Today: ICA

Next time: Trees for correlated data gathering

Noiseless linear transformations

$\underline{X} = A\underline{S}$ $\underline{X}, \underline{S}$ – i.i.d. vector random variables of dimension N

– zero-mean

A – separation matrix
- constant
- square $N \times N$
- invertible

if T available samples:

$\underline{X}(t) = A\underline{S}(t)$ $\underline{X}(t) = \begin{bmatrix} X_1(t) \\ \vdots \\ X_N(t) \end{bmatrix}$ $\underline{S}(t) = \begin{bmatrix} S_1(t) \\ \vdots \\ S_N(t) \end{bmatrix}$

$t = 1..T$

$\underline{X}$ has correlation matrix R_X

$$E[\underline{X}\underline{X}^T] = R_X$$

$$R_X(i,j) = E[X_i X_j]$$

Task: Given $\underline{X}$, and knowing that $\underline{S}$ are independent:

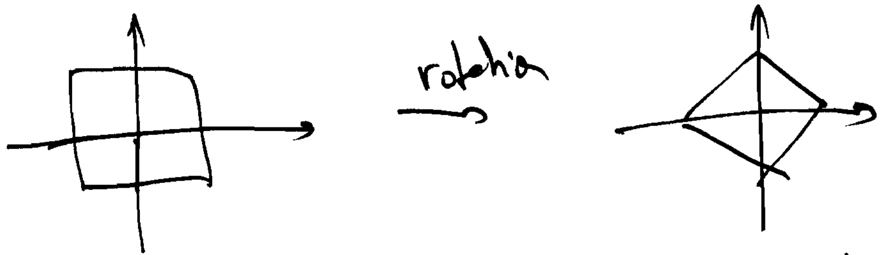
- find $\underline{S}$, A

$$\underline{S} = W \cdot \underline{X}, \quad A = W^{-1}$$

independence $\Rightarrow$ decorrelation

$E[xy] = E[x]E[y] = 0$

$\Leftarrow$ only for normal r.v.



1. Mixtures of normal r.v. : PCA (Karhunen-Loéve 47-48)
2. Mixtures of arbitrary (non-normal) distributions : ICA.

---

1. KLT

Assume x is m.v. Gaussian of correlation matrix R_x

$$x \sim \mathcal{N}(0, R_x)$$

$$x = \frac{1}{\sqrt{(2\pi)^N \det R_x}} \exp\left(-\frac{1}{2} x^T R_x^{-1} x\right)$$

R_x : - psd $\quad a^T R a \geq 0$

- symmetric $\quad E[x_i x_j] = E[x_j x_i]$

$\rightarrow$ it has a complete set of orthog. eigenvectors

- positive eigenvalues

$$R_x \cdot V = \lambda \cdot V$$

- normalize them to unit norm

$$V_x = [\underline{v}_1 \quad \dots \quad \underline{v}_N]$$

$$\Lambda_x = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{bmatrix}$$

Properties: $V_x^T \cdot V_x = V_x V_x^T = I$

$$R_x V_x = V_x \Lambda_x \quad \text{(by def)}$$

$$\Downarrow$$

$$V_x^T R_x V_x = \Lambda_x$$

Remember task find $W \cdot \underline{x} = \underline{s}$, such that $\underline{s}$ are independent

linear comb. of Gauss = Gaussian
indep. = correlated

$\Rightarrow$ need R_s diagonal

$$R_s = E[\underline{s}\,\underline{s}^T] = E[W \underline{x}\,\underline{x}^T W^T] = W\, E[\underline{x}\,\underline{x}^T] W^T =$$
$$= W R_x \cdot W^T$$

Take $W = V_x^T$. — KL transform (or PCA)

Example $N=2$ $\quad R_x = \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix}$

has eigenvectors $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$$W = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad R_s = \begin{bmatrix} r_0 + r_1 & 0 \\ 0 & r_0 - r_1 \end{bmatrix}$$

Prop.

KLT is best MSE approx. among orthonormal transforms (linear subspace)

$M < N$ $\qquad WW^T = I$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{W} \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix} \xrightarrow{\text{Proj.}} \begin{bmatrix} s_1 \\ \vdots \\ s_M \\ 0 \end{bmatrix} \xrightarrow{W^{-1}} \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_N \end{bmatrix}$$

$$\underline{x} \qquad\qquad \underline{s} \qquad\qquad \underline{\hat{s}} \qquad\qquad \underline{\hat{x}}$$

$$\underline{s} = W\underline{x}$$

$$\underline{\hat{x}} = W^{-1} \cdot \underline{\hat{s}} = W^T \cdot \underline{\hat{s}}$$

task $\quad \min\limits_{W} E\left(\|\underline{x} - \underline{\hat{x}}\|_2^2\right)$

