# ORIE 779: Functional Data Analysis

From last meeting

High Dimension Low Sample Size Statistical Analysis

- Motivated by Corpora Callosa data
- General trend in FDA
- High Dimensional Space is Strange
- New Conceptual Model
- Orthogonal Subspace Projection
- Inscribed Sphere Example

# High Dimensional Space Is Strange (cont.)

Eric Friedman idea: consider other "distances":

$L^1$ - norm: $\left\| \underrightarrow{X} \right\|_1 = \sum_{i=1}^{d} \left| X_i \right|$

$L^\infty$ - norm: $\left\| X \right\|_\infty = \max_{i=1,\ldots,d} \left| X_i \right|$

Limiting behavior as $d \to \infty$?

# High Dimensional Space Is Strange (cont.)

Limiting distribution for $L^p$ - norm ( $p < \infty$ ):

Assume $X_i, i = 1,\ldots,d$ are i.i.d., with $E|X_i|^p < \infty$

$$\frac{1}{d}\sum_{i=1}^{d}|X_i|^p = E|X_i|^p + O_p\left(d^{-1/2}\right)$$

$$\|\underrightarrow{X}\|_p = \left(\sum_{i=1}^{d}|X_i|^p\right)^{1/p} = \left(d\left(E|X_i|^p + O_p\left(d^{-1/2}\right)\right)\right)^{1/p}$$

$$\|\underrightarrow{X}\|_p = d^{1/p}\|X_i\|_p + O_p\left(d^{1/p-1/2}\right)$$

where *both* discrete and probabilistic versions of $\|\cdot\|_p$ are used

# High Dimensional Space Is Strange (cont.)

Conclusion: Data lie

"near surface of $L^p$ unit ball, magnified by $d^{1/p} \|X_i\|_p$"

Paradox 1: True for all $p > 1$

- Yields conflicting rates of divergence as $d \to \infty$ ????

Paradox 2: Unit balls seem to have *very different* "shapes"

# High Dimensional Space Is Strange (cont.)

Shapes of $L^p$ unit balls: $B_p = \{\underset{\rightarrow}{x} : \|\underset{\rightarrow}{x}\| \leq 1\}$

$p = 2$: $\|\cdot\|_2$ is "Euclidean norm", so have $B_2$ is "usual sphere"

$p = 1$: $B_1 = \left\{ \underset{\rightarrow}{x} : \sum_{i=1}^{d} |x_i| < 1 \right\}$

- "boundary" is points where $\sum_{i-1}^{d} |x_i| = 1$
- for $x_i \geq 0, i = 1, \ldots, d$, boundary is the "Unit Simplex"

# High Dimensional Space Is Strange (cont.)

Shape of $L^1$ Unit Ball:

- Unit Simplex $\subset$ plane through the $\underset{\rightarrow}{e_i} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ ($1$ in $i$th place)

- Unit Simplex nearest to $\underset{\rightarrow}{0}$ at $\underset{\rightarrow}{n} = \begin{pmatrix} 1/d \\ \vdots \\ 1/d \end{pmatrix}$ (Lagrange Mult's)

# High Dimensional Space Is Strange (cont.)

Shape of $L^1$ Unit Ball (cont.):

- $\underset{\to}{n}$ is also normal vector of plane containing Unit Simplex

- Plane is $\left\{\underset{\to}{x} : \left\langle \underset{\to}{x}, \underset{\to}{n} \right\rangle = \left\langle \underset{\to}{n}, \underset{\to}{n} \right\rangle\right\}$

- $\left\langle \underset{\to}{n}, \underset{\to}{n} \right\rangle = \sum_{i=1}^{d} \frac{1}{d^2} = \frac{1}{d}$

- Check unit vectors lie on plane: $\left\langle \underset{\to}{e_i}, \underset{\to}{n} \right\rangle = \frac{1}{d}$

# High Dimensional Space Is Strange (cont.)

Shape of $L^1$ Unit Ball (cont.):

- Other "faces" have normal vectors $\begin{pmatrix} \pm 1/d \\ \vdots \\ \pm 1/d \end{pmatrix}$

- $L^1$ Unit Ball is "polytope with simplices as faces"

- $L^1$ Unit Ball is convex hull of $\{\underrightarrow{e_i}, -\underrightarrow{e_i} : i = 1, \ldots, d\}$

- Shape is *very different* from $L^2$ Unit Ball

# High Dimensional Space Is Strange (cont.)

Interesting questions:

- how can shapes be so different?

- yet both are "where most of data are"?

- May be resolvable via volumes (Lebesgue) of unit balls

  (volume of $L^{\infty}$ Unit Ball = $2^d$ (very "fat"!))
  (volume of $L^{1}$ Unit Ball = $4/d!$ (very "thin"!))

