A Few Words on ICA

Prof. Sharon McNicholas

STATS 780/CSE 780

Introduction

- This is the second in a series of a selected topics to be covered between this "lecture" and the end of term..
- We have seen factor analysis and principal component analysis.
- Now we will take a look at independent component analysis (ICA).
- The first two figures are taken from Hastie et al. (2009)[a]

[a] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. Springer: NY.

Basic Idea

- A little like factor analysis and principal component analysis (PCA).
- In factor analysis, the latent factors are normally distributed and independent.
- In PCA, the components are uncorrelated.
- In ICA, the (latent) components are independent (and non-Gaussian).
- Recall that uncorrelated is not the same as independent, i.e., independent variables are necessarily uncorrelated but not vice versa.

Time Series Ex. (from Hastie et al.)

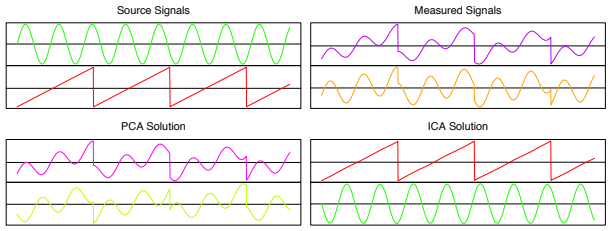


FIGURE 14.37. *Illustration of ICA vs. PCA on artificial time-series data. The upper left panel shows the two source signals, measured at* 1000 *uniformly spaced time points. The upper right panel shows the observed mixed signals. The lower two panels show the principal components and independent component solutions.*

Uniform Ex. (from Hastie et al.)

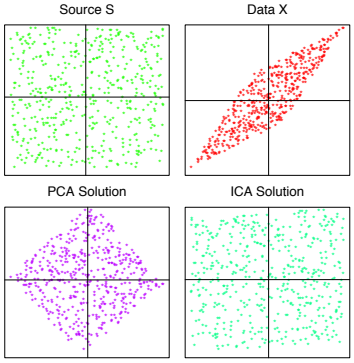
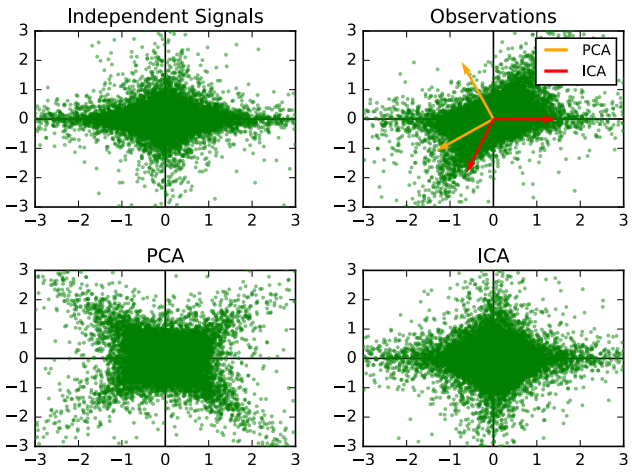


FIGURE 14.38. *Mixtures of independent uniform random variables. The upper left panel shows* 500 *realizations from the two independent uniform sources, the upper right panel their mixed versions. The lower two panels show the PCA and ICA solutions, respectively.*

Example Using Python



Comments

- Hastie et al. (2009, Ch. 14) discuss ICA and give some interesting examples.
- I really like the paper by Hyvärinen and Oja (2000)[a].
- I see education in data science / statistics as the assembly of a toolbox of techniques, and I think ICA is a good tool to understand.

[a] Hyvärinen, A. and Oja, E. (2000). 'Independent component analysis: Algorithms and applications'. *Neural Networks* **13**(4–5), 411–430.

Comments contd.

- In addition to taking a quick look at ICA, the previous and following figures are form Python.
- I like Python and I think it is perhaps underused in the statistics community.
- A good place to start with Phyton is through Anaconda, see `www.continuum.io/anaconda-overview`.
- There are many excellent resources that support Python, e.g., the code to produce the figures on the previous and following slides is based on code from scikit-learn (`scikit-learn.org/stable/`).

Another Phython Example: k-Means