Abstract

Independent component Analysis(ICA) is a widely used blind source separation method whose goal is to find a linear representation of nongaussian data so that the components are statistically independent, or as independent as possible. Other than that, it is also used to do feature extraction and signal separation.

In this paper we will first see how the algorithm works to separate the signals into its individual components and then use the algorithm on some actual data to see if it could do the desired work and would try to take some inference from the separated signals.

Introduction

We take the signals/waves that we originally get form the rows of the matrix **X**, and the independent components are **S**. as the signals we receive are some linear combinations of the independent components. We could define a mixing matrix **A**, where

$$\mathbf{X} = \mathbf{AS}$$

Suppose that we have two original signal and formed by the linear combination of two independent signal

$$\mathbf{x}_1 = \mathbf{a}_{11}\mathbf{s}_1 + \mathbf{a}_{12}\mathbf{s}_2$$

$$\mathbf{x}_2 = \mathbf{a}_{21}\mathbf{s}_1 + \mathbf{a}_{22}\mathbf{s}_2$$

Or we could define an unmixing matrix **W** that is basically what we have to calculate

$$\mathbf{S} = \mathbf{WX}$$

or

$$\mathbf{W} = \mathbf{A}^{-1}$$

Algorithm

Pre-processing

It involves two processes one is Centering, and another is Whitening.

Centering - it is a necessary pre-processing of making the mean of x to zero. By subtracting its mean.

$X \Rightarrow X - \text{mean}(X)$

After estimating the mixing matrix A we can regenerate the data by adding mean matrix of S back to the centered estimates of S.The mean matrix of S could be calculated by A^{-1}.mean(X)

Whitening – this process basically means that we will linearly transform a vector x such that the new vector $\overline{x}$ has its components uncorrelated and their variance equal one.

To do this one popular method is eigen value decomposition(EVD) in this the covariance matrix $E[xx^T] = UDU^T$

here U is a orthogonal matrix of eigenvectors of $E[xx^T]$ and D is the diagonal matrix of its eigenvalues, D = diag(d1,d2, ...dn)

now whitening could be done by $\overline{x} = UD^{-1/2} U^T x$

where the $D^{-1/2} = diag(d1^{-1/2}, d2^{-1/2}, \ldots\ldots dn^{-1/2})$

Fastica

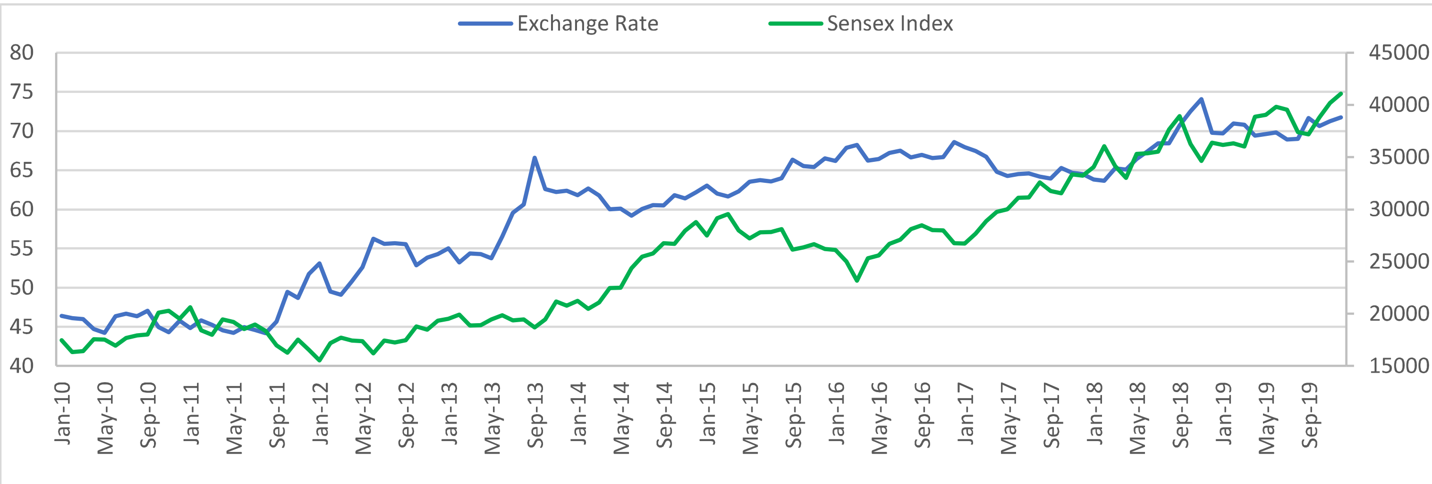
In Fastica algorithm to separate data we iteratively maximise nongaussianity of W^TX by taking $\tanh(w^Tx)$ as the measure of nongaussianity of a separated signal.

