From Last Meeting

Studying Independent Component Analysis (ICA)

References:

Hyvärinen and Oja (1999) Independent Component Analysis: A Tutorial, http://www.cis.hut.fi/projects/ica

Lee, T. W. (1998) Independent Component Analysis: Theory and Applications, Kluwer.

ICA for Corpora Collosa Data (cont.)

Last time got interesting results from:

“lazy man’s attempt at minimizing kurtosis”:

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

Used A. Kurt., Tanh, Gaus, random start and PC start

ICA for Corpora Collosa Data (cont.)

Untried 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 at most 2 dir’ns 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 did depend on start

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

  - 1^{st} 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 and Projection Pursuit

Question of Jerry Friedman, Stanford Univ.
(projection pursuit, CART, MARS, …)

Is ICA "well defined"?

Viewpoint: Projection Pursuit Density estimation

Model: "joint density function", $f(\underset{\rightarrow}{x}) = C \prod_{j=1}^{k} f_j\left(\underset{\rightarrow}{a}_j^t \underset{\rightarrow}{x}\right)$

For some "projection directions" $\underset{\rightarrow}{a}_j$

And some "marginal univariate densities" f_j

ICA and Projection Pursuit (cont.)

Interesting Properties:

Can have $k > d$:

- then all of the $\underrightarrow{a}_j$ can be viewed as “independent comp’s”
- clearly not all orthogonal
- so why should ICA algorithm restrict to orthogonal dir’ns?
- for k large enough, can approx. any f (tomography)
- smaller k is more interesting

ICA and Projection Pursuit (cont.)

Friedman's Projection Pursuit Algorithm

Step 1: Find $\underset{\rightarrow}{a}_1$ - "direction of maximal nonGaussianity"

Step 2: Transform data to Gaussianity *in that direction only*

Step 3: Iterate until "multivariate fit is good"

Useful for FDA?

ICA and Projection Pursuit (cont.)

Example: Directions: $\underset{\to}{a}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \underset{\to}{a}_2 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}, \quad \underset{\to}{a}_3 = \begin{pmatrix} \sqrt{3}/2 \\ -1/2 \end{pmatrix}$

(point in directions 120° apart)

and marginal densities $f_1, \ f_2, \ f_3$ Uniform(1/2,1)

Then $f(\underset{\to}{x}) = C \prod_{j=1}^{k} f_j\left(\underset{\to}{a}_j^t \underset{\to}{x}\right)$ is "uniform on an equilateral triangle"

- What does ICA find? (not well defined?)
- Projection Pursuit conveniently summarizes this dist'n
- Useful for FDA??? What do these directions tell us???

ICA and Projection Pursuit (cont.)

When is ICA well defined?

(Assume $E\underset{\rightarrow}{X} = \underset{\rightarrow}{0}$)

Sufficient Condition: $\exists$ matrix B,

so that $\underset{\rightarrow}{Y} = B^t \underset{\rightarrow}{X}$ is uncorrelated (i. e. $E\underset{\rightarrow}{Y}\underset{\rightarrow}{Y}^t = I_{d \times d}$)

- Recall: Independence $\Rightarrow$ uncorrelated
- So standard ICA independence assumption is sufficient
- Also get here by (nondegenerate) “sphering”

ICA and Projection Pursuit (cont.)

Under this assumption:

For which $\underline{a}_1$ and $\underline{a}_2$ is $\operatorname{cov}(\underline{a}_1^t \underline{Y}, \underline{a}_2^t \underline{Y}) = 0$?

$$0 = \operatorname{cov}(\underline{a}_1^t \underline{Y}, \underline{a}_2^t \underline{Y}) = E\left[\left(\underline{a}_1^t \underline{Y}\right)\left(\underline{a}_2^t \underline{Y}\right)^t\right] =$$

$$0 = E\left[\underline{a}_1^t \underline{Y}\,\underline{Y}^t \underline{a}_2\right] = \underline{a}_1^t E\left(\underline{Y}\,\underline{Y}^t\right)\underline{a}_2 = \underline{a}_1^t \underline{a}_2$$

Thus only uncorrelated when $\underline{a}_1$ and $\underline{a}_2$ are orthogonal

So enough to look for “directions of indep.” among ortho’l vectors

ICA and Projection Pursuit (cont.)

