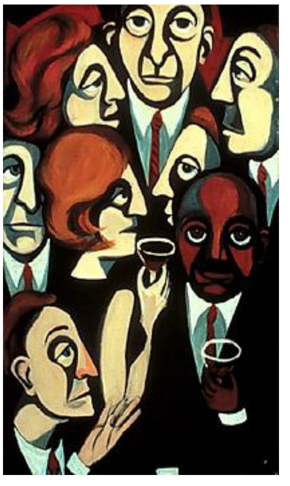
The Cocktail-Party Problem



Faith Riggold 1964 "Cocktail-Party"

ICA

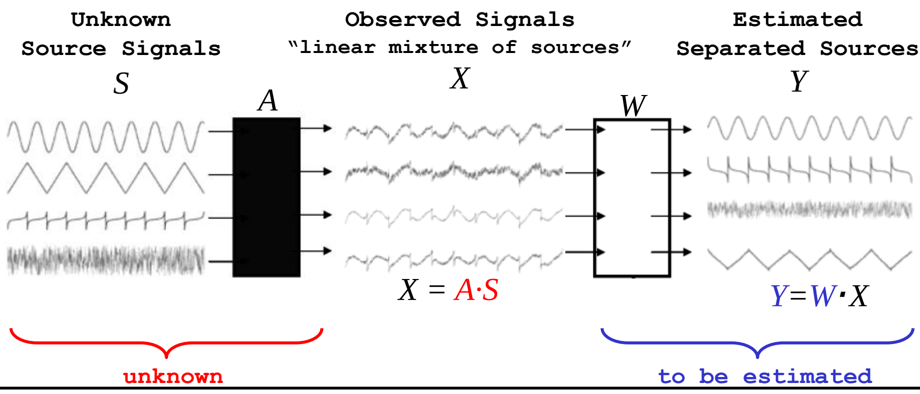
Independent Component Analysis

http://www.cis.hut.fi/aapo/papers/IJCNN99_tutorialweb/IJCNN99_tutorial3.html

ICA: the task

The "BSS: Blind Source Separation" or "Cocktail-Party" problem:

Given n Signals X that are a linear mixture of unknown source signals S, can we estimate the source signals ?



ICA: the problem

4

Given: $X = \{ x_i(t) \mid 1 \leq i \leq n, 1 \leq t \leq T \}$, a $n \times T$ data matrix

Problem: How to decompose the matrix into $X = AS$ with an unknown mixing matrix A and unknown source signal $S_{i\bullet}$

$$X = A \times \begin{pmatrix} \cdots S_{1\bullet} \cdots \\ \cdots S_{2\bullet} \cdots \\ \vdots \\ \cdots S_{m\bullet} \cdots \end{pmatrix}$$

So far only linearity is assumed. -> Many solutions & ambiguities !!!

Ambiguities:

1. The variance (scaling) of $S_{i\bullet}$ cannot be determined, either a scalar multiplication of A or of S -> we should normalize the sources.
2. The order of the sources is arbitrary.

5

ICA: the data model

X = A x $\begin{bmatrix} \cdots\ s_1 \cdots \\ \cdots\ s_2 \cdots \\ \vdots \\ \cdots\ s_m \cdots \end{bmatrix}$

Question:

What could be an assumption on the sources S that helps to decompose the X into A and S ?

Assumption:

1) All sources signals s_i , the rows of S, are statistically independent.
2) Since we can not estimate the magnitude of s_i , we fix it to $E[s_i s_i^T]$ =1 ==> $E[SS^T] = I$

6

ICA: statistical independence

Definition:

Random variables **(vectors)** y_i are statistically independent if

$$P(y_1, y_2, \ldots, y_n) = \prod P(y_i)$$

==> for any function g_i

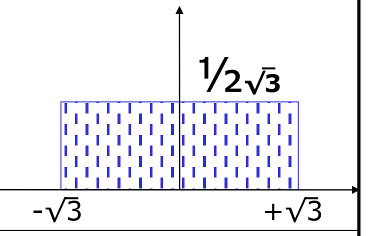
$$E[\, g_1(y_1)\, g_2(y_2) \cdots g_n(y_n)] = \prod E[\, g_i(y_i)\,]$$

ICA: what about PCA

uncorrelated versus independence

$p(s)$: uniform distribution

=> $E[s] = 0$, $E[s^2] = 1$



define $t = 2s^2$

=> $E[s \cdot t] = 2\,E[s^3] = 0$

=> s,t are uncorrelated

also $E[s]E[t] = 0$?? => s, t statistically independent??