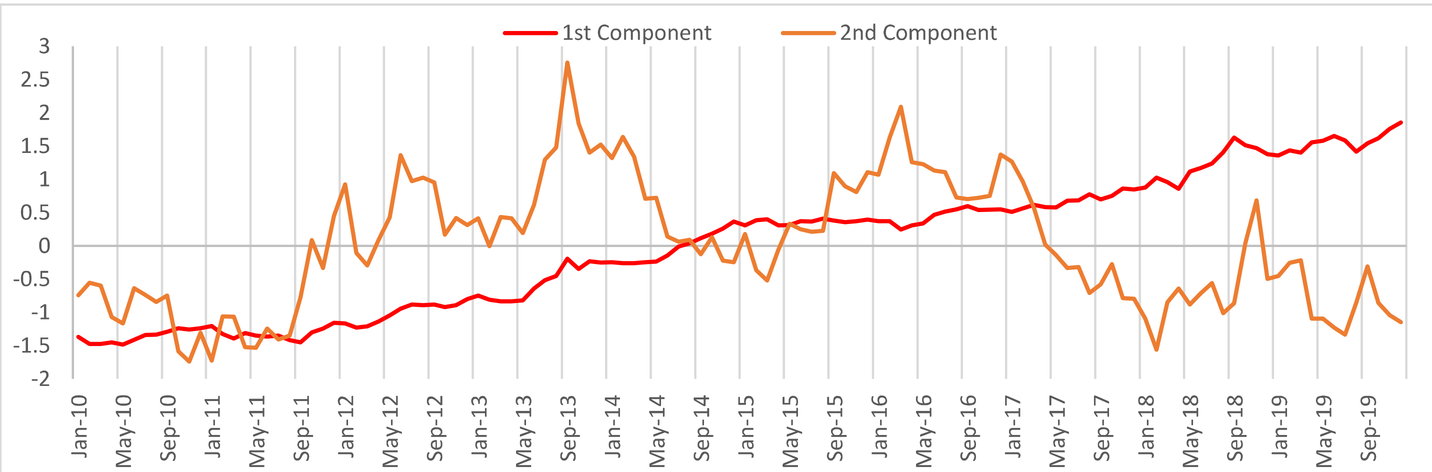
1. Choose and initial random vector weight w
2. $w^+ = E[xg(w^Tx)] – E[g'(w^Tx)]w$
3. $w = w^+/||w^+||$
4. if w does not converge to some finite vector, go back to 2 and repeat the process.

Application

Sensex and Exchange Rate

In this I used Independent component algorithm on the Sensex and Exchange Rate data from 2010 to 2019. Assuming that these signals are formed by 2 independent components we could separate them.