Is ICA (especially search over ortho'l dir'ns) well-defined?

- Yes, under ICA assumptions

- No, in general ??? (equilateral triangle example)

Conclusions:

- If there is a "translation to indep.", then ICA can find it

- If not, then ICA "restriction to orthogonal direction" can miss important structure that projection pursuit can find

Fun new data analysis:

From National Center for Atmospheric Research (last week)

Data from Enrica Bellone (NCAR)

- “Mass Flux” for quantifying “cloud types”

- Tried Standard PCA analysis

Show MassFlux\MassFlux1d1p1.ps

Mass Flux PCA

Mean: Captures "general shape"

PC1: Finds "overall height of peak"

- note 3 clusters in projections. "really there"?

PC2: Location of peak (2nd col. very useful here)

PC3: Describes "side lobes"?

Investigation of PC1 Clusters:

Main Question: “Important structure” or “sampling variability”?

Approach: SiZer (Significance of ZERo crossings of deriv.)

Show MassFlux\MassFlux1d1p1s.ps

Idea: at a “bump” $\hat{f}$ goes up then down, so highlight as

Blue when deriv. significantly > 0

Purple when deriv. not significant

Red when deriv. significantly < 0

For more on SiZer:

http://www.stat.unc.edu/faculty/marron/DataAnalyses/SiZer_Intro.html

Investigation of PC1 Clusters:

SiZer analysis: find 3 significant clusters!

- Correspond to 3 known "cloud types"

Improved view of PCA, highlight the clusters in the PCA

Show MassFlux/MassFlux1d1p2.ps

Draftsman's Plot: Can get "better separation" with

"better chosen directions"???

show MassFlux\MassFlux1d1p3.ps

Investigation of “better directions” for PC3 and PC4

Idea: “rotate” subspace gen’d by PC3 and PC4

To better “visually separate” colors

Axes shown in MassFlux\MassFlux2d1p1.ps

Result: “better separation”

Show MassFlux\MassFlux2d1p2.ps

Really useful direction????

Show MassFlux/MassFlux2d1p3.ps

Goodness of Approximation

I.e. how many basis elements to use

E.g. Corpora Callosa data

Recall "shape representations" are based on
$d = 80$ dimensional "feature vectors"

Show CCFrawAlls3.mpg

How big does d need to be?

A personal working assumption:

"shape is complicated, so need d large"

Major sticking point

For medical image shapes, usually have “few data points”,

$n < d$

Personal approach:

- that complicates matters
- but “shape” is “complex” and requires complex rep’n
- hence need to develop new statistical methods:

High Dimension Low Sample Size

Classical Approach

- Statistical Multivariate Analysis is based on "standardizing"
- Multiply by $\hat{\Sigma}^{-1/2}$ (for covariance matrix)
- Requires $n > d$ (else matrix inverse doesn't exist)
- For $n \leq d$, do "dimension reduction"
- For example, keep only the "1st few Principal Components"

Today's Questions:

Is dimension reduction (e.g. PCA based) "good enough"?

Or is it important to develop HDLSS methods?

Aside: how well do ANOVA sums of squares "capture shape"?

Study in context of corpus collosum data

Fourier Approximation Background:

Represent:

$$Shape = \sum_{j=1}^{d} c_j BE_j$$

where the c_j are the “Fourier Coefficients”

and where the BE_j are “basis element” shapes

Possible web site: http://www.vision.ee.ethz.ch/~brech/test_fourdem.html

Some examples of generated shapes:

Show CorpColl\CCFbasis.ps

Approximation 1: Raw Fourier Coefficients

View “goodness of approximation” of

$$k - approx.\, Shape = \sum_{j=1}^{k} c_j BE_j$$

for $k = 0,1,2,\ldots,d$

show CCFappFourAlls3C4.mpg

- $k = 0$ single point: the “zero function”
- $k = 1$ just a line
- $k = 2,3$ still a line (due to “shape normalization”)

Approximation 1: Raw Fourier Coefficients (cont.)

