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, Last Time (cont.)

Idea: Find “directions that maximize independence”

Parallel Idea: Find directions that maximize “non-Gaussianity”

Related concept: starting from independent coordinates

“most projections are Gaussian”

Mathematics behind this:

Diaconis and Freedman (1984*) Annals of Statistics*, 12, 793-815.

ICA, Last Time (cont.)

Examples:

- 2-d Toys

- Curves as Data

- Corpora Callosa Data

ICA Global Solutions

Interesting question:

How do "sol'ns found by FastICA relate to global sol'ns?"

Approach: ICA attempts to maximize absolute value of kurtosis

How good are these solutions?

Assess by showing kurtosis of projections

ICA, CurvDat Examples Revisited

E.g. Parabs Up and Down (two distant clusters)

Recall PCA:

Show CurvDat/ParabsUpDnCurvDat.ps

- Found clusters in PC1
- Other PCs found other clusters

ICA, CurvDat Examples Revisited (cont.)

Default ICA:

Show CurvDat/ParabsUpDnCurvDatICA.ps

- Recall found “unimportant directions”
- Driven by outliers (see projections)
- Kurtosises (6.7, 6.0, 2.5) seem OK
- Kurtosises driven by outliers

ICA, CurvDat Examples Revisited (cont.)

PCA start ICA:

Show CurvDat/ParabsUpDnCurvDatICAt2.ps

- Recall found "right direction"
- Wondered about local minima
- "Correct direction" had absolute kurtosis = 1.9
- Not global maximizer
- But not far from "previous best 3"

Careful look at Kurtosis

Recall for standardized (mean 0, var 1) data: $Z_1, ..., Z_n$,

$$\text{Kurtosis} = \frac{1}{n}\sum_{i=1}^{n} Z_i^4 - 3$$

- for $Z_i \sim N(0,1)$, Kurtosis = 0
- Kurtosis "large" for high peak, low flanks, heavy tails?
- Kurtosis "small" for low peak, high flanks, light tails?
- Can show Kurtosis $\geq$ -2
- Thus very assymetric? (see above examples)

Careful look at Kurtosis (cont.)

E.g. three point distribution, with probability mass function:

$$f_w(x) = \begin{cases} \dfrac{1-w}{2} & x = \dfrac{-1}{\sqrt{1-w}} \\ w & x = 0 \\ \dfrac{1-w}{2} & x = \dfrac{1}{\sqrt{1-w}} \end{cases}, \qquad \text{for} \quad w \in [0,1]$$

Some simple Calculations:

- $EX = 0, \quad \mathrm{var}(X) = 1, \quad EX^4 = \dfrac{1}{1-w}$

Careful look at Kurtosis (cont.)

Special Cases:

Show ICAeg4p1.ps

- $w = 0$ (no weight in middle), Kurtosis = -2 (minimum)
- $w = 1/3$ (uniform), Kurtosis = -1.5
- $w = 2/3$ Kurtosis = 0, (closest to Gaussian)
- $w > 2/3$ (heavy tails), Kurtosis > 0, (finally positive)
- $w \approx 1$ (2 outliers), Kurtosis very large

Note strong asymmetry in Kurtosis

Careful look at Kurtosis (cont.)

Aapo Hyvärinen comments:

Solve asymmetry problem with "different nonlinearities",

i.e. replace absolute kurtosis = $\left| E(\underset{\rightarrow}{w}^t \underset{\rightarrow}{Z})^4 - 3 \right|$ with:

1. "tanh": $\left(E\left|\underset{\rightarrow}{w}^t \underset{\rightarrow}{Z}\right| - \sqrt{\frac{2}{\pi}} \right)^2$ (since $E\left|N(0,1)\right| = \sqrt{\frac{2}{\pi}}$)

2. "gaus": $\left(E\varphi\left(\underset{\rightarrow}{w}^t \underset{\rightarrow}{Z}\right) - \frac{1}{2\sqrt{\pi}} \right)^2$ (since $E\varphi\left(N(0,1)\right) = \frac{1}{2\sqrt{\pi}}$)

Careful look at Kurtosis (cont.)

