Supervised Classification for Functional Data Using False Discovery Rate and Multivariate Functional Depth

Chong Ma[1] David B. Hitchcock[2]

[1]PhD Candidate
University of South Carolina
[2]Associate Professor
University of South Carolina

Joint Statistical Meetings, August 1, 2016

Outline

Forensic Casework

Subjects of Interets

- 12 blue acrylic fibers

Forensic Data

- **y**: Labels of the 12 blue acrylic fibers, i.e., $1, 2, \ldots, 12$.
- **x**: smoothed absorbance spectra. Each group (fiber) has 50 replicate measurements (smoothed curves) obtained from 5 laboratories.

Research Interest

- In a K-groups classification problem, given a new observed smoothed curve with unknown group membership, assume the prior probability of assigning it to each group is equally likely.
- Develop a probabilistic predictive classifier aiming to compute the posterior probabilities of assigning it to each group.

Forensic Casework

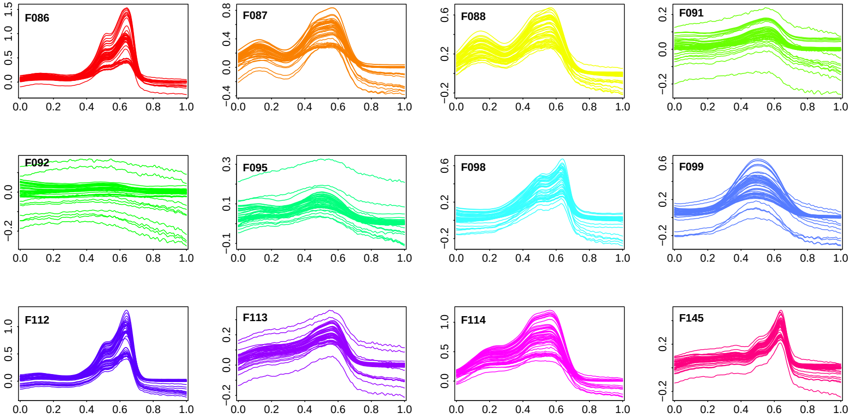

Figure 1: Smoothed curves of UV-visible absorbance spectra for 12 blue acrylic fibers. Each fiber has 50 replicate absorbance spectra obtained from 5 labs.

Outline

Statistical Data Depth

Statistical depth function serves as a tool providing a center-outward ordering of points in $\mathbb{R}^d$, i.e., $D(\cdot;\cdot) : \mathbb{R}^d \times \mathfrak{F} \to \mathbb{R}^+ \cup \{0\}$. It should ideally satisfy properties (Liu (1990), Zuo and Serfling (2000)):

1. Affine invariance.
2. Maximality at center.
3. Monotonicity relative to deepest point.
4. Vanishing at infinity.

Maximum Depth Classifier

Consider two groups of curves

$$x_1(t), \ldots, x_n(t) \overset{i.i.d}{\sim} \mathbb{F}_{X(t)}$$

and

$$y_1(t), \ldots, y_m(t) \overset{i.i.d}{\sim} \mathbb{F}_{Y(t)}$$

Given a new observed curve $z(t)$ with unknown class membership, the maximum depth classifier is to calculate the depth of $z(t)$ in each group separately and classify it to the group with the largest depth value.

$$I(z(t) \in \mathbb{F}_{X(t)}) = \begin{cases} 1, & \text{if } D(z(t); \mathbb{F}_{X(t)}) > D(z(t); \mathbb{F}_{Y(t)}) \\ 0, & \text{otherwise.} \end{cases}$$

Outline

Hypothesis test

Consider the depth ratios $T(z)$ for z in the above 0-1 classification problem,

$$\begin{aligned} T(z) &= \left(\frac{D(z;\mathbb{F}_X)}{D(z;\mathbb{F}_X)+D(z;\mathbb{F}_Y)}, \frac{D(z;\mathbb{F}_Y)}{D(z;\mathbb{F}_X)+D(z;\mathbb{F}_Y)}\right) \\ &= (T_X(z), T_Y(z)) \end{aligned}$$

and the hypothesis test,

$$\mathrm{H}_0 : z \sim \mathbb{F}_X \text{ vs. } \mathrm{H}_a : z \sim \mathbb{F}_Y$$