- $k = 4$ ellipse
- $k > 4$ more complicated shapes
- larger k get convergence towards full shape
- $k = 80 = d$ blue completely covers white

Approximation 1: Raw Fourier Coefficients (cont.)

ANOVA style Sums of Squares:

$$Signal\ Power(k - Approx.) = \sum_{j=1}^{k} c_j^2$$

Measures “goodness of fit”, on scale of “energy”

Energy decomposition: c_j^2 is “power in signal in direction BE_j”

Show upper left of CCFappFourAlls3.ps

Approximation 1: Raw Fourier Coefficients (cont.)

Useful scales:

- log scales

Show bottom row of CCFappFourAlls3.ps

- relative scale: $c_j^2 / \sum_{j'=1}^{d} c_{j'}^2$

Show center of CCFappFourAlls3.ps

- cumulative relative scale: $\sum_{k=1}^{k} c_k^2 \Big/ \sum_{j'=1}^{d} c_{j'}^2$

Show right of CCFappFourAlls3.ps

Approximation 1: Raw Fourier Coefficients (cont.)

What does "cumulative relative signal power" really measure?

Again show CCFappFourAlls3C4.mpg

- $k = 2$ line alone is 93%
- $k = 6$ nearly elliptical is 95%???
- $k = 12$ 99%, but still "misses lots of shape"
- $k = 25$ 99.9%, still don't have all of this "shape"?

Have looked at some others: similar lessons

Approximation 2: Centered Fourier Coefficients

Main idea: subtract out the mean first

- standard in ANOVA (often huge part of Sums of Squares)
- results in much different interpretation (of relative SS)

When is "90% of SS explained"?

- Case 29: 31 terms: all of shape

show CCFappCFourAlls3C3.mpg

- Case 2: 11 terms: missed a lot of shape

show CCFappCFourAlls3C1.mpg

Approximation 2: Centered Fourier Coefficients (Cont.)

Paradox of cumulatives ("data compression" plots):

- Case 2 has "great compression" (high curve), yet needs ~50 terms (99.8% explained) for "good shape rep'n"
- Case 29 has "poor compression" (low curve), yet needs only ~32 terms (92.53% explained) for "good shape rep'n"

Personal conclusion:

"shape" manifestations of Sum of Square Analysis is "slippery"

Approximation 3: Principal Component Analysis

Recall Ideas:

- Find "directions of greatest variability"
- Will "maximize signal compression"
- Works in an "average sense", not individually
- Use 1st k for "dimensionality reduction"

Approximation 3: Principal Component Analysis (cont.)

Overlay of cumulatives:

Show CCFappPCAAlls3.ps

- cumulative eigenvalues (“average”) shown in yellow
- much better signal compression than centered Fourier

flip back to CCFappCFourAlls3.ps

- colored cases are extremes of signal compression:
- Case 2 is “great”, Case 13 and Case 29 are “poor”
- Case 35 is “closest to average”

Approximation 3: Principal Component Analysis (cont.)

How well does "90%" capture "shape"?

- Case 2: poor (happens at $k = 1$)

Show CCFappPCAAlls3C1.mpg

- Case 13 and Case 29 good (happens at $k = 16$ and $k = 17$)

Show CCFappPCAAlls3C2.mpg and CCFappPCAAlls3C3.mpg

- Case 35 not quite (happens at $k = 6$)

Show CCFappPCAAlls3C4.mpg

- k is more useful than "% variability"?

Approximation 3: Principal Component Analysis (cont.)

How many terms are needed to capture shape?

- Case 2: $k = 17$?

Show CCFappPCAAlls3C1.mpg

- Case 13 $k = 15$?

Show CCFappPCAAlls3C2.mpg

- Case 29 $k = 16$?

Show CCFappPCAAlls3C3.mpg

- Case 35 $k = 15$?

Show CCFappPCAAlls3C4.mpg

Personal conclusions

- “Sums of Squares” are very crude surrogate for “shape”
- Not enough to “just work with 1st k PCs”
- Not enough to “just work with PCs with top 95% of signal”
- Careful about “average fit” (as in PCA), vs. “individuals”
- 15 – 20 PCs “captures shape for Corpus Callosum data”
- Expect more needed for higher dim’nal objects

Show GreggTracton.html

- Still worth developing HDLSS