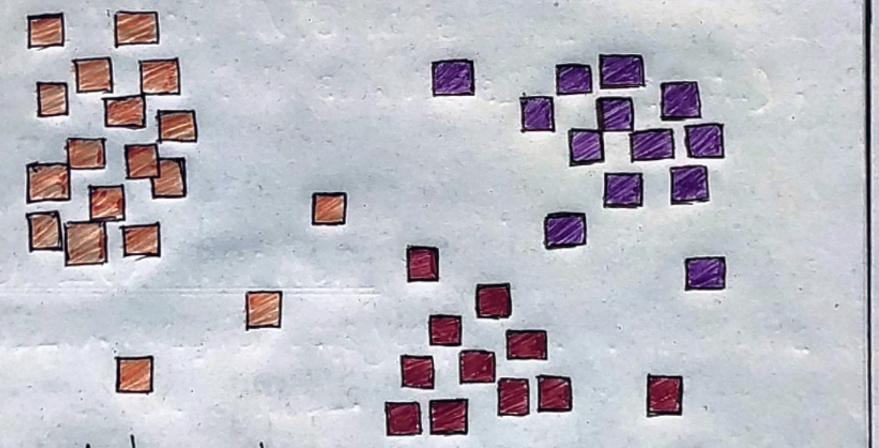
Multivariate Analysis

Multivariate analysis is used to study more complex sets of data than what univariate analysis methods can handle.

Multivariate analysis can reduce the likelihood of Type I errors. Sometimes, univariate analysis is preferrd as multivariate techniques can result in difficulty interpreting the results of the test. For example, group differences on a linear combination of dependent variables in MANOVA can be unclear.



Cluster analysis showing three groups.

There are more than 20 different ways to perform multivariate analysis. Below are the types of multivariate analysis :

- Additive Tree.
- Canonical Correlation Analysis.
- Cluster Analysis.
- Correspondence Analysis/Multiple Correspondence Analysis.
- Factor Analysis.

.......Cont'd

- Generalized Procrustean Analysis.
- Homogeneity of Covariance.
- Independent Component Analysis.
- MANOVA.
- Multidimensional scaling.
- Multiple Regression Analysis.
- Partial Least Square Regression.
- Principal Component Analysis/Regression/PARAFAC.
- Redundancy Analysis.

For example, if you have a single data set you have several choices.

- Additive trees, multidimensional scaling, cluster analysis are appropriate for when the rows and columns in your data table represent the same units and the measure is either a similarity or a distance.

- Principal Component analysis (PCA) decompose a data table with correlated measures into a new set of uncorrelated measures.

- Correspondence analysis is similar to PCA. However, it applies to contingency tables.

Independent Component Analysis

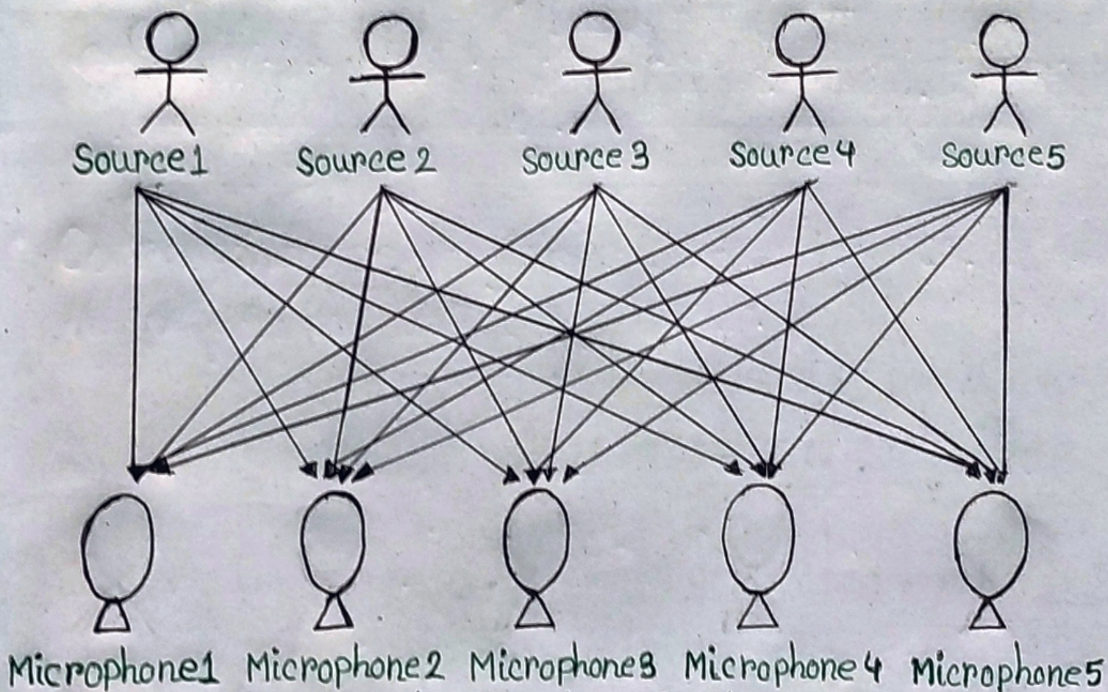
Independent component analysis is used in statistics and signal processing to express a multivariate function by its hidden factors or subcomponents. These component signals are independent non-Gaussian signals, and the intention is that these independent sub-components accurately represent the composite signal.

