ORIE 779: Functional Data Analysis

From last meeting

Robust Functional Data Analysis

Several “multivariate medians”

i. Coordinate-wise median
   - not rotation invariant (2-d data uniform on “L”)
   - can lie on convex hull of data (same example)
   - thus poor notion of “center”

ii. Simplicial depth (a. k. a. “data depth”)
   - slow to compute

From last meeting (cont.)

iii. Huber's L^p M-estimate:

[toy data illustration]

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

Approaches to Robust PCA:

1. Robust Estimation of Covariance Matrix
   Major problem 1: hard to get non-negative definiteness
   Major problem 2: High Dimension Low Sample Size

2. Projection Pursuit
   Major problem: slow to compute

From last meeting (cont.)

Approaches to Robust PCA (cont.):

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

Problem: "whitening" (i.e. magnificat'n of high frequency terms)

From last meeting (cont.)

Solution: "Elliptical Analysis"

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

Simple Implementation: 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)

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

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

Robust Functional Data Analysis (cont.)

Elliptical PCA: Do PCA on data projected to ellipse

Recall Toy Example, Parabolas with an outlier

[Conventional PCA] [Spherical PCA] [Elliptical PCA]

- Spherical & Elliptical are very nearly the same
- since scales nearly same in coordinate directions
- but a few small differences (study SS's)

Robust Functional Data Analysis (cont.)

Elliptical PCA for cornea data:

Original PC1, Elliptical PC1

- Still finds "overall curvature & correlated astigmatism"
- Minor edge effects almost completely gone

Original PC2, Elliptical PC2

- Huge edge effects dramatically reduced
- Still find "steeper superior vs. inferior"

Robust Functional Data Analysis (cont.)

Elliptical PCA for cornea data (cont.):

Original PC3, Elliptical PC3

- Edge effects greatly diminished
- But some of “against the rule astigmatism” also lost
- Price paid for robustness

Original PC4, Elliptical PC4

- Now looks more like variation on astigmatism???

Robust Functional Data Analysis (cont.)

Another population of corneas:

Kerataconus: “regions (cones) of very sharp curvature”

Raw Data (bright red cones)

Structure of the population?

Try PCA

Robust Functional Data Analysis (cont.)

Conventional PCA for Kerataconic Population

- Mean shows an “average cone”
- Mean affected by edge effects
- Major edge effects in all PC directions
- Thus strong need for robust PCA (e.g. elliptical)
- PC1 is “width of cone” (& outlier)
- PC2 is “horizontal location of cone”
- PC3 & PC4 also about location of cone

Robust Functional Data Analysis (cont.)

Elliptical PCA for Kerataconic Population:

- Edge Effects greatly diminished
- Mean similar, but narrower (in middle) & no edge effect
- PC1 is “width and strength of cone”
- PC2 is “narrow sharpness of cone”
- PC3 is “horizontal location”
- PC4 is outlier driven
- So elliptical PCA is not perfect

Robust Functional Data Analysis (cont.)

Disclaimer about robust analysis of Cornea Data:

Critical parameter is "radius of analysis", R_0

$R_0 = 4\ mm$: Shown above, Elliptical PCA very effective

$R_0 = 4.2\ mm$: Stronger edge effects, Elliptical PCA less useful

$R_0 = 3.5\ mm$: Edge effects weaker, don't need robust PCA

Robust Functional Data Analysis (cont.)

One more Toy Example: Same Parabolas, with 2 outliers

Raw Data

Outliers chosen to be "orthogonal"

- 1^{st} is "highest Fourier frequency"
- 2^{nd} outlier is (rescaled) "2^{nd} highest Fourier frequency"
- (will explain this later)

Robust Functional Data Analysis (cont.)

Toy Example: Parabolas, with 2 outliers (cont.)

Standard PCA:

- PC1 still "vertical shift, affected by 1st outlier"
- PC2 still 1st outlier
- Makes sense since "less variation" in 2nd outlier (than 1st)
- PC3 & PC4 a "muddling" of 2nd outlier and "slant"
- "Dir'n of greatest var'n", may not be "most useful dir'n"
- Also unclear desired dir'ns are orthogonal (but PCA is)

Robust Functional Data Analysis (cont.)

Toy Example: Parabolas, with 2 outliers (cont.)

