CS456: Machine Learning

Independent Component Analysis

Jakramate Bootkrajang

Department of Computer Science
Chiang Mai University

March 12, 2020

Objectives

- To understand general idea of independent component analysis
- To understand Blind Source Separation (BSS) problem
- To understand general idea to unmix signal in BSS

Outlines

- Motivation
- Blind source separation problem
- Mixing and unmixing signal
- Maximsing Kurtosis

Motivation: Blind Source Separation (BSS) problem

There are two speakers, and two microphone in the room. Each microphone picks up sound of two speakers. Want to unmix the mixture signal into original voices.

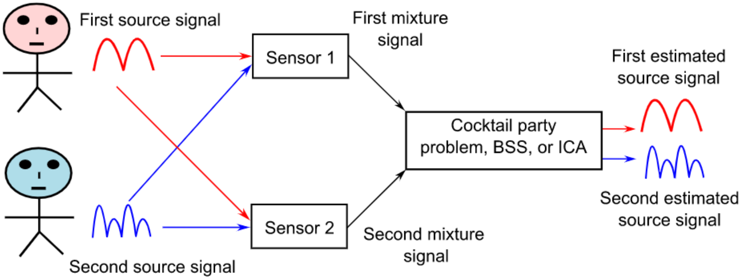


Figure: Alaa Tharwat, Independent component analysis: An introduction

BSS problem (1/2)

- Let $\mathbf{s}_i = (s_i^1, s_i^2, \ldots, s_i^m)$ denotes source signal
  - m is the length of signal
  - s_i^t can be amplitude of s_i at time step t
- Let $\mathbf{x}_i = (x_i^1, x_i^2, \ldots, x_i^m)$ denotes mixture of signals
- Time is not important here, we see signal as a vector of amplitudes rather than a time-series data

BSS problem (2/2)

- Suppose there are two source signals $\mathbf{s}_1$ and $\mathbf{s}_2$, and there are two mixtures $\mathbf{x}_1, \mathbf{x}_2$.
- $\mathbf{x}_1, \mathbf{x}_2$ is a linear combination of source signals say

$$\mathbf{x}_1 = a\mathbf{s}_1 + b\mathbf{s}_2$$

$$\mathbf{x}_2 = c\mathbf{s}_1 + d\mathbf{s}_2$$

- We can write the mixing procedure in matrix form

$$X = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} a\mathbf{s}_1 + b\mathbf{s}_2 \\ c\mathbf{s}_1 + d\mathbf{s}_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} = AS \tag{1}$$

Sources mixing process

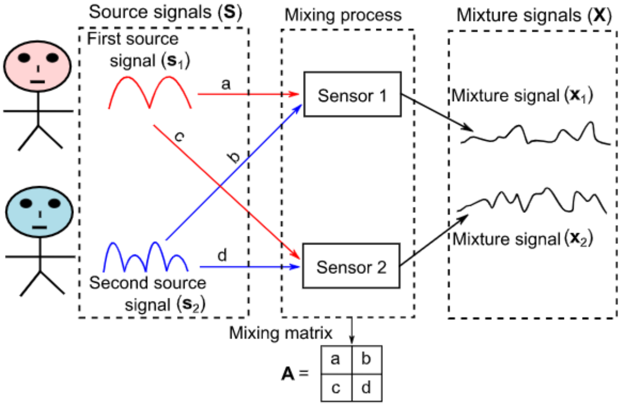


Figure: Alaa Tharwat, Independent component analysis: An introduction

Unmixing signals

- Given mixture of signals $\mathbf{x}_i = (x_i^1, x_i^2, \ldots, x_i^m), i = 1 : N$
- Want to extract source signals from these mixtures
- This can be done by searching for unmixing coefficients $\alpha, \beta, \gamma, \delta$

$$\alpha \mathbf{x}_1 + \beta \mathbf{x}_2 = \mathbf{y}_1 \approx \mathbf{s}_1$$

$$\gamma \mathbf{x}_1 + \delta \mathbf{x}_2 = \mathbf{y}_2 \approx \mathbf{s}_1$$

- In matrix form

$$Y = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \alpha \mathbf{x}_1 + \beta \mathbf{x}_2 \\ \gamma \mathbf{x}_1 + \delta \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = WX \tag{2}$$

Unmixing process

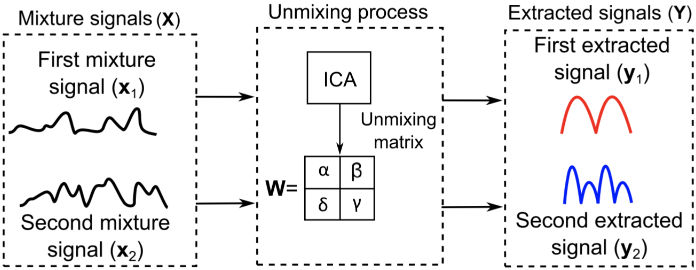


Figure: Alaa Tharwat, Independent component analysis: An introduction

Is it possible to unmix signals ?