Comparison via 3 point example:

Show ICAeg4p2.ps

- upper left: noncomparable scales
- upper right: max rescaling is better
  - tanh and gaus "less asymmetric" than A. Kurt.
- lower left: still shows all are asymmetric
- lower right: "best scale"
  - A. Kurt. has pole at left, but "best for small w"

ICA, CurvDat Examples Revisited (cont.)

E.g. Parabs Up and Down (two distant clusters)

Tanh:

Show CurvDat\ParabsUpDnCurvDatICAt3.ps

- Only IC2 finds an outlier
- IC1 and IC3 have kurt. < 0
- IC3 finds most of 2 clusters
- but not so well as PC1

ICA, CurvDat Examples Revisited (cont.)

Gaus:

Show CurvDat\ParabsUpDnCurvDatICAt4.ps

- IC1 is classical “heavy tail kurtosis”
- IC2 nicely finds clusters
- IC3 is another bimodal direction (no insights abut data)

Conclusion: tanh and gaus seem to work as expected

ICA for Corpora Collosa Data Revisited

Recall: shapes of “window” between brain halves

Show CorpColl\CCFrawAlls3.mpg

Discrimination problem: Schizophrenics vs. Controls

Show CorpColl\CCFrawSs3.mpg & CCFrawCs3.mpg

Recall PCA gave 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)

ICA results: Found “outliers” (and identified them)

Show CorpColl\CCFicaSCs3IC1.mpg, CCFicaSCs3IC2.mpg, CCFicaSCs3IC3.mpg and CCFicaSCs3IC4.mpg

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 1^{st} 4 PCs:

- No help, similar to PCA directions

Sol’n 2: Use PCA “starting values” (as 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 (recall Cornea Data)

Show CorneaRobust\SphericalPCA.ps

Show CorpColl\CCFicaSCs3IC1v4.mpg, CCFicaSCs3IC2v4.mpg, CCFicaSCs3IC3v4.mpg, CCFicaSCs3IC4v4.mpg

- 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

Show CorpColl\CCFicaSCs3IC1v5.mpg, CCFicaSCs3IC2v5.mpg, CCFicaSCs3IC3v5.mpg, CCFicaSCs3IC4v5.mpg

- Similar lessons (2 old, 2 new outliers)

ICA for Corpora Collosa Data (cont.)

Sol’n 5: Different “ICA non-linearities”,

5a: tanh, random start:

Show CorpColl\CCFicaSCs3IC1v6.mpg, CCFicaSCs3IC2v6.mpg, CCFicaSCs3IC3v6.mpg, CCFicaSCs3IC4v6.mpg

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

5b: tanh, PC start:

Show CorpColl\CCFicaSCs3IC1v7.mpg, CCFicaSCs3IC2v7.mpg, CCFicaSCs3IC3v7.mpg, CCFicaSCs3IC4v7.mpg

- all outlier driven

ICA for Corpora Collosa Data (cont.)

5c: gauss, random start:

Show CorpColl\CCFicaSCs3IC1v8.mpg, CCFicaSCs3IC2v8.mpg, CCFicaSCs3IC3v8.mpg, CCFicaSCs3IC4v8.mpg

- same lessons (even same outliers) as above

5d: gauss, PC start:

Show CorpColl\CCFicaSCs3IC1v9.mpg, CCFicaSCs3IC2v9.mpg, CCFicaSCs3IC3v9.mpg, CCFicaSCs3IC4v9.mpg

- 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. A. Kurt., random start:

Show CorpCol\CCFicaSCs3allv21.ps

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

b. A. Kurt., PC start:

Show CorpCol\CCFicaSCs3allv22.ps

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

ICA for Corpora Collosa Data (cont.)

c. Tanh, random start:

Show CorpCol\CCFicaSCs3allv23.ps

- 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:

Show CorpCol\CCFicaSCs3allv24.ps

- OK, but not so good as (c)

ICA for Corpora Collosa Data (cont.)

e. Gaus, random start:

Show CorpCol\CCFicaSCs3allv25.ps

- similar to above

f. Gaus, PC start:

Show CorpCol\CCFicaSCs3allv26.ps

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

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