Notes on ICA Tutorial

June 15, 2017

1 Component Analysis

The model underlying the different types of component analyses is

$$X = As \qquad \rightarrow \qquad \hat{s} = Wx$$

Where for this simple linear generative model, the Matrix funciton transforming x into $\hat{s}$ is linear, too, i.e. a matrix W.

The Idea now is to find W such that $\hat{s}$ is *interesting* the differen types of analysis are different instantiations of what is considered "interesting".

Properties of $\hat{s}$

- Maximum variance: $\rightarrow$PCA, Dimensionality reduction
- Independence: MI($s_1, s_2, ...$) $\rightarrow$Source separation/identification
- non-Gaussianity: (Sparseness, Kurtosis) $\rightarrow$Projection Pursuit, Sparse coding

Cost functions: "quantifications of interestingess"

- Variance (+Orthogonality) $\rightarrow$PCA
- independence: $\rightarrow$MI Entropy
- non-Gausianity $\rightarrow$kurtosis, negentropy (see independence)

Algorithms: How to optimize the costfunctions

- Analytical solutions (Eigenvalue problme for PCA)
- Gradient Methods (Oja's Rule, plain gradient, natural gradient)
- Fixed Point methods $\rightarrow$`fastICA`

2 Non-Gaussian means independent

Central limit theorem: distribution of a sum of RVs is more gaussian than the individual RV distributions. Sum of 2 RVs is more gaussian than the individual ones.

$$x = a_1 s_1 + a_2 s_2 \qquad |X = as$$

with $Z = WA \rightarrow W \approx \mathbf{P}A^{-}$ and therefore

$$\hat{S} = Wx = WAs = Zs \qquad \rightarrow \hat{s}_1 = z_{1,1}s_1 + z_{1,2}s_2$$

this means that $\hat{s}$ is maximally non-gaussian if Z is a permutation matrix $\mathbf{P}$, i.e. the identity matrix (potentially with permuted rows).

Measures of non-Gaussianity

Kurtosis

$$kurt(y) = E(y^4) - 3E(y^2)^2$$

which is simply $kurt(y) = E(y^4) - 3$ for whitened variables. Note that $kurt(\mathcal{N}(\mu, \sigma)) = 0$ and furthermore, for independent random variables y_1 and y_2.

$$k(y_1 + y_2) = k(y_1) + k(y_2) \qquad \text{and} \qquad k(ay_1) = a^4 k(y_1)$$

therefore we have:

$$\hat{s} = Wx = WAs = Zs \rightarrow k(\hat{s}) = z_{1,1}^4 k(s_1) + z_{1,2}^4 k(s_2)$$

where we need to impose the constraint $var(\hat{s}_1) = z_{1,1}^2 + z_{1,2}^2 = 1.$ and find the z^* maximizing the absolute value of the kurtosis. One can then show that z is a row of the identity matrix, i.e. one entry is 1 and all other entries are zero.

Finding this z can be implemented using the gradient of the kurtosis wrt. W.

Problem: Kurtosis is non-robust

3 Alternative Measures of Non-Gaussianity

Entropy of a continuous random variable is defined as:

$$H(x) = \int f(x) \log(f(x) dx$$

where $f(x)$ is the density of x. Ideally, one would like to use this directly, but this requires estimation of the density of X. This is even harder than estimating kurtosis reliably.

Negentropy is defined as:

$$J(Y) = H(\nu) - H(Y)$$

where ν is a Gaussian Random variable with the same mean and covariance matrix.

Because the Gaussian is the maximum Entropy distribution for a RV with fixed variance, $J(Y) \geq 0$.

Although $J(Y)$ is difficult to compute, there are ways to efficiently approximate it.

- $\hat{J}(Y)_1 = \frac{1}{2}E(X^3)^2 + \frac{1}{48}k(x)^2$
- $\hat{J}(Y)_2 = \sum_i k_i[EG_i(X) - EG_i(\nu)]^2$

the first one clearly has the problem of being at least as unstable as kurtosis itself but the second one can be good for good choices of the *contrast function* G. and in the simplest case of N=1

$$\hat{J}(Y)_2 = [EG_i(X) - EG_i(\nu)]^2$$

Popular choices (implemented in fastICA) yielding good & robust estimates are

$$G(x) = \frac{1}{a} \log \cosh au \qquad \text{or} \qquad G(x) = -\exp(-\frac{u^2}{2})$$

optimize $\hat{J}(Wx)$ wrt. x yields fast and robust algorithms to minimize gaussianity