- And suitably normalizing (e.g. to volume 1)

- Main lessons:

  We *don't understand* high dimensional space

# Fisher Linear Discrimination (cont.)

Toy example with “expanding dimension” view:

- Population 1: $n = 20$, $N(-1,1)$
- Population 2: $n = 20$, $N(1,1)$
- Dimension $d = 1$: Everything happens on dashed line
- Substantial overlap of sub-populations
- Iterate: add a new dimension with indep. $N(0,0.6^2)$
- In all cases dashed line is “best discriminator”

# Fisher Linear Discrimination (cont.)

Toy example with "expanding dimension" view (cont.):

- Dimension $d = 2$: Scatterplot & FLD Normal Vector
- Dimension $d = 3$: Show 2-d Subspace (of $\Re^3$)

  Generated by dashed line & FLD

- Higher Dim: Show same 2-d subspace
- Shows how FLD normal vector "strays from optimal"
- An expected HDLSS effect
- But didn't see "perfect separation" as for Corp. Coll. FLD?

# Fisher Linear Discrimination (cont.)

Why no "perfect separation"?

Careful Investigation Showed:

- Used "Global Covariance Matrix", $\hat{\Sigma}$, in FLD calc'n
- Not "Within Class Covariance", $\hat{\Sigma}^w$, as intended
- Recall difference is in "centerpoint" (of Sum of Squares)
- Matters much???? Recall means very similar….
- Check by using $\hat{\Sigma}$ in Toy Example

# Fisher Linear Discrimination (cont.)

Same Toy example with Global Covariance $\hat{\Sigma}$ based FLD:

- Similar to correct $\hat{\Sigma}^w$ FLD, for lower dimensions
- But get "perfect alignment" for $d = 39$
- And "growing gap" for larger $d$
- Explains above observations
- But both versions still find "spurious directions"
- Could explain by relating FLD direction to:
  "direction that maximizes the t-statistic"???

# Revisit ICA (from discrimination viewpoint)

Recall find directions to:

- Maximize Independence
- Minimize Gaussianity

Recall:

Lecture 3/11/02

Lecture 3/25/02

Lecture 4/1/02

# Revisit ICA (cont.)

Slanted Clouds Toy Example

- Seek "direction" that separates red and blue projections
- PCA is poor (neither PC1, nor PC2 works)
- ICA is excellent (since "bimodal" = "most non-Gaussian")
- *No class information* used by ICA! (unlike FLD)
- Thus "useful preprocessing" for discrimination????
- Note non-orthogonal basis directions
- Which is "right", spherical or original scales????

# Revisit ICA (cont.)

Split X Discrimination Toy Example

- PCA leaves lots of overlap
- ICA gives excellent separation
- IC1 has “more kurtosis” (for total pop’n, not classes)
- but IC2 is best for discrimination
- Useful preprocessing for e.g. CART

# ICA for Corpora Collosa Data

Recall: Shapes of “window” between brain halves

Discrimination problem:

Schizophrenics [Shapes] vs. Controls [Shapes]

PCA: recall poor separation

# ICA for Corpora Collosa Data (cont.)

ICA Problem: HDLSS, $71 = n < d = 80$

Solution: Work only with 1st 20 Principal Components

(Reason for 20 discussed later?)

IC1: Seems to find an "outlier"

- Outlier is Case 50 [Numbered Population]

# ICA for Corpora Collosa Data (cont.)

Similarly for other ICs:

| IC | Outlier Case |
|---|---|
| 2 | 60 |
| 3 | 2 (“biggest outlier” only 3rd IC? By sphering?) |
| 4 | 26 |

Reason:

Outlier distributions have high kurtosis, thus found by ICA

# ICA for Corpora Collosa Data (cont.)

Solutions to “ICA driven by outliers” problem?

Sol’n 1: Reduce to only 1st 4 PCs: IC1 IC2 IC3 IC4

- Same as 4 - PC subspace above, not good discrimination

Sol’n 2: Use PCA “starting values” IC1 IC2 IC3 IC4

- Recall worked well for Parabs Up - Dn
- found some different outliers – Cases 2, 30, *, 22
- found a “bimodal direction”
- but weak discrimination???

