A Direct Approach to ICA(I)

Zongchao Liu

10/25/2020

A Direct Approach to ICA

- R package: ProDenICA

By definition, indepent components have a joint product density:

$$f_S(s) = \prod_{p}^{j=1} f_j(s_j)$$

In the next slides, we present an approach that estimates this density directly using generalized additive models.

A Direct Approach to ICA

In the spirit of representing departures from Gaussianity, we represent each f_j as

$$f_j(s_j) = \phi(s_j)e^{g_j(s_j)}$$

a tilted Gaussian density.

- ϕ: standard Gaussian density
- g_j satisfies the normalization conditions required of a density

A Direct Approach to ICA

Assuming X is pre-whitened, the log-likelihood for the observed data $\mathbf{X} = \mathbf{AS}$ is:

$$\ell\left(\mathbf{A}, \{g_j\}_1^p ; \mathbf{X}\right) = \sum_{i=1}^{N} \sum_{j=1}^{p} \left[\log \phi_j\left(a_j^T x_i\right) + g_j\left(a_j^T x_i\right)\right]$$

We wish to maximize the likelihood to the constraints that:

- $\mathbf{A}$ is orthogonal
- g_j results in densities $f_j(s_j) = \phi(s_j) e^{g_j(s_j)}$.

A Direct Approach to ICA

Without imposing any further restrictions on g_j, the model is over-parametrized. Therefore, we would instead maximize a regularized version:

$$\ell\left(\mathbf{A}, \{g_j\}_1^p ; \mathbf{X}\right) = \sum_{i=1}^{N} \sum_{j=1}^{p} \left[\log \phi_j\left(a_j^T x_i\right) + g_j\left(a_j^T x_i\right)\right]$$

$$\sum_{j=1}^{p} \left[\frac{1}{N} \sum_{i=1}^{N} \left[\log \phi\left(a_j^T x_i\right) + g_j\left(a_j^T x_i\right)\right] - \int \phi(t) e^{g_j(t)} dt - \lambda_j \int \left\{g_j'''(t)\right\}^2 (t) dt\right]$$

A Direct Approach to ICA

$$\sum_{j=1}^{p}\left[\frac{1}{N}\sum_{i=1}^{N}\left[\log\phi\left(a_j^T x_i\right)+g_j\left(a_j^T x_i\right)\right]-\int\phi(t)e^{g_j(t)}dt-\lambda_j\int\left\{g_j'''(t)\right\}^2(t)dt\right]$$

- $\int\phi(t)e^{g_j(t)}dt$: enforces the density constraint $\int\phi(t)e^{g_j(t)}dt = 1$ on any solution $\hat{g}_j$
- $\lambda_j\int\left\{g_j'''(t)\right\}^2(t)dt$: a roughness penalty, which guarantees that the solution $\hat{g}_j$ is a quartic-spline with knots at the observed values of $s_{ij} = a_j^T x_i$

A Direct Approach to ICA

$$\sum_{j=1}^{p}\left[\frac{1}{N}\sum_{i=1}^{N}\left[\log\phi\left(a_j^T x_i\right)+g_j\left(a_j^T x_i\right)\right]-\int\phi(t)e^{g_j(t)}dt-\lambda_j\int\left\{g_j'''(t)\right\}^2(t)dt\right]$$

- It can be shown that the solution densities $\hat{f}_j = \phi e^{\hat{g}_j}$ each have mean zero and variance one.

As we increases λ_j, the solutions approach the standard Gaussian ϕ

Steps for the Product Density ICA Algorithm

- ProDenICA

$$\sum_{j=1}^{p}\left[\frac{1}{N}\sum_{i=1}^{N}\left[\log\phi\left(a_j^T x_i\right)+g_j\left(a_j^T x_i\right)\right]-\int\phi(t)e^{g_j(t)}dt-\lambda_j\int\left\{g_j'''(t)\right\}^2(t)dt\right]$$

We fit the functions g_j and directions a_j by optimizing the objective function in an alternating fashion, as described in the following:

1. Initialize **A**(random Gaussian matrix followed by orthogonalization).
2. Alternate until convergence of **A**:
   (a) Given **A**, optimize the objective function w.r.t. g_j(separately for each j)
   (b) Given g_j, $j = 1, ...p$, perform one step of a fixed point algorithm towards finding the optimal A

Steps for the Product Density ICA Algorithm

- Step 2(a) amounts to a semi-parametric density estimation, which can be solved using a novel application of generalized additive models.
- For convenience, we extract one of the p separate problems:

$$\frac{1}{N}\sum_{i=1}^{N}\left[\log\phi\left(s_i\right)+g\left(s_i\right)\right]-\int\phi(t)e^{g(t)}dt-\lambda\int\left\{g'''(t)\right\}^2(t)dt$$

Steps for the Product Density ICA Algorithm

- For approximation, we construct a fine grid of L values s_ℓ^* in increments Δ covering the observed values s_i, and count the number of s_i in the resulting bins:

$$y_\ell^* = \frac{\# s_i \in (s_\ell^* - \Delta/2, s_\ell^* + \Delta/2)}{N}$$

- Typically, we pick L to be 1000, which is more than adequate

Steps for the Product Density ICA Algorithm

We can then approximate the objctive function by:

$$\sum_{\ell=1}^{L} \left\{ y_i^* \left[\log \left(\phi \left(s_\ell^* \right) \right) + g \left(s_\ell^* \right) \right] - \Delta \phi \left(s_\ell^* \right) e^{g\left(s_\ell^* \right)} \right\} - \lambda \int g'''^2(s) ds$$

- can be seen to be proportional to a penalized Poisson log-likelihood with response y_ℓ^* / Δ and penalty parameter λ / Δ, and mean $\mu(s) = \phi(s) e^{g(s)}$
- a generalized additive spline model with an offset term $log \phi(s)$ (Hastie and Tibshirani, 1990; Efron and Tibshirani, 1996)
- can be fit using a Newton algorithm