- Independent Component Analysis (ICA) can be used to solve BSS

$$Y = WX$$

- We only know X, can we find W ?
- Even if we can find W, how do we know Y are the original signals ?

ICA assumption

In order to unmix the signal ICA makes some assumptions

- Source signals must be independent
- Source signal must not have Gaussian distribution $\star$
- Source signal must be less complex than the mixture of signal

Non-Gaussianity assumption

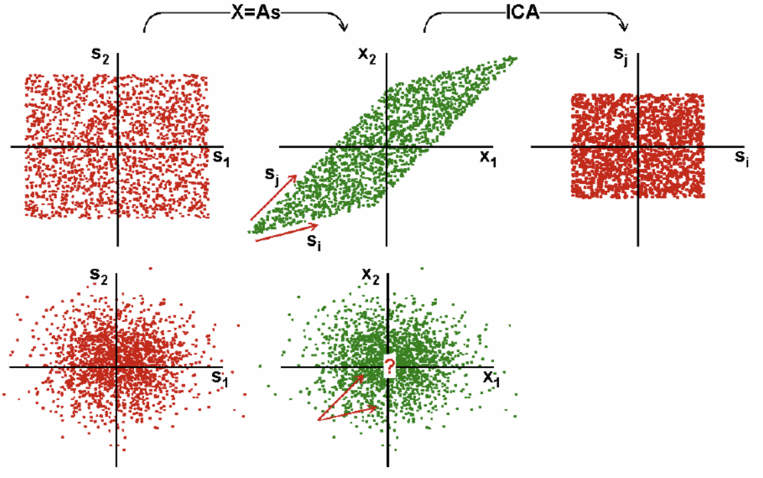


Figure: Gutierrez-Osuna

Sources are not Gaussian

Histogram of source signal

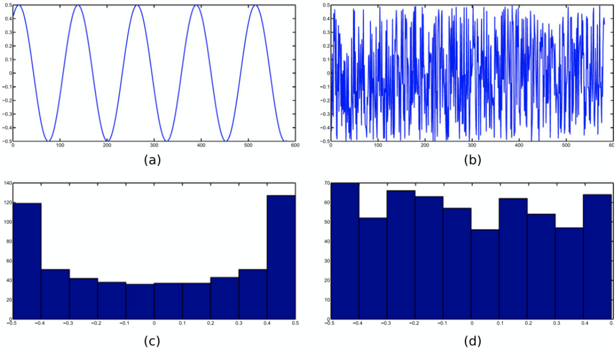


Figure: Alaa Tharwat, Independent component analysis: An introduction

Mixture can be Gaussian

Histogram of mixture of signals

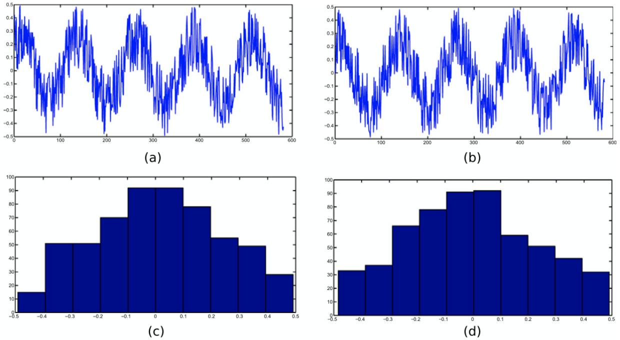


Figure: Alaa Tharwat, Independent component analysis: An introduction

ICA ideas

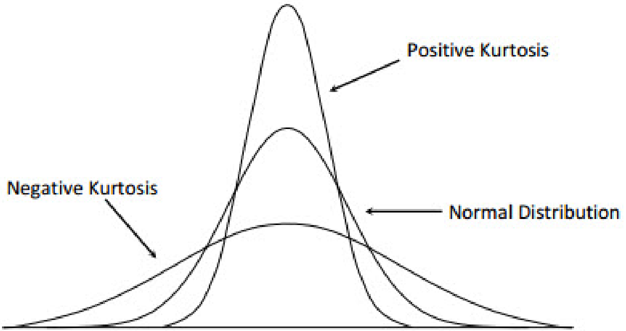
- Let us consider one independent component
- Since $\mathbf{y}_i = \mathbf{w}_i X$ is like projecting data using projection vector $\mathbf{w}_i$
- And we know that after projection $\mathbf{y}_i$ is vector with non Gaussian distribution
- The quality of the projection can then be measured by how non-Gaussian $\mathbf{y}_i$ is[1]

