Universidad Politécnica de Yucatán

Robotics Engineering

Ortiz Victor

Machine Learning

Independent Component Learning

by

Christian Dzul

Ari Quintal

Ricardo Hernandez

Francisco Hernandez

School of Computing Science and Engineering

UPY

Mérida, Yucatán.

September 2023

Dimensionality Reduction

Dimensionality reduction refers to techniques that reduce the number of input variables in a dataset. More input features often make a predictive modeling task more challenging to model, more generally referred to as the curse of dimensionality.

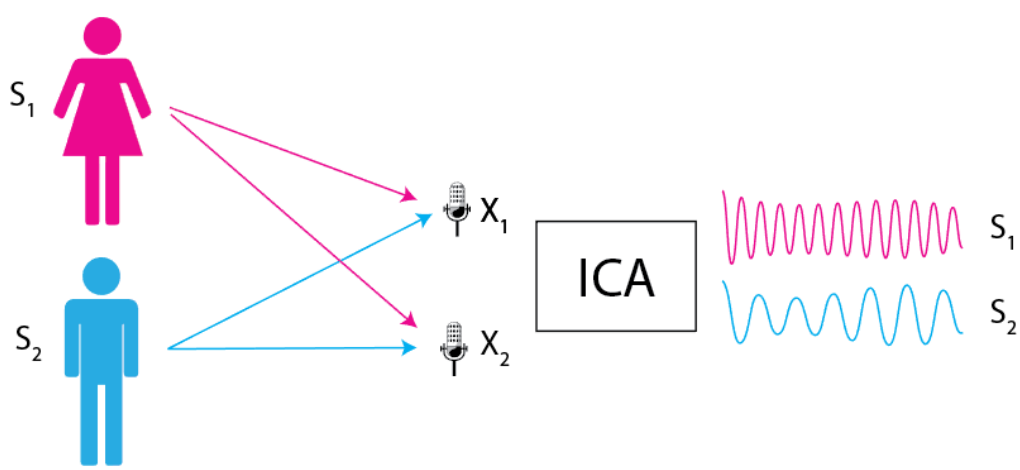
High-dimensionality statistics and dimensionality reduction techniques are often used for data visualization. Nevertheless, these techniques can be used in applied machine learning to simplify a classification or regression dataset in order to better fit a predictive model.

Independent Component Analysis

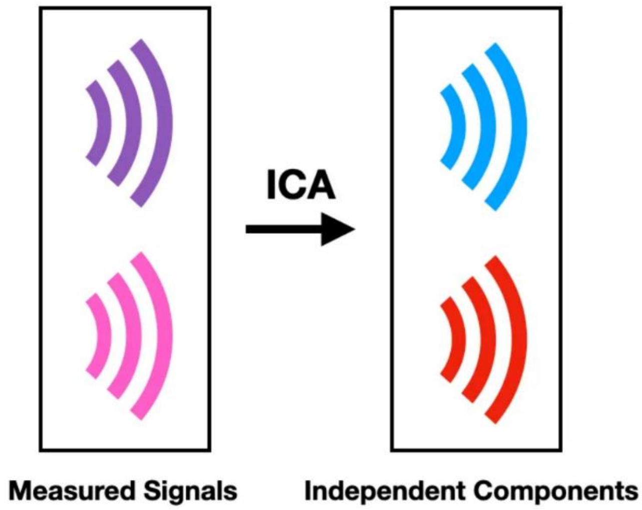
Independent Component Analysis (ICA) is a statistical and computational technique used in machine learning to separate a multivariate signal into its independent non-Gaussian components. It is a powerful technique used for a variety of applications, such as signal processing, image analysis, and data compression. ICA has been used in a wide range of fields, including finance, biology, and neuroscience.

Example:

The standard problem used to describe ICA is the “Cocktail Party Problem”. let’s imagine that two people have a conversation at a cocktail party. Let’s 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 we can separate two voices from noisy recordings and is it even possible?



