

Independent Component Analysis

Erkki Oja
Helsinki University of Technology,
Neural Networks Research Centre,
P.O.B. 5400, 02015 HUT, Finland

Independent Component Analysis (ICA) is a computational technique for revealing hidden factors that underlie sets of measurements or signals. ICA assumes a statistical model whereby the observed multivariate data, typically given as a large database of samples, are assumed to be linear or nonlinear mixtures of some unknown latent variables. The mixing coefficients are also unknown. The latent variables are nongaussian and mutually independent, and they are called the independent components of the observed data. By ICA, these independent components, also called sources or factors, can be found. Thus ICA can be seen as an extension to Principal Component Analysis and Factor Analysis. ICA is a much richer technique, however, capable of finding the sources when these classical methods fail completely.

In many cases, the measurements are given as a set of parallel signals or time series. Typical examples are mixtures of simultaneous sounds or human voices that have been picked up by several microphones, brain signal measurements from multiple EEG sensors, several radio signals arriving at a portable phone, or multiple parallel time series obtained from some industrial process. The term blind source separation is used to characterize this problem.

The lecture will first cover the basic idea of demixing in the case of a linear mixing model and then take a look at the recent nonlinear demixing approaches.

Although ICA was originally developed for digital signal processing applications, it has recently been found that it may be a powerful tool for analyzing text document data as well, if the documents are presented in a suitable numerical form. A case study on analyzing dynamically evolving text is covered in the talk.