Kernel Methods for Independent Component Analysis

Independent component analysis (ICA) corresponds to a class of methods with the objective of recovering underlying latent factors present on the data. The observed variables are linear mixtures of the factors which are assumed to be mutually independent [3, 4, 6, 7]. There is a wide range of applications of ICA such as exploratory data analysis, signal processing, feature extraction, noise reduction and blind source separation. A well-known application is the "cocktail party problem" where several signals are mixed together (people talking in a room simultaneously, background music, noise) and the objective is to obtain the original source signals using only the mixtures.

Kernel methods (e.g. support vector machines) are a class of learning techniques that make use of the kernel trick to construct nonlinear algorithms in an efficient way. These methods are known to generalize in high dimensional spaces and have been successfully applied in classification, regression, feature extraction, etc. A new class of kernel methods for ICA have been shown to produce better results than classical ICA algorithms in terms of robustness to noise and performance. Some kernel-based functions for ICA are the constrained covariance (COCO) [5], the kernel generalized variance (KGV) [2], the kernel mutual information (KMI) [5], the kernel canonical correlation (KCC) [2] and the kernel regularized correlation (KRC) [1].

![Image](image-url)

Figure 1: ICA applied on mixed images. First row: Original images. Second row: Observed mixtures. Third row: Estimated independent components using the KRC [1].

The objective of this thesis is to apply several kernel methods for ICA to mixed images and gene expression data emphasizing on key issues such as scalability to large-scale data, robustness to noise and computational complexity. Several public image datasets and gene expression data can be found on the internet.

- **Requirements**: Basic knowledge of linear algebra (e.g. eigenvalue decomposition) and Matlab.
Type of work: Literature: 30%
Implementation: 40%
Experiments and results: 30%

Daily advisors:
Carlos Alzate, <carlos.alzate@esat.kuleuven.be>

Promotors: Johan Suykens

Number of students: 1.

References