Universidad Politécnica de Yucatán

Teacher´s name:
Víctor Alejandro Ortiz Santiago

Subject: Machine Learning

Team Members:
Jose de la Cruz Pat Ramirez
Ángel Ariel Huerta Diaz
Rafael Marí Reyna
Hernando Enrique Te Bencomo

Date: 15/09/2023

Dimensionality reduction; independent component analysis

Independent Component Analysis (ICA) is a computational technique used to separate a multivariate signal into independent non-Gaussian signals. In simpler terms, it is a method to find independent sources given only their mixtures. It is often cited in the context of the "cocktail party problem," where several people are speaking simultaneously, and you want to separate their voices from a set of overlapping recordings.

The aim of ICA is to extract useful information or source signals from data (a set of measured mixture signals). This data can be in the form of images, stock markets, or sounds. Hence, ICA was used for extracting source signals in many applications such as medical signals, biological assays, and audio signals. ICA is also considered as a dimensionality reduction algorithm when ICA can delete or retain a single source. This is also called filtering operation, where some signals can be filtered or removed. ICA assumes that the observed data is a linear combination of independent, non-Gaussian signals. The goal of ICA is to find a linear transformation of the data that results in a set of independent components.

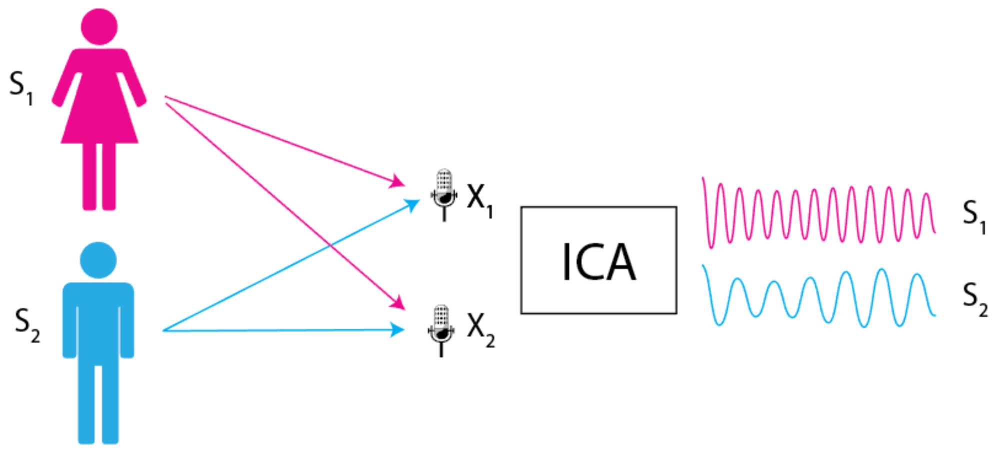
The term non-Gaussian data refers to response variables that are assumed not to have a normal distribution. In this chapter, the “more technically correct” Gaussian distribution is used instead of the more common term “normal distribution.” This is to dispel the bad impression that non-Gaussian data is somehow abnormal.

The basic idea behind ICA is to identify a set of basic functions that can be used to represent the observed data. These basis functions are chosen to be statistically independent and non-Gaussian. Once these basis functions are identified, they can be used to separate the observed data into its independent components. It is often used in conjunction with other machine learning techniques, such as clustering and classification. For example, ICA can be used to pre-process data before performing clustering or classification, or it can be used to extract features that are then used in these tasks.

ICA has some limitations, including the assumption that the underlying sources are non-Gaussian and that they are mixed linearly. Additionally, ICA can be computationally expensive and can suffer from convergence issues if the data is not properly preprocessed. Despite these limitations, ICA remains a powerful and widely used technique in machine learning and signal processing.

Consider a simple example where two audio signals (for example, one of a piano and the other of a guitar) are played simultaneously and recorded using two microphones. Each microphone captures a mixture of both sounds. ICA can be employed to separate these two mixed signals back into the original piano and guitar sounds.

The simplest way to understand the ICA technique and its applications is to explain one problem called the “cocktail party problem.” In its simplest form, let us imagine that two people have a conversation at a cocktail party. Let us assume that there are two microphones near both people. Microphones record both people as they are talking but at different volumes because of the distance between them. In addition to that, microphones record all noise from the crowded party. The question arises, how can we separate two voices from noisy recordings and is it even possible?



One technique that can solve the cocktail party problem is ICA. Independent component analysis (ICA) is a statistical method for separating a multivariate signal into additive subcomponents. It converts a set of vectors into a maximally independent set.

Following the image above, we can define the measured signals as a linear combination:

$$X_i = a_{i1}S_1 + a_{i2}S_2 = \sum_j a_{ij}S_j$$

Where Sy are independent components or sources and aij are some weights. Similarly, we can express sources S1 as a linear combination of signals X1:

$$S_i = \sum_j w_{ij}X_j$$

Where wij are weights.

Using matrix notation, source signals S would be equal to S = W X where W is a weight matrix, and X are measured signals. Values from X are something that we already have and the goal is to find a matrix W such that source signals S are maximally independent. Maximal independence means that we need to:

- Minimize mutual information between independent components or
- Maximize non-Gaussian between independent components.

In ICA, the goal is to find the unmixing matrix and then project the whitened data onto that matrix for extracting independent signals. This matrix can be estimated using three main approaches of independence, which result in slightly different unmixing matrices. The first is based on the non-Gaussian. This can be measured by some measures such as negentropy and kurtosis, and the goal of this approach is to find independent components which maximize the non-Gaussian. In the second approach, the ICA goal can be obtained by minimizing the mutual information. Independent components can be also estimated by using maximum likelihood (ML) estimation. All approaches simply search for a rotation or unmixing matrix. Projecting the whitened data onto that rotation matrix extracts independent signals. The preprocessing steps are calculated from the data, but the rotation matrix is approximated numerically through an optimization procedure. Searching for the optimal solution is difficult due to the local minima exists in the objective function. In this section, different approaches are introduced for extracting independent components.

4.1 Measures of non-Gaussian

ICA has many algorithms such as FastICA, projection pursuit, and Infomax . The main goal of these algorithms is to extract independent components by maximizing the non-Gaussian, minimizing the mutual information, or using maximum likelihood (ML) estimation method. However, ICA suffers from a number of problems such as over-complete ICA and under-complete ICA.

References

A. Tharwat, “Independent component analysis: An introduction,” Applied Computing and Informatics, vol. 17, Aug. 2018, doi: https://doi.org/10.1016/j.aci.2018.08.006.

“ML | Independent Component Analysis,” GeeksforGeeks, May 21, 2019. https://www.geeksforgeeks.org/ml-independent-component-analysis/

E. Zvornicanin, “What Is Independent Component Analysis (ICA)?,” Baeldung, May 04, 2023. https://www.baeldung.com/cs/independent-component-analysis

Stroup, W. W. (2018). Analysis of Non-Gaussian data. En ASSA, CSSA and SSSA https://doi.org/10.2134/appliedstatistics.2015.0081.c16

Tharwat, A. (2020). Independent component Analysis: an Introduction. Applied Computing and Informatics, 17(2), 222-249. https://doi.org/10.1016/j.aci.2018.08.006