Under H_0, a set of significance regions are

$$\Gamma(t) = \{T_X(z) \leq t\}$$

Assume the prior probability of assigning z to each group is equally likely, i.e., $\pi_0 = \pi_a = 0.5$. The depth ratio for z w.r.t $\mathbb{F}_X$ is a test statistic, which is a scaler random variable related to z. The smaller the $T_X(z)$ is, the more evidence we reject it belonging to $\mathbb{F}_X$.

$$FDR(t) = P(\mathrm{H}_0|\Gamma(t)) = \frac{\pi_0 P(\Gamma(t)|\mathrm{H}_0)}{\pi_0 P(\Gamma(t)|\mathrm{H}_0) + \pi_a P(\Gamma(t)|\mathrm{H}_a)}$$

$$NPV(t) = P(\mathrm{H}_0|\Gamma(t)^c) = \frac{\pi_0 P(\Gamma(t)^c|\mathrm{H}_0)}{\pi_0 P(\Gamma(t)^c|\mathrm{H}_0) + \pi_a P(\Gamma(t)^a|\mathrm{H}_a)}$$

Remark: FDR is usually used in simultaneous hypothesis tests such as testing the significant genes from tens of thousands of gene expressions. In the classification scenario, we might consider both of FDR and NPV (Storey (2007),Storey (2002)).

Multi-group classification

Consider K groups of curves

$$x_1^1, \ldots, x_{n_1}^1 \overset{i.i.d}{\sim} \mathbb{F}_{X_1}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_1^K, \ldots, x_{n_K}^K \overset{i.i.d}{\sim} \mathbb{F}_{X_K}$$

The depth ratios $T(z)$ for z in the multi-groups is

$$T(z) = \left(\frac{D(z;\mathbb{F}_{X_1})}{D(z;\mathbb{F}_{X_1}) + \ldots + D(z;\mathbb{F}_{X_K})}, \ldots, \frac{D(z;\mathbb{F}_{X_K})}{D(z;\mathbb{F}_{X_1}) + \ldots + D(z;\mathbb{F}_{X_K})}\right)$$
$$= (T_{X_1}(z), \ldots, T_{X_K}(z))$$

The observed depth ratios for z is denoted by

$$\mathbf{t}(z) = (t_1, \ldots, t_K)$$

Multi-group classification

In the multi-group classification, we are more interested in the negative predictive value (NPV) for z as a "posterior probability" of assigning z to each group, given the prior probability of assigning z to each group is equally likely, i.e., $\pi_1 = \ldots = \pi_K = \frac{1}{K}$.
Consider the hypothesis test

$$\mathrm{H}_0^j : z \sim \mathbb{F}_{X_j} \text{ vs. } \mathrm{H}_a^j : z \nsim \mathbb{F}_{X_j}$$

The negative predictive value is

$$NPV_j(t) = P(\mathrm{H}_0^j | \Gamma(t)^c) = \frac{\pi_0 P(T_{X_j}(z) \geq t_j | \mathrm{H}_0^j)}{\pi_0 P(T_{X_j}(z) \geq t_j | \mathrm{H}_0^j) + \pi_a P(T_{X_j}(z) \geq t_j | \mathrm{H}_a^j)}$$

Remark: The probability distribution of $T_{X_j}(z)$ is unknown under H_0^j and H_a^j but can be estimated using the training data. Here $\pi_0 = \pi_j$ and $\pi_a = 1 - \pi_0$.

Multi-group classifier

By performing K hypothesis tests, we can get the "posterior probability" of assigning z to each group, i.e., the negative predictive value NPV_j. The predictive class membership for z is

$$\arg\max_j NPV_j = \arg\max_j P(\mathrm{H}_0^j | T_{X_j}(z) \geq t_j)$$

Remark: There are two main factors determining the performance of the negative predictive value (NPV) classifier. One is the choice of statistical depth function and the other one is the estimation of probability distribution of $T_{X_j}(z)$ under H_0^j and H_a^j.