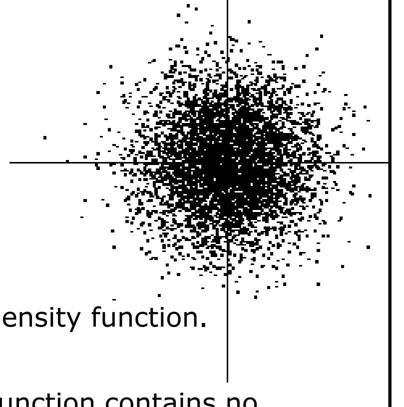
define $f(x) = x^2$ and $g(x) = x$

=> $E[f(s)\,g(t)] = 2E[s^4] = \frac{2(\sqrt{3})^4}{5}$

and $\underbrace{E[f(s)]}_{E[s^2]}\ \underbrace{E[g(t)]}_{2E[s^2]} = 2$

=> s,t are stat. dependent

ICA: gaussian signals are of no use



Let us assume:
all source signals are Gaussian,
uncorrelated and of unit variance.

Then an orthogonal mixing matrix would generate signals X_i with a completely symmetrical Gaussian joint density function.

A completely symmetrical joint density function contains no information on the structure of the mixing matrix A.

If more than one source signal is Gaussian we can not separate the sources with ICA.

ICA: the data model continued

9

Assumptions:

1) non-Gaussian source signals s_i (except possibly one).

2) All sources signals s_i , the rows of S are statistically independent.

3) Since we can not estimate the magnitude of s_i , we fix it to $E[s_i s_i^T] = 1$ ==> $E[SS^T] = I$

ICA: approach

10

We search for a matrix $\boldsymbol{W}$ (ideally $\boldsymbol{W} = \boldsymbol{A}^{-1}$) so that the rows $\boldsymbol{y_i}$ of

$$\boldsymbol{Y} = \boldsymbol{WX}$$

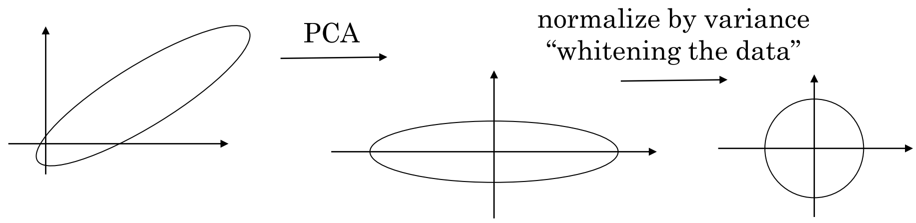
1. are maximal statistically independent,
2. are maximal non Gaussian,
3. and of variance $E[y_i y_i^T] = 1$.

ICA: PCA as preprocessing

PCA
- computes the axis of maximum variance
- these axis are uncorrelated (but only for Gaussian data statistically independent).

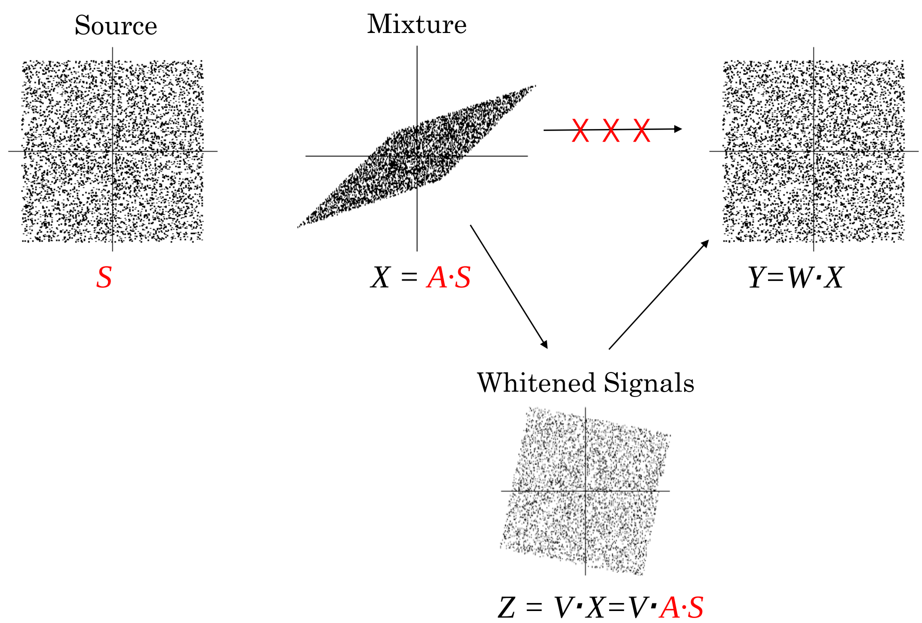
Now we can normalize the axis by their variance

=> $Z = V \cdot X = V \cdot A \cdot S$ *with* $E[z_i z_i^T] = 1$.



for non Gaussian data →

ICA: the procedure



ICA: the problem reformulated twice

13

1.) Non-Gaussian approach:

By the central limit theorem, the PDF of a sum of n independent random variables tends to a Gaussian random variable.

1. Find a measure of non-Gaussianity.
2. Find $\boldsymbol{W}$ such that the outputs PDF are as different as possible from the Gaussian function.

