ORIE 779: Functional Data Analysis

From last meeting

Principal Component Analysis of Cornea Data

Try alternate “static view” (better projection?)

PC1: Overall curvature & “with the rule” astigmatism

PC2: Steeper superior vs. inferior & clear outlier effect

PC3: With the rule vs. against the rule astigmatism & outlier

From last meeting (cont.)

Clear problem with PCA:

Edge effects $\Rightarrow$ "outliers" $\Rightarrow$ "pulls off PC direction"

Outlier Deletion: still had problems with 10% of data deleted

Motivated alternate approach: Robust Statistical Methods

Main idea: downweight (instead of delete) outliers

Robust Functional Data Analysis

What is "multivariate median"?

There are *several*!   ("median" generalizes in different ways)

i. Coordinate-wise median $\begin{pmatrix} median(X_{i,1}) \\ \vdots \\ median(X_{i,d}) \end{pmatrix}$

- often worst
- not rotation invariant (2-d data uniform on "L")
- can lie on convex hull of data (same example)
- thus poor notion of "center"

Robust Functional Data Analysis (cont.)

What is “multivariate median” (cont.)?

ii. Simplicial depth (a. k. a. “data depth”): Liu, R. Y. (1990) “On a notion of data depth based on random simplices”, *Annals of Statistics*, 18, 405-414.

- “paint thickness” of $d+1$ dim “simplices” with corners at data
- Nice idea
- Good invariance properties
- slow to compute

Robust Functional Data Analysis (cont.)

What is “multivariate median” (cont.)?

iii. Huber’s L^p M-estimate:

Given data $X_1,...,X_n \in \Re^d$, estimate “center of population” by

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \left\| X_i - \theta \right\|_2^p$$

where $\left\| \cdot \right\|_2$ is the usual Euclidean norm.

Here: use only $p = 1$ (minimal impact by outliers)

Robust Functional Data Analysis (cont.)

L^1 M-estimate (cont.):

A view of minimizer: solution of

$$0 = \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \left\| X_i - \theta \right\|_2 = \sum_{i=1}^{n} \frac{X_i - \theta}{\left\| X_i - \theta \right\|_2}$$

A useful viewpoint is based on:

$P_{Sph(\theta,1)}$ = “Proj’n of data onto sphere cent’d at θ with radius 1”

And representation:

$$P_{Sph(\theta,1)} X_i = \theta + \frac{X_i - \theta}{\left\| X_i - \theta \right\|_2}$$

Robust Functional Data Analysis (cont.)

L^1 M-estimate (cont.):

Thus the solution of

$$0 = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\left\|X_i - \theta\right\|_2 = \sum_{i=1}^{n}\frac{X_i - \theta}{\left\|X_i - \theta\right\|_2}$$

is the solution of:

$$0 = avg\left\{P_{Sph(\theta,1)}X_i - \theta : i = 1,\ldots,n\right\}$$

So $\hat{\theta}$ is "location where projected data are centered"

[toy data illustration]

"slide sphere around until mean (of projected data) is at center"

Robust Functional Data Analysis (cont.)

L^1 M-estimate (cont.):

Additional literature:

Called "geometric median" (long before Huber) by: Haldane (1948) Note on the median of a multivariate distribution. *Biometrika*, 35, 414-415.

Shown unique for $d > 1$ by: Milasevic and Ducharme (1987) Uniqueness of the spatial median, *Annals of Statistics*, 15, 1332-1333.

Useful iterative algorithm: Gower (1974). The mediancentre. *Applied Statistics*, 23, 466-470 (see also Sec. 3.2 of Huber).

(Cornea Data experience: works well for $d = 66$)

Robust Functional Data Analysis (cont.)

L^1 M-estimation for Cornea Data

Sample Mean L^1 M-estimate

- Definite improvement
- But outliers still have some influence
- Improvement? (will suggest one soon)

Robust Functional Data Analysis (cont.)

Now have robust measure of “center”, how about “spread”?

I.e. how can we do robust PCA?

Recall:

- 2-d Toy Example
- Parabolas with 1 outlier

Robust Functional Data Analysis (cont.)

Approaches to Robust PCA:

1. Robust Estimation of Covariance Matrix

2. Projection Pursuit

3. Spherical PCA

Robust Functional Data Analysis (cont.)

