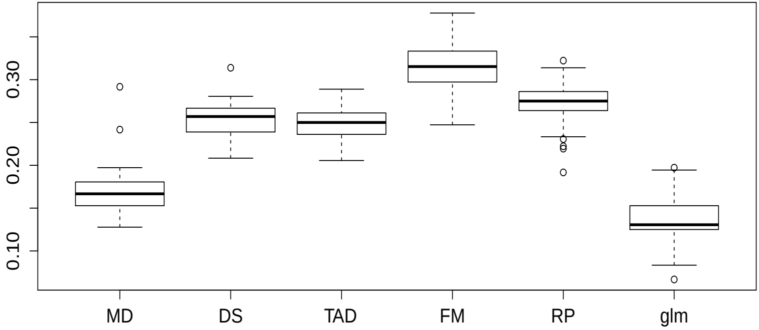

Figure: Misclassification rate for forensic fiber data.