Multivariate Functional Depth

We propose to augment the observed curve (after appropriate smoothing preprocessing), creating a set of p functions, by successively taking up to $p-1$ derivatives.

$$\mathbf{x} = \left(x^{(0)}, x^{(1)}, \ldots, x^{(p-1)}\right)$$

Data augmentation could obtain more powerful information regarding the depth calculation (Claeskens et al. (2014) *et al.*). In our work, we consider two types of multivariate functional depth, one is based on the integrated data depth (Fraiman and Muniz (2001), Cuevas et al. (2007) *et al.*), and the other is based on h-mode depth (Ferraty and Vieu (2004), Cuevas et al. (2006)*et al.*).

Multivariate Functional Depth

- **Multivariate functional integrated data depth**

$$D(\mathbf{X}; F_{\mathbf{Y}}) = \int_0^1 Z(t) \cdot w(t)\, dt$$

where

$$Z(t) = HD(\mathbf{X}(t); F_{\mathbf{Y}(t)}) = \inf_{\mathbf{u} \in \mathbb{R}^{p+1}, ||\mathbf{u}||=1} P(\mathbf{u}'\mathbf{Y}(t) \geq \mathbf{u}'\mathbf{X}(t)), \mathbf{X}(t) \in \mathbb{R}^{p+1}$$

- **Multivariate functional h-mode depth**

$$D(\mathbf{X}; F_{\mathbf{Y}}) = E_{\mathbf{Y}}[K_h(m(\mathbf{X}, \mathbf{Y}))]$$

where

$$m(\mathbf{X}, \mathbf{Y}) = \sqrt{||X^{(0)} - Y^{(0)}||^2 + ||X^{(1)} - Y^{(1)}||^2 + \cdots + ||X^{(p-1)} - Y^{(p-1)}||^2}$$

Outline

Simulation

main effects

$$x_1(t) = 0.4\phi(\frac{t-0.52}{0.125}) + 0.6\phi(\frac{t-0.75}{0.224}) + \epsilon(t) \quad (1)$$

$$x_2(t) = 0.4\phi(\frac{t-0.35}{0.141}) + 0.6\phi(\frac{t-0.73}{0.1}) + \epsilon(t) \quad (2)$$

$$x_3(t) = 300t^6(1-t)^2 + \epsilon(t) \quad (3)$$

batch effects

$$b_1(t) = \sin(t+U_{11})\log(t+U_{12}) \quad (4)$$

$$b_2(t) = -U_{21}t^2 + U_{22}t \quad (5)$$

$$b_3(t) = \phi(\frac{t-U_{31}}{0.316}) + U_{32} \quad (6)$$

Simulation

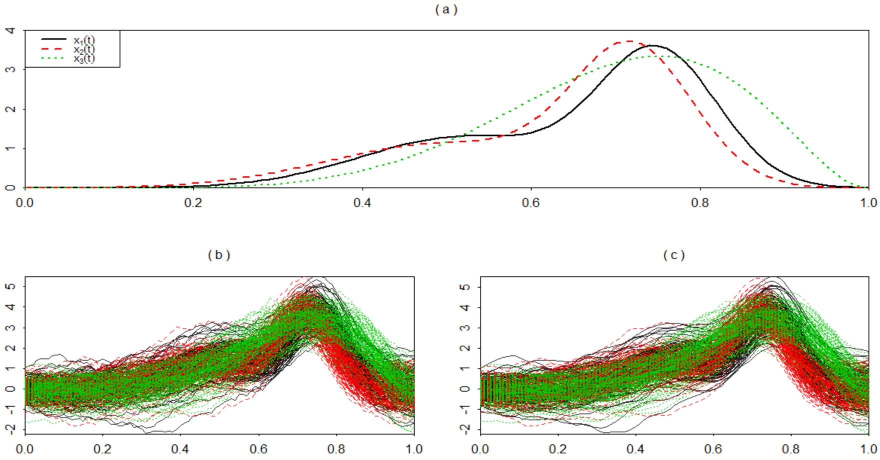

Figure 2: Simulation raw and smoothed curves for three groups

Simulation (FM depth)