This problem is solved easily with Independent Component Analysis (ICA) which transforms a set of vectors into a maximally independent set. For this problem, ICA will convert the two mixed audio recordings (represented by purple and pink waveforms) into two unmixed recordings of each individual speaker (represented by blue and red waveforms)

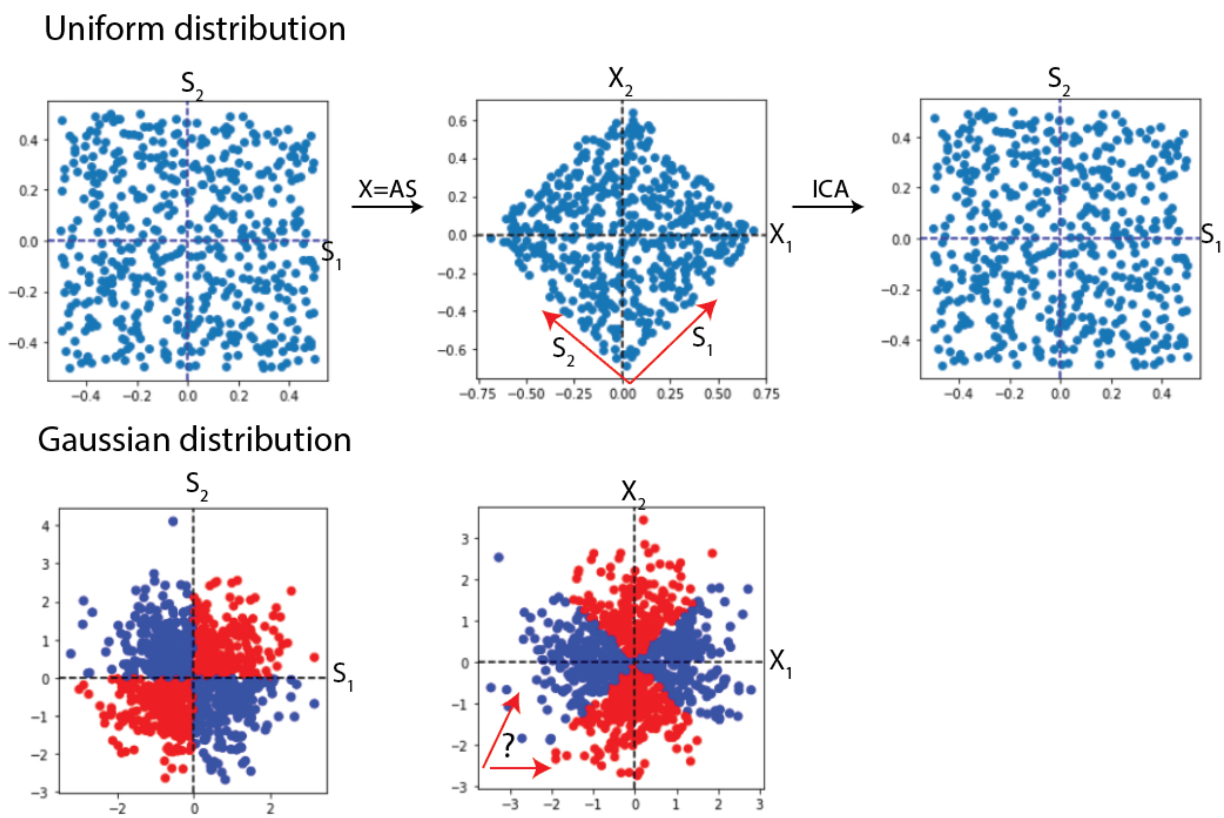


To successfully apply ICA, we need to make three assumptions:

- Each measured signal is a linear combination of the sources.
- The source signals are statistically independent of each other.
- The values in each source signal have non-Gaussian distribution.

Two signals x and y are statistically independent of each other if their joint distribution p(x, y) is equal to the product of their individual probability distributions p(x) and p(y).

From the central limit theorem, a linear combination between two random variables will be more Gaussian than either individual variable. If our source signals are Gaussian, their linear combination will be even more Gaussian. The Gaussian distribution is rotationally symmetric, and we wouldn't have enough information to recover the directions corresponding to original sources. Hence, we need the assumption that the source signal has non-Gaussian distribution:



References

[1]. What is Independent Component Analysis? (2023). Helsinki.fi. https://www.cs.helsinki.fi/u/ahyvarin/whatisica.shtml#:~:text=Independent%20component%20analysis%20(ICA)%20is,a%20large%20database%20of%20samples.

[2]. ML Independent Component Analysis. (2019, May 21). GeeksforGeeks; GeeksforGeeks. https://www.geeksforgeeks.org/ml-independent-component-analysis/

[3]. Talebi, S. (2021, March 17). Independent Component Analysis (ICA) - Towards Data Science. Medium; Towards Data Science. https://towardsdatascience.com/independent-component-analysis-ica-a3eba0ccec35#:~:text=The%20standard%20problem%20used%20to,purple%20and%20pink%20microphones%20above).

[4]. Brownlee, J. (2020, May 5). Introduction to Dimensionality Reduction for Machine Learning - MachineLearningMastery.com. MachineLearningMastery.com. https://machinelearningmastery.com/dimensionality-reduction-for-machine-learning/