[1] In PCA, the quality after projection is measured by variance

Kurtosis: Measuring non-Gaussianity

- Kurtosis: measures flatness of distribution compared to Gaussian (Kurtosis for Gaussian $= 3$)
- Kurtosis for a random variable x is $K(x) = E[x^4] - 3[E[x^2]]^2$
- Some people define *excess* Kurtosis $K(x) - 3$, so the excess Kurtosis for Gaussian $= 0$

Kurtosis



So absolute value of (excess) Kurtosis can be used as measure of good projection direction for ICA

Maximising kurtosis

- ICA objective $\mathbf{y}_i = \mathbf{w}_i X$
- If we standardise X to have unit variance, call the new data Z
- And normalise weight vector $\mathbf{w}_i \in W$ to unit length
- Kurtosis of $\mathbf{y}_i$ is

$$K(\mathbf{y}_i) = E[(\mathbf{w}_i Z)^4] - 3 \tag{3}$$

- To maximise we find the gradient of the Kurtosis

$$\frac{\partial K(\mathbf{w}_i Z)}{\partial \mathbf{w}_i} = cE[Z(\mathbf{w}_i Z)^3] \tag{4}$$

Gradient descent (one-unit ICA)

- We can update $\mathbf{w}_i$ using gradient descent

$$\mathbf{w}_i^{new} = \mathbf{w}_i^{old} + \eta \frac{\partial K(\mathbf{w}_i Z)}{\partial \mathbf{w}_i} \tag{5}$$

- And normalise $\mathbf{w}_i$ to have unit length

$$\mathbf{w}_i = \frac{\mathbf{w}_i}{||\mathbf{w}_i||} \tag{6}$$

Ensuring each $\mathbf{w}_i$ is uncorrelated

- We can estimate several projection vectors
- We need to prevent them from converging to the same solution
- This can be done based on Gram-Schmidt process
  - Estimate each independent component one by one, say we get $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p$
  - Using one-unit ICA to estimate $\mathbf{w}_{p+1}$ using gradient descent
  - after each iteration uncorrelate $\mathbf{w}_{p+1}$ from $\mathbf{w}_j, j = 1 : p$, and normalise $\mathbf{w}_{p+1}$

$$
\begin{aligned}
\mathbf{w}_{p+1} &= \mathbf{w}_{p+1} - \sum_{i=1}^{p} (\mathbf{w}_{p+1}^T \mathbf{w}_j)\mathbf{w}_j \\
\mathbf{w}_{p+1} &= \frac{\mathbf{w}_{p+1}}{||\mathbf{w}_{p+1}||}
\end{aligned}
$$

ICA limitation

- The order of independent components are not guaranteed
- The sign of independent components can be flipped
- It cannot estimate Gaussian component signal

ICA vs PCA

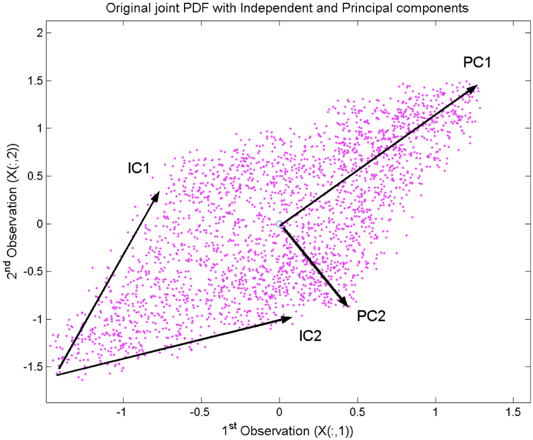


Figure:
http://labs.seas.wustl.edu/bme/raman/Lectures/Lecture14_ICA.pdf

Objectives: revisited

- To understand general idea of independent component analysis
- To understand Blind Source Separation (BSS) problem
- To understand general idea to unmix signal in BSS