2.) Independence approach :

1. Measure the independence between the signals.
2. Find the signals that maximize this independence.

ICA: measure of non-Gaussian

14

There exist several approaches to measure if a pdf is Gaussian or not!

However, we do not know the full pdf!

→ it is more reasonable to use more global measures of the distribution such as mean, variance,...

Remember, moments and cumulants (semi-invariants) are easy to compute!

i[th] moment: $m_i = E[x^i] = \sum_{I=1}^{N} x^i P(x),\ i = 1, 2, \ldots$

i[th] central moment: $\mu_i = E\left[(x - E\left[x\right])^i\right] = \sum_{I=1}^{N} \left(x - m_1\right)^i P(x)$

ICA: measure of non-Gaussian

15

The cumulants of distribution are:

$$\kappa_1(x_i) = E[x_i] = m_1$$

$$\kappa_2(x_i x_j) = E[x_i x_j] = m_2 - m_1^2 = \sigma^2$$

....

$$\kappa_4(x_i x_j x_k x_l) = E[x_i x_j x_k x_l] - E[x_i x_j]E[x_k x_l] - E[x_i x_k]E[x_j x_l] - E[x_i x_j]E[x_k x_l] - E[x_i x_l]E[x_j x_k]$$

The Kurtosis is then defined:

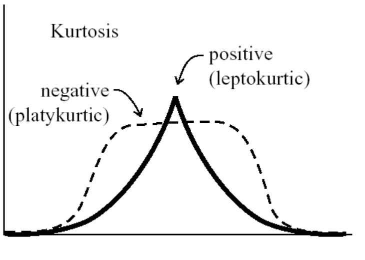
$$kurt(x_i) = \frac{\kappa_4}{\kappa_2^2} = E[x_i^4] - 3(E[x_i^2])^2$$

ICA: measure of non-Gaussian

16

Both cumulants and kurtosis are good to measure the deviation of a distribution from being Gaussian:

For ICA the Kurtosis is commonly applied:



For finding the independent components:

optimize W so that $\sum_j^n \left\| kurt(Y_j) \right\|$ is maximum:

ICA: 2nd approach: independence

Remember:

Definition of stat. independent $p(y_1, y_2, \ldots, y_n) = \prod p(y_i)$

Kullbach-Leibler divergence measures a distance between two pdf's (! not symmetric):

$$L(p_a, p_b) = -\int p_a(x) \ln \frac{p_b(x)}{p_a(x)} dx$$

Now we measure $L\left(p(y_1, y_2, \ldots, y_n) \, , \, \prod p(y_i) \right)$

$$L\left(p(Y), \prod_i^n p(y_i) \right) = -\int p(Y) \ln \frac{\prod p(y_i)}{p(Y)} dY$$

ICA: measure of independence

Now we measure $L\left(p(y_1, y_2, \ldots, y_n) \, , \, \prod p(y_i) \right)$

$$L\left(p(Y), \prod_i^n p(y_i) \right) = -\int p(Y) \ln \frac{\prod p(y_i)}{p(Y)} dY$$

$$= \int p(Y) \ln p(Y) dY \; - \; \sum_i^n \int p(Y) \ln p(y_i) dY$$

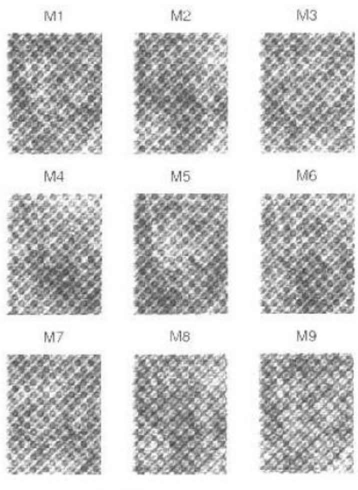
$$= - \; H(Y) \; - \; \sum_i^n \int p(Y) \ln p(y_i) dY$$

$$= - \; H(Y) \; + \; \sum_i^n H(y_i)$$

Def.: Mutual Information: $I(y_1, y_2, \ldots y_n) = -H(Y) + \sum_i^n H(y_i)$

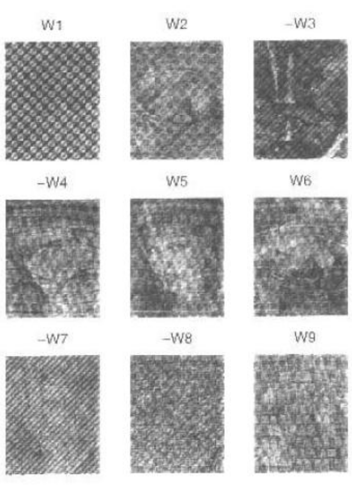
ICA: application to images

Mixtures



ICA: application to images

PCA



ICA

ICA: application to images

ICA

Originals