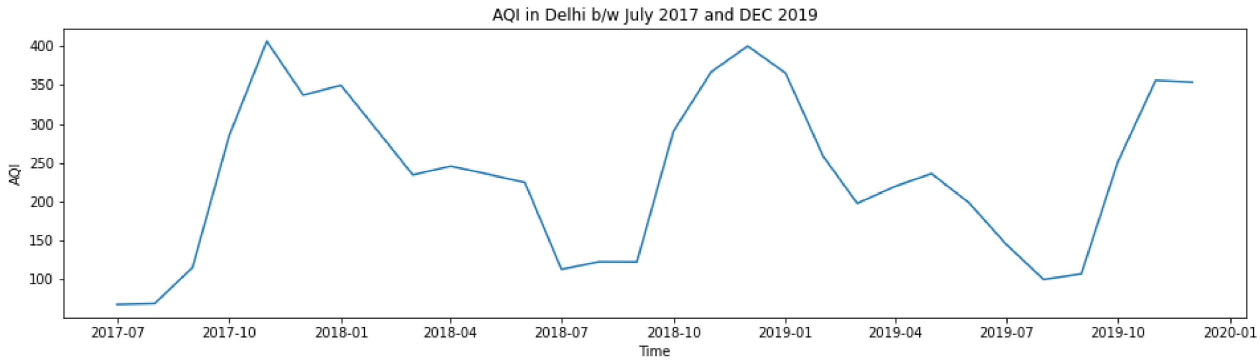


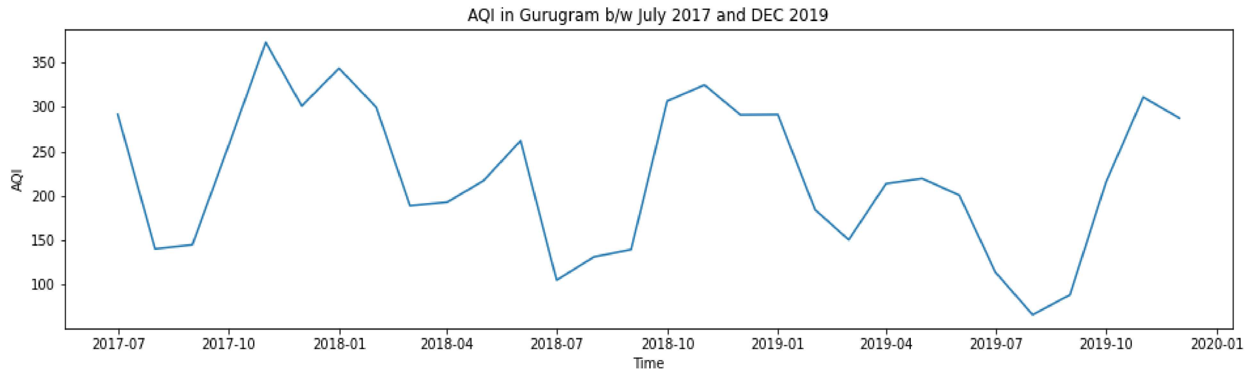
On applying ICA on this data, we get the following independent component. The 1st component could be interpreted as trend variation in exchange rate and Sensex which are growing the throughout years.

The 2nd component could be interpreted as yearly variation in the exchange rate and Sensex. i.e. low in some years and high in other years.

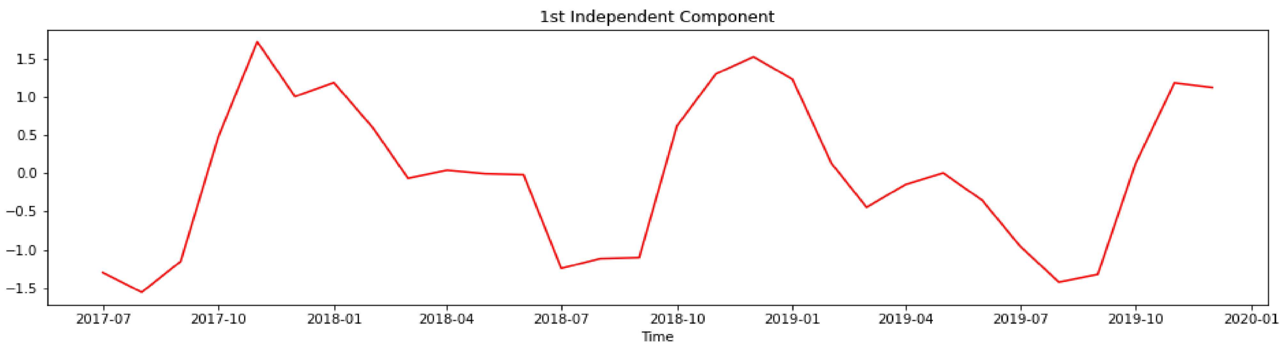
AQI near Delhi

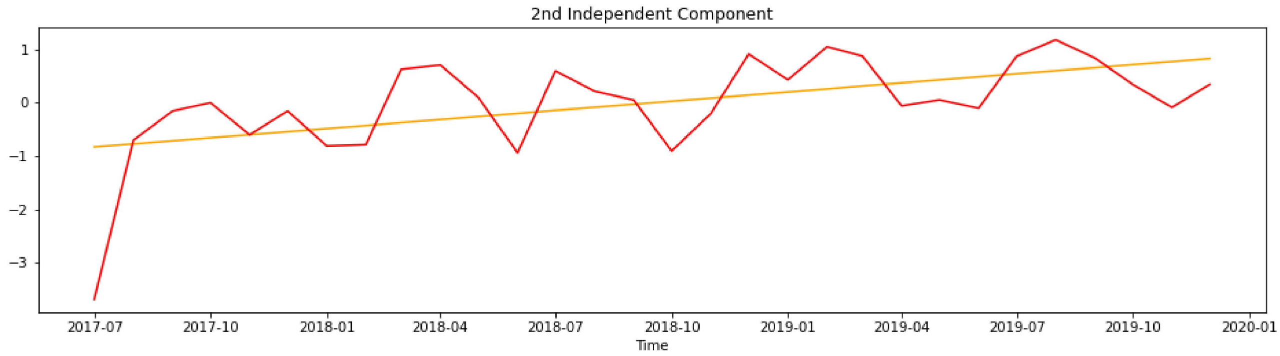
These are the data on the monthly average AQI level from Delhi and Gurugram.