# ICA for Corpora Collosa Data (cont.)

Sol'n 3: Project Data to surface of sphere IC1 IC2 IC3 IC4

- Recall Toy Example Motivation
- Still outlier driven (same outliers)
- But outliers "not so far out"
- Still doesn't separate Schizophrenics and Controls

Sol'n 4: Projection to sphere and PCA start

- Similar lessons (2 old, 2 new outliers)

# ICA for Corpora Collosa Data (cont.)

Sol'n 5: Different "ICA non-linearities",

5a: tanh, random start: IC1 IC2 IC3 IC4

- most are outlier (same as usual) driven
- IC2 maybe a little more interesting

5b: tanh, PC start: IC1 IC2 IC3 IC4

- all outlier driven

# ICA for Corpora Collosa Data (cont.)

5c: gauss, random start: IC1 IC2 IC3 IC4

- same lessons (even same outliers) as above

5d: gauss, PC start: IC1 IC2 IC3 IC4

- same as above

# ICA for Corpora Collosa Data (cont.)

Idea for improvement: find “directions to minimize kurtosis”

(not absolute value of kurtosis)

Implementation (short of recoding ICA):

1. Look in all 20 ICA directions (for some choice of opt’s)
2. Compute kurtosis for each
3. Sort in increasing kurtosis order

# ICA for Corpora Collosa Data (cont.)

Attempts:

a. Absolute Kurtosis, random start [combined graphic]:

- all kurtoses > 0, found *no* “useful directions”

b. Absolute Kurtosis, PC start [combined graphic]:

- found a bimodal direction (discovered earlier)
- and a $2^{nd}$ direction with kurtosis < 0
- “Start” is still an important issue

# ICA for Corpora Collosa Data (cont.)

c. Tanh, random start [combined graphic]:

- found 4 directions with kurtosis < 0
- none give "magic bullet" discrimination
- maybe "4 together" (e.g. input to CART) can do well?

d. Tanh, PC start [combined graphic]:

- OK, but not so good as (c)

# ICA for Corpora Collosa Data (cont.)

e. Gaus, random start [combined graphic]:

   - similar to above

f. Gaus, PC start [combined graphic]:

   - again 4 directions with strongly negative kurtosis
   - quite different directions from those in (c)?

# ICA for Corpora Collosa Data (cont.)

Another variation: Replace “sequential direction finding”

By “simultaneous maximization”

Results not so different from before, but:

- Best 1 dir’n separation, may be Gaus

show CorpColl\CCFicaSCs3allv35.ps & CCFicaSCs3allv36.ps

- Found only 1 dir’n with kurtosis < 0

- Thus not same as minimizing kurtosis

- Gaus directions nearly independent of start

Flip back and forth between CorpColl\CCFicaSCs3allv35.ps & CCFicaSCs3allv36.ps

# ICA for Corpora Collosa Data (cont.)

- Tanh directions similar

Flip back and forth between CorpColl\CCFicaSCs3allv33.ps & CCFicaSCs3allv34.ps

- Abs. Kurt. did not converge
- Oscillated between local solutions?
- Tried reducing 20-d eigenspace to 15,12,10
- Finally got convergence using 5-dim eigenspace

show CorpColl\CCFicaSCs3allv37.5d.ps & CCFicaSCs3allv38.5d.ps

  - 1st 2 directions look good for discrimination
  - Not dependent on starting value

flip back and forth

# ICA for Corpora Collosa Data (cont.)

- Found 8-d converged, but 9-d didn't

show CorpColl\CCFicaSCs3allv37.ps & CCFicaSCs3allv38.ps

  - Found 3 (out of 8) directions with Kurtosis < 0
  - Doesn't look so good for discrimination
  - Independent of starting value

- Get better results from more eigen-space reduction???

# ICA for Corpora Collosa Data (cont.)

Some conclusions and ideas:

i. Starting point is critical (and poorly understood)

ii. Should try “global optimization”, vs. “sequential”

iii. Seems a promising direction

iv. Will use these directions later (with SVM)

v. Would like to try explicitly minimizing kurtosis

To Do:

Connect up OSP with CV

Not yet used material:

After discrimination:

CC Sequential ICA vs. Simultaneous ICA   3-22-01:  3-5

ICA and Projection Pursuit    3-22-01:  6-13

Poly embedding:  4-19-01:   pg. 2-13

Kernel machines:  4-19-01:  pg. 14-22

SVM:   4-19-01:   pg.  23-27,    5-2-01:  pg. 2-20

Validation of discrimination: 5-2-01: pg. 21-32

Somewhere mention other approaches: neural nets, again reference

Duda, R. O., Hart, P. E. and Stork, D. G. (2001) *Pattern Classification*, Wiley.

To do later (???):

1. PCA time series – chemometrics data
2. ICA in discrimination
3. In vector space, orthogonal basis introduction
4. Fourier basis 3-22-01
5. Legendre basis
6. Tensor product Fourier Legendre basis

7. Zernike basis
8. Revisit cornea data?   (compare “raw image” with “fit images”, fiddle with Cornean power map? (do this at home?), use Figure from LMTZ paper, see directories D:\DellInspiron7000\SW30\Docs\Steve and D:\DellInspiron7000\SW30\Pictures)
9. Elliptical Fourier bases  4-05-01
10. Complex plane representation (no simple real valued basis)
11. Corpora Collosa Approximation
12. Support Vector Machines
13. Polynomial Embedding
14. Micro-Array Data analysis
15. Normal KerCli discrimination (in Cornean/demo)