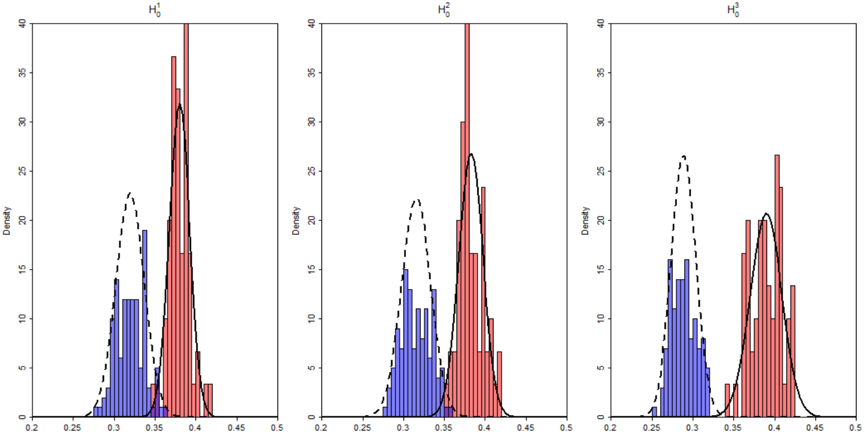

Figure 3: Distributions of depth ratios for the training curves under $H_0^j : z \sim \mathbb{F}_{X_j}$. The red stands for the true "negatives" and the blue for the true "discoveries".

Simulation (h-mode depth)

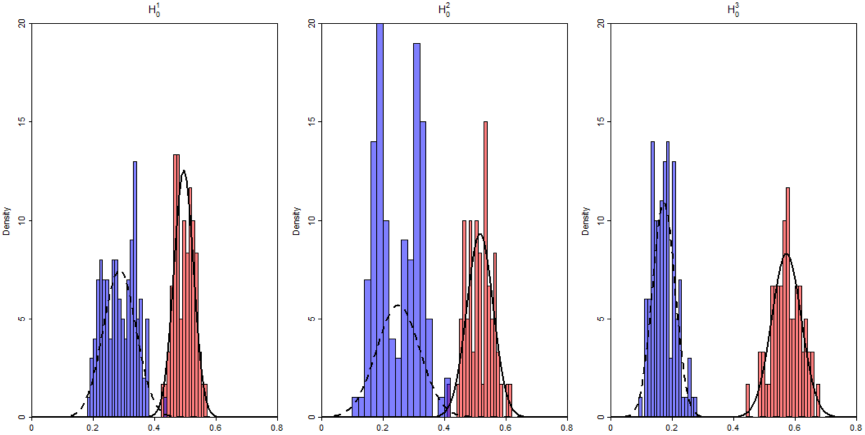

Figure 4: Distributions of depth ratios for the training curves under $H_0^j : z \sim \mathbb{F}_{X_j}$. The red stands for the true "negatives" and the blue for the true "discoveries".