W unitary $\Rightarrow \|x - \hat{x}\|_2 = \|s - \hat{s}\|_2$

$$\|s - \hat{s}\|_2^2 = \sum_{m=M+1}^{N} s_m^2 = \sum_{m=M+1}^{N} s_m s_m^T = \sum_{m=M+1}^{N} \underline{w}_m^T \underline{x}\,\underline{x}^T \underline{w}_m$$

$$E\left[\|x - \hat{x}\|_2^2\right] = E\left[\|s - \hat{s}\|^2\right] = \sum_{m=M+1}^{N} \underline{w}_m^T R_x \cdot \underline{w}_m$$

$\min \Leftrightarrow \max \sum\limits_{m=1}^{M} \underline{w}_m^T R_x \underline{w}_m$ $\quad$ u.c. $w_m^T w_m = 1$ $\Rightarrow$ eig. vectors

$$\frac{d}{dw} \underline{w}_m^T R_x \underline{w}_m - \lambda(w_m^T w_m - 1)$$

$$2 R_x \underline{w}_m - 2\lambda w_m = 0$$

- under some conditions, good for compression
- if $\underline{x}$ are block samples of a Markov stationary process

$$x_n = \rho x_{n-1} + w_n \Rightarrow R_x = \begin{bmatrix} 1 & \rho & \cdots & \rho^{N-1} \\ \vdots & \ddots & & \vdots \\ \rho^{N-1} & \cdots & & 1 \end{bmatrix}$$

if $\rho \to 1$, eigen vectors do not depend on $\rho \Rightarrow$
$\Rightarrow$ DCT transform
"~~all~~-use KLT"

- pre-processing in ICA.
(ortho and space reduction)

2. ICA (Herault and Jutten '84)
Comon '94

$\boxed{\text{Post EEG}}$

what if components are not Gaussian?
decorrelation $\neq$ independence

Scene model: $X = AS$
priors: — S independent
— noiseless linear transform.

$$\underline{x} = \begin{bmatrix} a_1 \end{bmatrix} \cdot s_1 + \begin{bmatrix} a_2 \end{bmatrix} \cdot s_2 + \begin{bmatrix} a_3 \end{bmatrix} \cdot s_3 + \dots + \begin{bmatrix} a_N \end{bmatrix} s_N$$

Whitening (decorrelation) using KLT
whitened version:

$$\underline{x}^w = \Lambda_x^{-\frac{1}{2}} V_x^T \underline{x}$$

$$E[\underline{x}^w \underline{x}^{wT}] = E\left[\Lambda_x^{-\frac{1}{2}} \cancel{V_x}^T \underline{x}\,\underline{x}^T \cancel{V_x} \cdot \Lambda_x^{-\frac{1}{2}}\right] =$$

$$= \Lambda_x^{-\frac{1}{2}} \cdot \Lambda_x \cdot \Lambda_x^{\frac{1}{2}} = I$$

$$X = A \cdot S \qquad A^w = \Lambda_x^{-\frac{1}{2}} V_x^T \cdot A \quad ; x^w = A^w S$$

to recover A: $A = V_x \Lambda_x^{\frac{1}{2}} A^w$

- can also reduce dimension, by choosing only largest eig. v.

$X \to x^w$, $A \to A^w$.

Moreover

$\underline{S}$ independent $\Rightarrow$ $\underline{S}$ uncorrelated

Assume $E[\underline{S}\,\underline{S}^T] = I$

$$\underline{X} = W\underline{S} \Rightarrow E(\underline{X}\,\underline{X}^T) = W\,E[SS^T]\cdot W^T \Rightarrow$$

$$\Rightarrow WW^T = I$$

W/out loss of gen., look for orthonormal transforms. (rotations)

Correlation removed

use the other prior: data comes from a linear combination

Idea: Central limit theorem.

m indep. r.v. $z_1, \ldots, z_m$, of 0-mean 1-variance

$$z = \frac{1}{m}\sum_{i=1}^{m} z_i \qquad \lim_{m \to \infty} z \to \mathcal{N}(0,1)$$

what if m is finite?

e.g. $m=2$: $z_1 + z_2$ is "more Gaussian" than z.