Standard PCA (cont.): Some approaches:

- Interactive choice of some basis directions:
  - E.g. choose 3rd direction by 2nd outlier
  - Could also choose 1st dir'n to be exactly "vertical shift"
  - Generally useful, but will require "art" (i.e. expertise)

- VARIMAX rotation of subspaces
  - From Section 6.3.3 of Ramsey and Silverman (1997)
  - A good topic for student presentation

- Robust Spherical PCA

Robust Functional Data Analysis (cont.)

Toy Example: Parabolas, with 2 outliers (cont.)

Robust Spherical PCA

- Center: L1 M-est. avoids “sharp corners”
- PC1: same good job, with “vertical shift”
- PC2: still nicely shows “slant”
- Remaining PCs: Outliers now “separated”
- Again PC directions “not best”?
- Which PCs do explain outliers?

Robust Functional Data Analysis (cont.)

Toy Example: Parabolas, with 2 outliers (cont.)

Spherical PCA, PC5-9:

- Outliers spread over all remaining PCs
- So goal of “downweighting outliers” was achieved
- I.e. outliers have nearly no effect on PC directions
- Good? Or bad?
- Depends on context (outliers worthwhile or distracting?)

Deeper look at PCA

Now know usefulness of PCA, so let's "look under the hood"

Three important (and interesting) viewpoints:

1. Mathematics
2. Numerics
3. Statistics

First: Review linear algebra and multivariate probability

Linear Algebra Review

Vector Space:

- set of “vectors”, $\underset{\rightarrow}{x}$,
- and “scalars” (coefficients), a
- “closed” under “linear combination” ($\sum_i a_i \underset{\rightarrow}{x}_i$ in space)
- e.g. $\Re^d = \left\{ \underset{\rightarrow}{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} : x_1, \ldots, x_d \in \Re \right\}$, “$d$ dim Euclid’n space”

Linear Algebra Review, (cont.)

Subspace:

- subset that is again a vector space
- i.e. closed under linear combination
- e.g. lines through the origin
- e.g. planes through the origin
- e.g. subspace “generated by” a set of vectors
  (all linear combos of them =
  = containing hyperplane through origin)

Linear Algebra Review, (cont.)

Basis of subspace: set of vectors that:

- "span", i.e. everything is a linear combo of them

- are "linearly independent", i.e. linear combo is unique

- e.g. $\Re^d$ "unit vector basis" $\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right\}$

- e.g. $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_d \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$

Linear Algebra Review, (cont.)

Basis Matrix, of subspace of $\Re^d$

Given a basis, $\underrightarrow{v_1},...,\underrightarrow{v_n}$, create “matrix of columns”:

$$B = \left(\underrightarrow{v_1} \quad \cdots \quad \underrightarrow{v_n}\right) = \begin{pmatrix} v_{11} & & v_{n1} \\ \vdots & \cdots & \vdots \\ v_{1d} & & v_{nd} \end{pmatrix}_{d \times n}$$

Then “linear combo” is a matrix multiplication:

$$\sum_{i=1}^{n} a_i \underrightarrow{v_i} = B\underrightarrow{a} \quad \text{where} \quad \underrightarrow{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Check sizes: $d \times 1 = (d \times n) \leftrightarrow (n \times 1)$

Linear Algebra Review, (cont.)

Aside on matrix multiplication: (linear transformation)

For matrices $A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{k,1} & \cdots & a_{k,m} \end{pmatrix}, \quad B = \begin{pmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{pmatrix}$

Define the "matrix product" $AB = \begin{pmatrix} \sum_{i=1}^{m} a_{1,i} b_{i,1} & \cdots & \sum_{i=1}^{m} a_{1,i} b_{i,n} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} a_{k,i} b_{i,1} & \cdots & \sum_{i=1}^{m} a_{k,i} b_{i,n} \end{pmatrix}$

("inner products" of columns with rows)
(composition of linear transformations)

Often useful to check sizes: $k \times n = k \times m \leftrightarrow m \times n$

Linear Algebra Review, (cont.)

Matrix trace:

For a square matrix $A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{pmatrix}$,

Define $tr(A) = \sum_{i=1}^{m} a_{i,i}$

Trace commutes with matrix multiplication:

$$tr(AB) = tr(BA)$$

Linear Algebra Review, (cont.)

Dimension of subspace (a notion of "size"):

- number of elements in a basis (unique)
- $\dim\left(\Re^{d}\right) = d$ (use basis above)
- e.g. dim of a line is 1
- e.g. dim of a plane is 2
- dimension is "degrees of freedom"