If we apply ICA on this set of data, we get the following independent Components



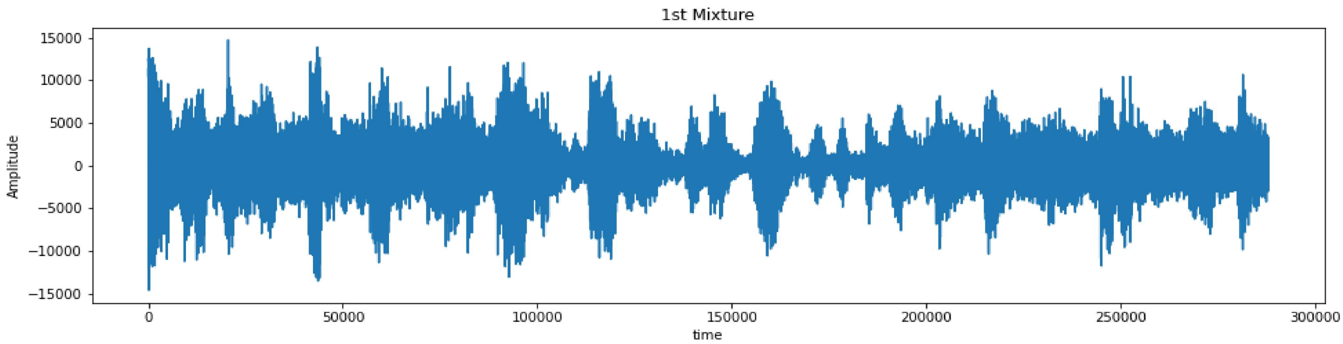


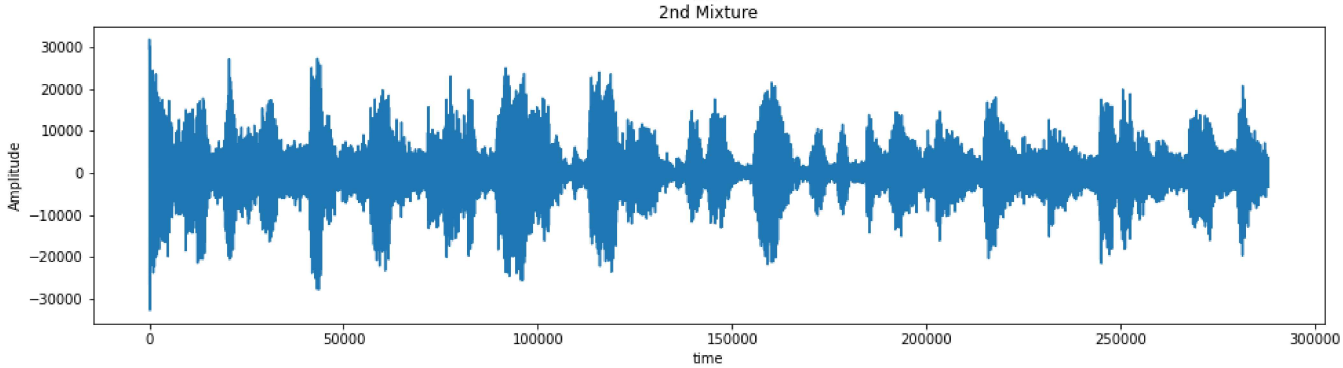
In this case the 1st independent component could be interpreted as seasonal variation in the AQI level i.e less AQI in the months b/w July and September and High AQI in winter months b/w November and January.

The 2nd independent component could be interpreted as the trend variation as the AQI level is increasing or decreasing throughout the years.