Robust PCA 1: Robust Estimation of Covariance Matrix

A. Component-wise Robust Covariances:

- Major problem: hard to get non-negative definiteness

B. Minimum Volume Ellipsoid: Rousseeuw, and Leroy (1987) *Robust regression and outlier detection*, Wiley, New York.

- Requires $n > d$ (at least in available software)
- Needed for simple definition of “affine invariant”

Important Aside

Major difference between Functional Data Analysis

& Classical Multivariate Analysis

High Dimension, Low Sample Size Data

(sample size n < dimension d)

Classical Multivariate Analysis:

- start with "sphering data" (multiply by $\Sigma^{-1/2}$)
- but $\Sigma^{-1/2}$ doesn't for HDLSS data

Important Aside (cont.)

Classical Approach to HDLSS data:

"Don't have enough data for analysis, go get more"

Unworkable (and getting worse) for many modern settings:

- Medical Imaging (e.g. Cornea Data, $n = 43$ and $d = 66$)
- Micro-arrays & gene expression (e.g. $n = 86$ and $d = 400$)
- Chemometric spectra data (e.g. $n = 81$ and $d = 1500$)

$\vdots$

Robust Functional Data Analysis (cont.)

Robust PCA 2: Projection Pursuit

Idea: focus “finding direction of greatest variability”

Reference: Li, G. and Chen, Z. (1985) “Projection pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo”, *Journal of the American Statistical Association*, 80, 759-776.

Problems:

- Robust estimates of “spread” are nonlinear
- Results in many local optima
- Makes search problem very challenging
- Especially in very high dimensions
- Most examples have $d = 4, 5$
- Guoying Li: “I’ve heard of $d = 20$, but 60 seems too big”

Robust Functional Data Analysis (cont.)

Robust PCA 3: Spherical PCA

Idea: use “projection to sphere” idea from L^1 M-estimation

In particular *project data to centered sphere*

[toy conventional PCA] [toy spherical PCA]

- “hot dog” of data becomes “ice caps”
- easily found by PCA
- outliers “pulled in to reduce influence”
- radius of sphere unimportant

Robust Functional Data Analysis (cont.)

Spherical PCA:

Toy example: parabolas, with 1 outlier [recall raw data]

Recall conventional PCA:

- Outlier had small effect on mean
- Outlier had some effect on PC1 direction
- Outlier completely dominated PC2 direction
- Original (with no outlier) PC2 became PC3

Robust Functional Data Analysis (cont.)

Spherical PCA (cont.):

Toy example: parabolas, with 1 outlier (cont.)

Spherical PCA:

- Mean looks "smoother"
- PC1 nearly "flat" (unaffected by outlier)
- PC2 is nearly "tilt" (again unaffected by outlier)
- PC3 finally strongly driven by outlier
- OK, since all other directions "about equal in variation"

Robust Functional Data Analysis (cont.)

Spherical PCA (cont.):

Toy example: parabolas, with 1 outlier (cont.)

- Neither PC3 nor PC4 “caught all of the outlier”
- An effect of this down-weighting method
- PC R^2 lines for “sphered data”
- PC R^2 symbols show “SS for curves” (visual impression)
- Latter are not monotonic!
- Reflects “reduced influence” property of spherical PCA

Robust Functional Data Analysis (cont.)

Spherical PCA for Cornea Data:

Some improvement, but still had outlier influence

Reason: projection onto sphere "distorts the data"

Problem is visible in Parallel Coordinate Plot for Cornea Data

Top Plot: Zernike Coefficients

- All $n = 43$ very similar
- Most action in a few low frequencies

Robust Functional Data Analysis (cont.)

Parallel Coordinate Plot for Cornea Data (cont.)

Middle Plot: Zernike Coefficients – median

- Most Variation in “lowest frequencies”
- E.g. as in “Fourier compression of smooth signals”
- Projecting on sphere will destroy this
- By magnifying high frequency behavior

Bottom Plot: discussed later

Robust Functional Data Analysis (cont.)

Solution: "Elliptical Analysis"

Main idea: project data onto "suitable ellipse", not sphere

Which ellipse? (in general, this is problem that PCA solves!)

Simplification: Consider ellipses "parallel to coordinate axes"

Robust Functional Data Analysis (cont.)

Elliptical Analysis (cont.):

Simple Implentation: via coordinate axis rescaling

1. Divide each axis by MAD

2. Project Data to sphere (in transformed space)

3. Return to original space (mul'ply by orig'l MAD) for analysis

Where MAD = Median Absolute Deviation (from median)

$$= median\left|x_i - median(x_i)\right|$$

(simple, high breakdown, outlier resistant measure of "scale")

Robust Functional Data Analysis (cont.)

Elliptical Estimate of “center”:

Do L^1 M-estimation in transformed space (then transform back)

Results for cornea data:

Sample Mean Spherical Center Elliptical Center

- Elliptical clearly best
- Nearly no edge effect