Problem: To extract independent sources' signals from a mixed signal composed of the signals from those sources.

Given: Mixed signal from five different independent sources.

Aim: To decompose the mixed signal into independent sources:

- Source 1, • Source 2, • Source 3, • Source 4
- Source 5



Cocktail Party Problem

Here, There is a party going into a room full of people. There is 'n' number of speakers in that room and they are speaking simultaneously at the party. In the same room, there are also 'n' microphones placed at different distances from the speakers which are recording 'n' speakers' voice signals. Hence, the number of speakers is equal to the number must of microphones in the room. Now, using these microphones' recordings, we want to separate all the 'n' speakers' voice signals in the room given each microphone recorded the voice signals coming from each speaker of different intensity due to the difference in distances between them. Decomposing the mixed signal of each microphones' recording into an independent source's speech signal can be done by using the machine learning technique, independent component analysis. [X1, X2, ..., Xn] => [Y1, Y2, ..., Yn] where, X1, X2, ..., Xn are the original signals present in the mixed signal and Y1, Y2, ..., Yn are the new features and are independent components which are independent of each other.

Assumptions

1. The independent components generated by the ICA are assumed to be statistically independent of each other.

2. The independent components generated by the ICA must have non-gaussian distribution.

3. The number of independent components generated by the ICA is equal to the number of observed mixtures.

Merits

1. Non-Gaussianity: ICA assumes that the source signals are non-Gaussian, which makes it well-suited for separating signals that are not easily separable by other methods, such as linear regression or PCA.

2. Blind Source Separation: ICA is capable of separating signals without any prior knowledge about the sources or their relationships. This is useful in many applications where the sources are unknown, such as in speech separation or EEG signal analysis.

3. Computationally Efficient: ICA algorithms are computationally efficient and can be applied to large datasets.

4. Interpretability: ICA provides an interpretable representation of the data, where each component represents a single source signal. This can help in understanding the underlying structure of the data and in making informed decisions about the data.

Demerits

1. Non-uniqueness: There is no unique solution to the ICA problem, and the estimated independent components may not match the true sources. This can lead to suboptimal results or incorrect interpretations.

2. Non-deterministic : Some ICA algorithms are non-deterministic, meaning that they can produce different results each time they are run on the same data.

3. Limitations of Gaussianity : If the source signals are not non-Gaussian, then ICA may not perform well, and other methods such as PCA or linear regression may be more appropriate.

Principal Component Analysis	Independent Component Analysis
It reduces the dimensions to avoid the problem of overfitting.	It decomposes the mixed signal into its independent sources' signals.
It deals with the principal components.	It deals with the independent components.
It focuses on maximizing the variance.	It doesn't focus on the issue of variance among the data points.
It focuses on the mutual orthogonality property of the principal components.	It doesn't focus on the mutual orthogonality of the components.
It doesn't focus on the mutual independence of the components.	It focuses on the mutual independence of the components.