ORIE 779: Functional Data Analysis

From last meeting

Independent Component Analysis

Idea: Find "directions that maximize independence"

ICA, motivating example

"Cocktail party problem":

Toy Example

Mixed version of above toy example

Scatterplot View (signal processing): plot

- scatterplot for signals $\{(s_1(t), s_2(t)) : t = 1, \ldots, n\}$
- scatterplot for data $\{(x_1(t), x_2(t)) : t = 1, \ldots, n\}$

Resulting ICA Decomposition

ICA algorithm

Parallel Idea (and key to algorithm):

Find directions that maximize "non-Gaussianity"

Reason: starting from independent coordinates

"most projections are Gaussian"

(since projection is "linear combo")

Mathematics behind this:

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

ICA, Algorithm (cont.)

Identifiability problem 1: Generally can't order rows of S_S (& S)

Identifiability problem 2: Can't find scale of elements of $\underrightarrow{s}$

So choose scale so each signal $s_i(t)$ has "unit average energy":

$$\frac{1}{n}\sum_t s_i(t)^2$$

ICA Toy Examples

E.g. Two sine waves [combined graphic]

- PCA finds wrong direction
- ICA is good

E.g. Two realizations of Gaussian noise [combined graphic]

- PCA finds "axis of ellipse" (happens to be "right")
- ICA is "wrong" (different noise realization)

ICA, Toy Examples (cont.)

E.g. Long parallel points clouds [combined graphic]

- PCA finds PC1: "noise" PC2: "signal"
- ICA finds signal in IC1 (most non-Gaussian), noise in IC2
- ICA again loses scale

E.g. Sine wave and Gaussian noise [combined graphic]

- PCA finds "diagonal of parallelogram"
- ICA gets it right
- but magnifies the noise

ICA, Toy Examples (cont.)

E.g. Balanced Sine and Noise [combined graphic]

- Note PCA gives “even split of sine wave”

- Thus fails poorly

- ICA gives excellent “denoising”

- OK for one direction to be Gaussian (just not *both*)

ICA, Toy Examples (cont.)

Now try ICA for FDA analysis (*not* the time series view)

E.g. Recall PCA for "Parabolas" [graphic]

- Mean captured "parabola" shape
- PC1 is "vertical shift"
- PC2 is "tilt" (hard to see visually)
- Remaining PCs are "Gaussian noise"

ICA, Toy Examples (cont.)

Corresponding ICA for “Parabs” [graphic]

- mean and centered data as before

- sphered data has “no structure” (i.e. this structure is “all in covariance”, i.e. have Gaussian point cloud)

- sphered ICs choose “random non-Gaussian” directions

- sphered ICs seem to find outliers

- Original scale versions capture some “vertical shift”

- Non-orthogonality on original scale $\Rightarrow$ hard to interpret

ICA, Toy Examples (cont.)

E.g. Recall PCA for “Parabs with 2 outliers” [graphic]

- Mean captured “parabola” shape
- PC1 is “vertical shift affected by hi-freq outlier”
- PC2 is “most of high freq.outlier”
- “low freq outlier” and “tilt” are mixed between PC3 & PC4
- hope ICA can “separate these”???

ICA, Toy Examples (cont.)

Corresponding ICA for “Parabs with 2 outliers” [graphic]

- ICA finds both outliers well (non-Gaussian direction)
- ICA still misses “shift” and “tilt”
- Since these are *elliptical point cloud properties*, that are ignored through sphering.
- ∃ analysis which keeps “both kinds of features”????
- apply one to the “residuals” of the other?
- E.g. ICA after 1^{st} two robust PCs removed?

ICA, Toy Examples (cont.)

E.g. Recall PCA for “3 bumps, with 2 independent” [graphic]

- Finds both sets of bumps in PC1 and PC2
- Slight mixing of clusters

Corresponding ICA for “3 bumps, with 2 independent” [graphic]

- Bumps not found (since are “Gaussian” features)
- sphering eliminated bumps

ICA, Toy Examples (cont.)

E.g. Recall PCA for “Parabs Up and Down” (2 clusters) [graphic]

- PC1 finds clusters
- Others find usual structure (vertical shift and tilt)

Corresponding ICA for “Parabs Up and Down” [graphic]

- Clusters not found???? (seems very “non-Gaussian”)
- sphering killed clusters????
- Problem with numerical search algorithm????

ICA, Toy Examples (cont.)

Attempted fix 1: Change of “nonlinear function” [graphic]

- similar results
- same happened for other choices

Attempted fix 2: use PCA directions as “starting value” [graphic]

- Gives good solution
- Is this a general problem????
- How generalizable is this solution????

ICA, Toy Examples (cont.)

Aapo Hyvärinen comments:

- “Random start” is deliberate choice
- Even though it is might give “sub-optimal” solutions
- Shows “local minina”, by different answers on replication
- Thus you find out when there are local minima

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, Toy Examples Revisited

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

Recall PCA: [graphic]

- Found clusters in PC1
- Other PCs found other structure

Recall Default ICA: [graphic]

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

ICA, Toy Examples Revisited (cont.)

Recall PCA start ICA: [graphic]

- Recall found “right direction”
- Wondered about local minima
- “Correct direction” had absolute kurtosis = 1.9
- *Not* global maximizer
- so random start ICA was “OK”
- But not far from “previous best 3”

Careful look at Kurtosis

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

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

- 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 (point masses at +-1)
- Thus very assymetric? (see above examples)

Careful look at Kurtosis (cont.)

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

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

Some simple Calculations:

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

Careful look at Kurtosis (cont.)

Special Cases: [graphic]

- $w = 0$ (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 = $\left| E(\underrightarrow{w}^t \underrightarrow{Z})^4 - 3 \right|$ with:

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

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

Careful look at Kurtosis (cont.)

Comparison via 3 point example: [graphic]

- 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"
    - tanh and gaus have different zeros than A. Kurt.