Simulation ROC curves

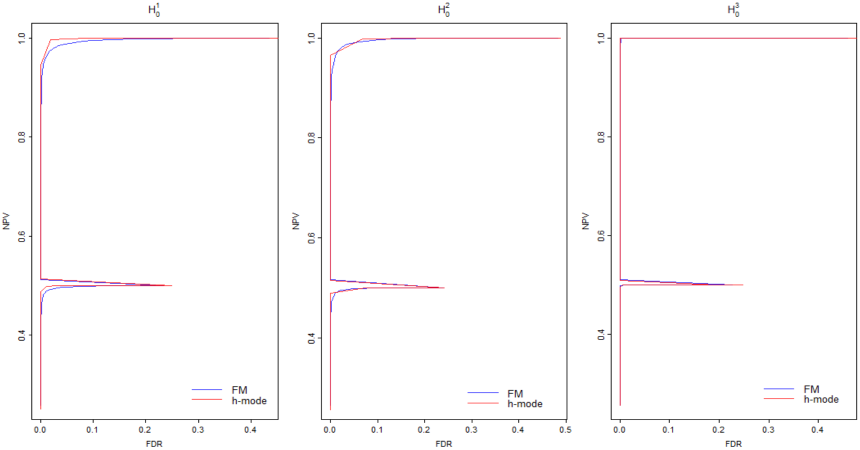

Figure 5: ROC curves under H_0^j $(j = 1, 2, 3)$ using FM depth and h-mode depth respectively.

Simulation Results

	maximum depth		normal	
m	FM	h-mode	FM	h-mode
1	7.11	0.16	6.73	0.16
	(3.41)	(0.25)	(3.69)	(0.25)
2	4.94	5.78	4.30	1.44
	(2.74)	(4.66)	(2.59)	(0.83)
3	5.29	60.2	4.23	48.5
	(2.78)	(4.97)	2.50)	(3.45)
4	5.48	67.4	4.39	67.7
	(2.78)	(2.54)	(2.60)	(3.19)

Table 1: Mean misclassification rate and standard deviation(in parenthesis) (in percent) obtained using maximum depth classifier and NPV based on Normal fitted model. *m* stands for the maximum order of derivatives in the augmented set of curves.

Outline

Distribution of FM depth ratio

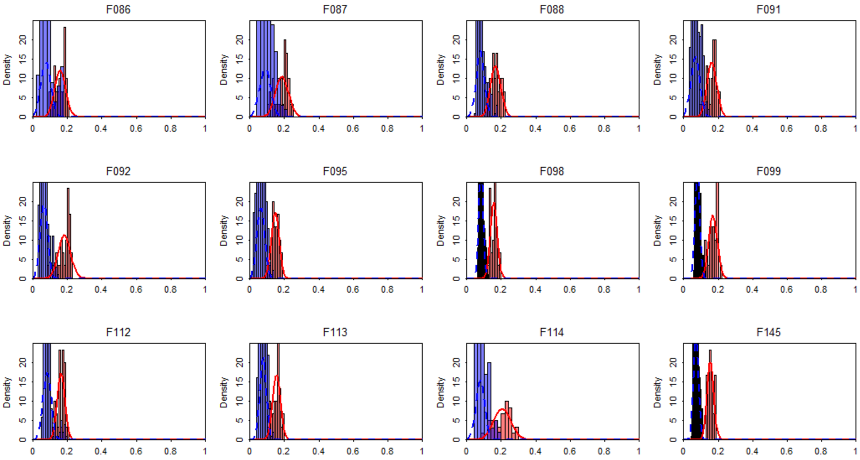

Figure 6: Distributions of FM depth ratios fitted by Normal distribution.

Distribution of h-mode depth ratio

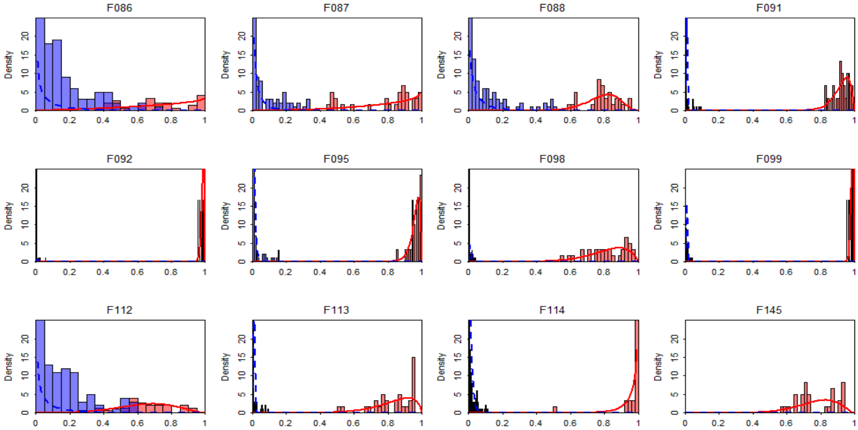

Figure 7: Distributions of h-mode depth ratios fitted by Beta distribution.