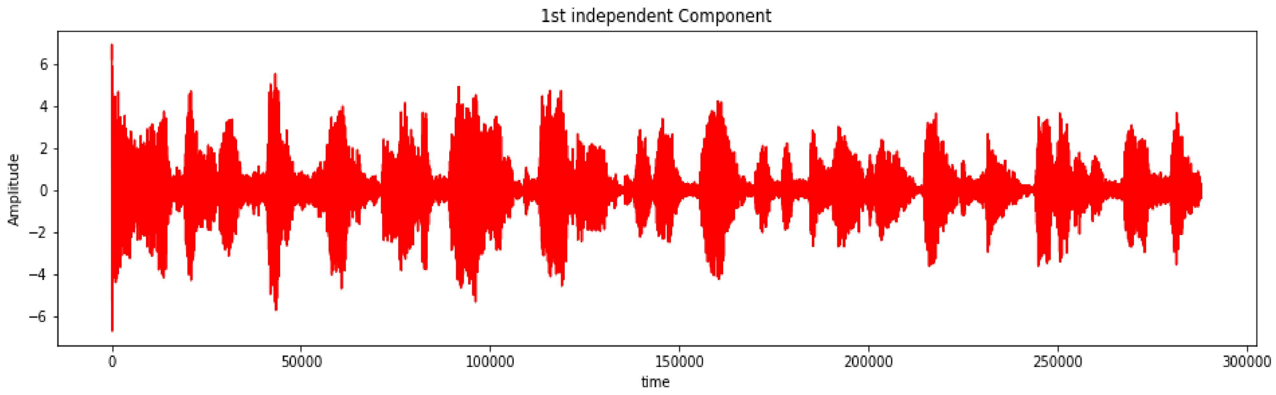
Mixture of Sounds

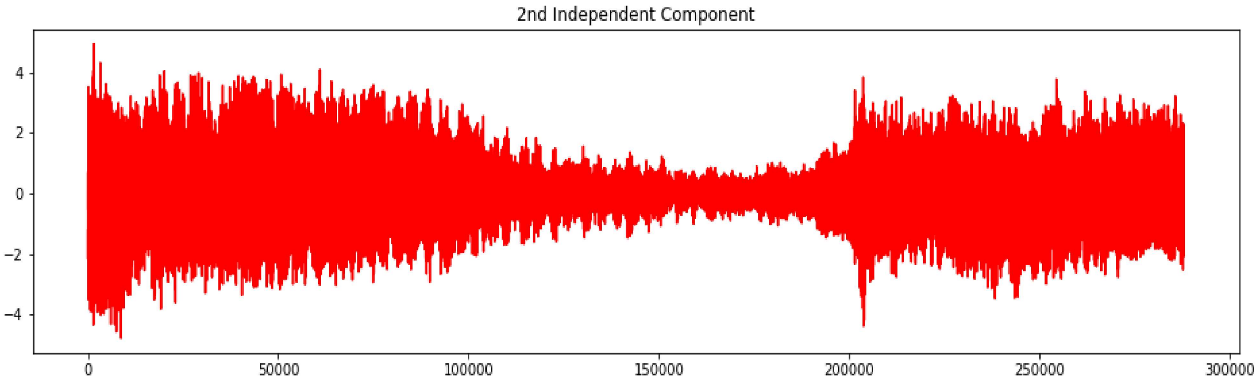
In this ICA is used to separate a mixture of sounds recorded by a pair of mics into its original components. (double click on the icons to play them)





After applying ICA on these set of mixtures, we get the following independent components.





Here the independent components could be interpreted as the original sounds that are spoken be people whose arbitrary combinations are received by the mics as 1^{st} and 2^{nd} mixture.

Limitations

There are some limitations of this algorithm namely

1. If one believes that a signal is formed by m independent components, then to identify them with the number of signals received(n) should be greater equal to m i.e m>n.
2. The signals received should be perfectly synchronised. If not synchronised the result would be completely wrong.
3. It is not necessary that a independent component received after the algorithm could be understood as something in physical world and may come out to be completely random.
4. The sign of the independent component(+ve or -ve) could also not be said with assurance, as it may change upon the changing the random initial value in the iteration process.

References and Data Sources

1. Aapo Hyvärinen and Erkki Oja , Independent Component Analysis: Algorithms and Applications ,Finland Neural Networks, 13(4-5):411-430, 2000
2. Alaa Tharwat, Independent Component Analysis: An Introduction, Applied Computing and informatics, Volume 17 Issue 2, 29/04/21
3. Monthly air Quality Data of Delhi and Gurugram, Air Quality Data in India Kaggle
4. BSE Sensex data b/w 2010 and 2019, Exchange rate Data b/w 2010 and 2019
5. Sound data was self-generated with a news sound and another a music sound.