Ex. Consider 2 non-normal i.i.d. r.v. z_1, z_2, var. 1 independent

$$0 < \alpha_1, \alpha_2 < 1$$

$$z_1' \triangleq \alpha_1 \cdot z_1 + \sqrt{1-\alpha_1^2}\; z_2$$
$$z_2' \triangleq \alpha_2 z_2 + \sqrt{1-\alpha_2^2}\; z_2$$

Q: Given z_1', z_2', and not knowing α_1, α_2 $\overset{?}{\Rightarrow} z_1, z_2$

Denote $\alpha = \begin{bmatrix} \alpha_1 & \sqrt{1-\alpha_1^2} \\ \alpha_2 & \sqrt{1-\alpha_2^2} \end{bmatrix}$

$\begin{bmatrix} z_1' \\ z_2' \end{bmatrix} = \alpha \cdot \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ choose $\beta \neq \alpha^{-1}$

$$\begin{bmatrix} \hat{z}_1' \\ \hat{z}_2' \end{bmatrix} = \beta \cdot \alpha \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

if $\beta \neq \alpha^{-1}$ $\begin{bmatrix} \tilde{z}_1' \\ \tilde{z}_2' \end{bmatrix}$ "more gaussian" than $\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$

$\boxed{\text{Laplace figure}}$

So, from z_1, z_2, need to find reverse transform that maximizes <u>non-gaussianity</u>

Back to $\underline{X}, \underline{S}$, find $\underline{w^T X}$ maximizing non-gauss

Measure?

~~Negentropy~~:

Entropy: – Gaussian has largest entropy
– Sparse densities have smallest entropy.

Def Diff. entropy of s w. pdf f_s

$$H(s) = -\int f_s(y) \log f_s(y)\, dy$$

Def. Negentropy (min entropy ⇒ max neg.)

$$J(s) = H(S_{Gauss}) - H(s)$$

same ~~covar.~~ matrix as S.

– invariant to linear transformations
– non-negative
– $\equiv 0$ only for Gaussian
– largest for sparse

Def. Mutual information

$$I(s_1 \ldots s_N) = \sum_{i=1}^{N} H(s_i) - H(s_1 \ldots s_N) =$$

Negentropy:

$$I(S_1 \ldots S_N) = \sum_{i=1}^{N} - H(S_{i_{Gauss}}) - \cancel{H(S_i)} + H(S_1, \ldots, S_{N_{Gauss}}) -$$

$$- H(S_1 \ldots S_N) = J(s) - \sum_{i=1}^{N} J(S_i)$$

$s = Wx \Rightarrow J(s) = J(x)$ ed.

min. <u>mutual info</u> $\Rightarrow$ $\max \sum_i J(s_i)$

$$\max_{\{w_i\}} \sum_{i=1}^{N} J(w_i^T \underline{x})$$

u.c. $WW^T = I$

<u>Negentropy needs pdf.</u>

A good approximation

$$J(s) = \frac{1}{2} k_3^2 + \frac{1}{48} k_4^2(s)$$

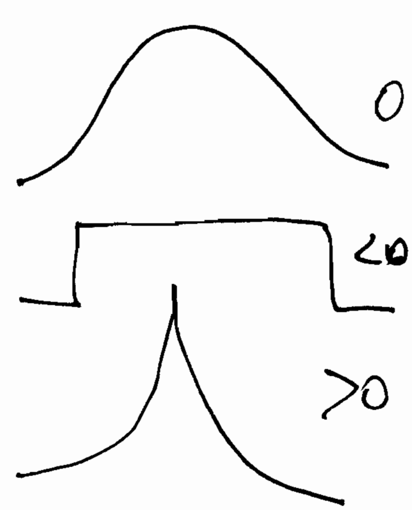
skewness kurtosis

non-zero for most non-gaus.

$k_4(s) = E[s^4] - 3E[s^2]^2$

$$J(s) = k_4^2(s)$$

k_4



or, contrast ~~functions~~

$$\cancel{J_G(s) = \left(E[G(s)] - E[G(\nu)] \right)^2}$$

- general purpose : $G(u) = \log \cosh (au)$
- ~~super G~~ : $\cancel{\exp\left(-a \frac{u^2}{2}\right)}$
- sub. G : kurtosis

Contrast function G

fixed-point ICA (for 1 comp.)

$$\max E[G(\underline{w}^T \underline{x})] \quad \text{under } \|\underline{w}\|^2 = 1$$

$$G' = g$$

$$\frac{\partial}{\partial w} \left(E[G(\underline{w}^T \underline{x})] - \beta(\|\underline{w}\|^2 - 1) \right) = 0$$

$$E[x g(\underline{w}^T \underline{x})] - \beta \underline{w} = 0$$

so $\beta = E[\underline{w}^T \underline{x}\, g(\underline{w}^T \underline{x})]$

Newton meth. $\quad w^+ = w - \frac{f}{f'}$

$$w^+ = w - \frac{E[x g(w^T x)] - \beta w}{E[g'(w^T x)] - \beta} \Rightarrow$$

$$\Rightarrow w^+ = E[x g(w^T x)] - E[g'(w^T x)]\, w$$

$$w^* = \frac{w^+}{\|w\|}$$