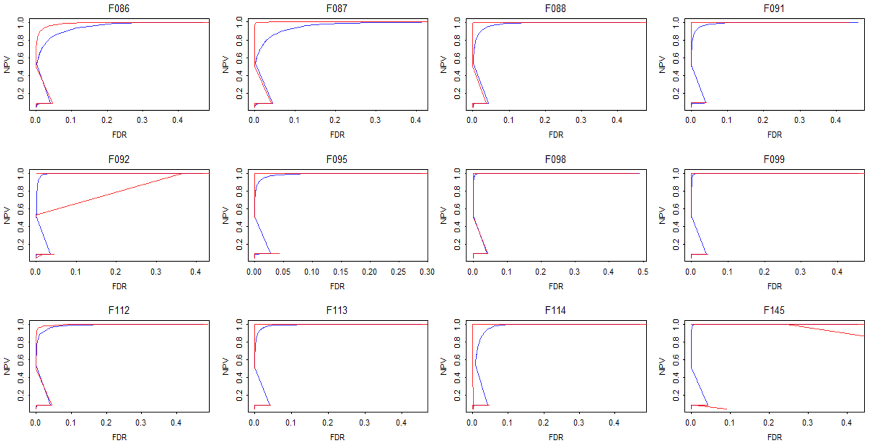

Figure 8: (red)ROC curve: h-mode depth fitted by Beta distribution; (blue)ROC curve: FM-depth fitted by Normal distribution.

Forensic Results

	Maximum Depth		Beta Model	
m	FM	h-mode	FM	h-mode
1	28.9	9.32	27.7	8.65
	(2.01)	(2.61)	(2.19)	(1.75)
2	24.2	24.9	23.3	21.4
	(2.29)	(3.99)	(2.47)	(3.42)
3	24.4	48.4	22.8	46.0
	(2.27)	(5.95)	2.26)	(4.49)
4	25.0	59.3	22.8	55.9
	(2.65)	(7.08)	(2.33)	(4.29)

Table 2: Mean misclassification rate and standard deviation(in parenthesis) (in percent) obtained using maximum depth classifier and NPV based on Normal fitted model. *m* stands for the maximum order of derivatives in the augmented set of curves.

Outline

Conclusion

1. We propose the NPV classifier based on the depth notion which obtains better consistency and efficiency than the maximum depth classifier.
2. The NPV classifier gives a more statistical interpretation in terms of the posterior probability of assigning a new curve to each group.
3. The NPV classifer has a potential ability to statistically compare the performance of different depth functions using ROC curves.
4. The performance of NPV classifier depends on the estimation of probability distribution of depth ratios under the H_0 and H_a(empirical distribution). A further study could be to consider a Bayesian model to fit the empirical distribution by capturing the variability with and between groups (Storey (2003)).

Outline

Reference I

Claeskens, G., Hubert, M., Slaets, L., and Vakili, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.*, 109(505):411–423.

Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Comput. Statist. Data Anal.*, 51(2):1063–1074.

Cuevas, A., Febrero, M., and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.*, 22(3):481–496.

Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *J. Nonparametr. Stat.*, 16(1-2):111–125. The International Conference on Recent Trends and Directions in Nonparametric Statistics.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer Science, New York.

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test*, 10(2):419–440.

Reference II

Ghosh, A. K. and Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scand. J. Statist.*, 32(2):327–350.

Hastie, T., Buja, A., and Tibshirani, R. (1995). Penalized discriminant analysis. *Ann. Statist.*, 23(1):73–102.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.*, 18(1):405–414.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer Science, New York.

Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):479–498.

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.*, 31(6):2013–2035.

Storey, J. D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(3):347–368.

Zuo, Y. and Serfling, R. (2000). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.*, 28(2):483–499.

The End

Thank you!