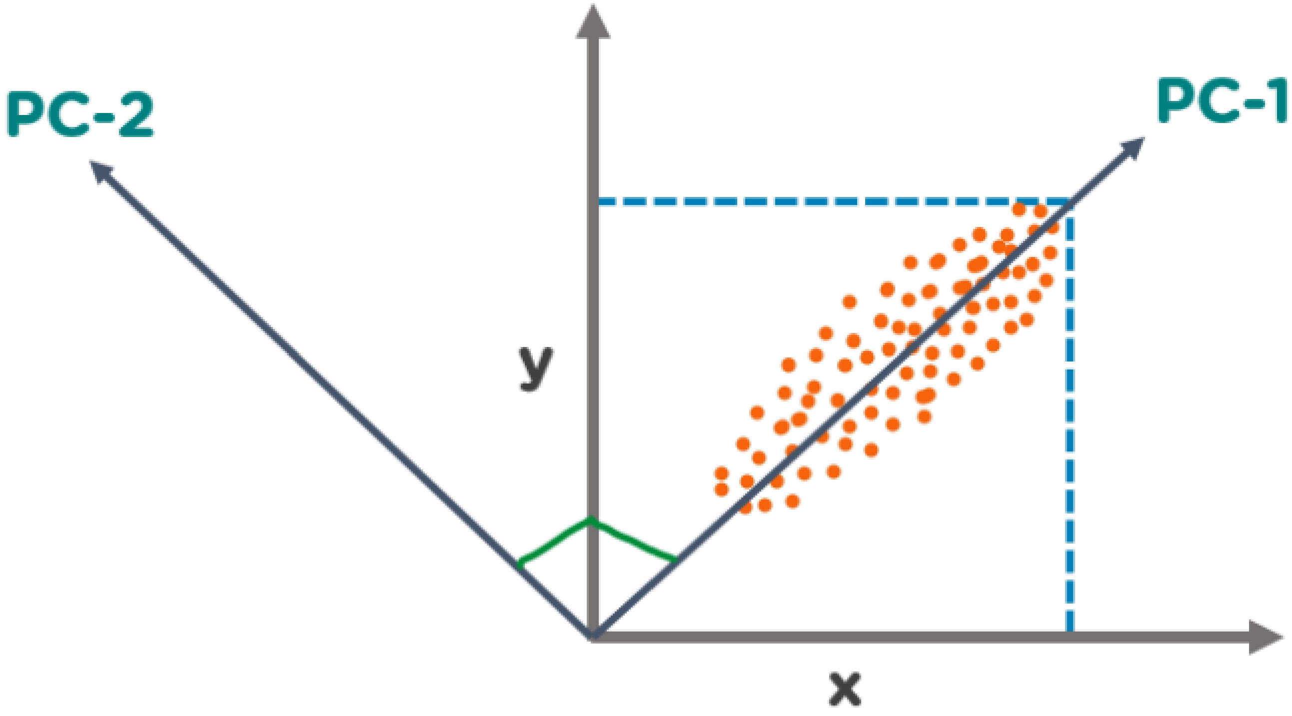
Curves and Surfaces – Independent Component Analysis

Principal Component Analysis

The Principal Component Analysis is a popular unsupervised learning technique for reducing the dimensionality of data. It increases interpretability yet, at the same time, it minimizes information loss. It helps to find the most significant features in a dataset and makes the data easy for plotting in 2D and 3D. PCA helps in finding a sequence of linear combinations of variables.



Principal Component

The Principal Components are a straight line that captures most of the variance of the data. They have a direction and magnitude. Principal components are orthogonal projections (perpendicular) of data onto lower-dimensional space.

Principal curves

Principal curves are smooth one-dimensional curves that pass through the middle of a p-dimensional data set. They minimize the distance from the points and provide a non-linear summary of the data. The curves are non-parametric and their shape is suggested by the data.

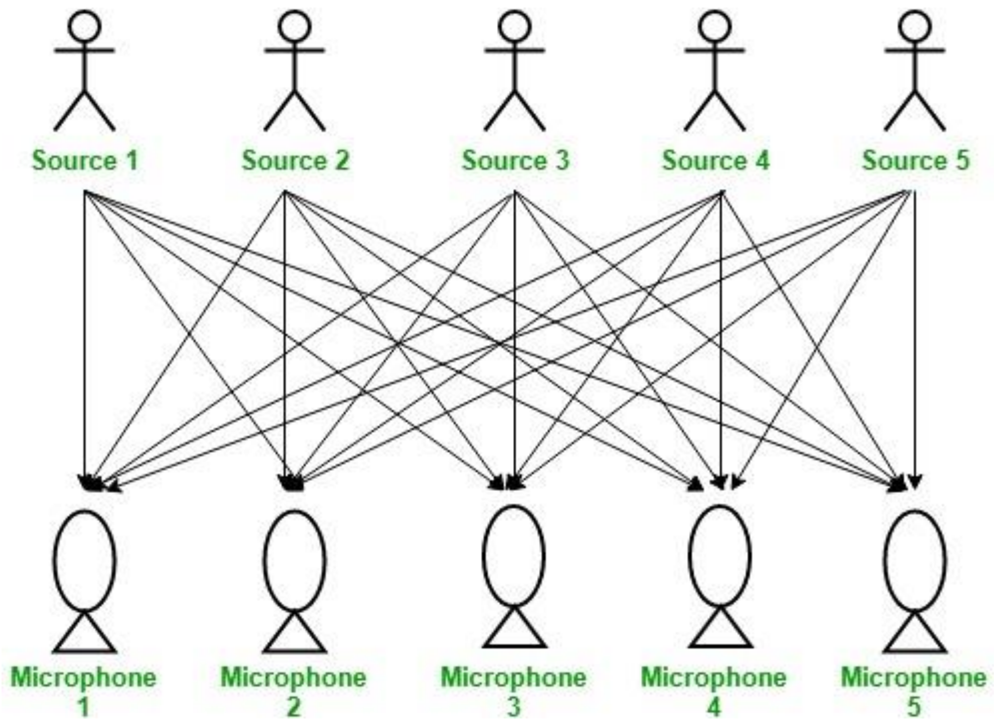
principal surfaces

- principal surfaces are two-dimensional surfaces that pass through the middle of the data. The curves and surfaces are found using an iterative procedure which starts with a liner summary such as the usual principal component line or plate.
- Each successive iteration is a smooth or local average of the-dimensional points, where local is based on the projections of the points onto the curve or surface of the previous iteration.
- Several linear techniques, such as factor analysis and errors in variables regression, end up using the principal components as their estimates after a suitable scaling of the coordinates.

Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is a machine learning technique to separate independent sources from a mixed signal. Unlike principal component analysis which focuses on maximizing the variance of the data points, the independent component analysis focuses on independence, i.e. independent components.

Consider *Cocktail Party Problem* or *Blind Source Separation* problem to understand the problem which is solved by independent component analysis.



Here, There is a party going into a room full of people. There is an ‘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 microphone’s 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, …, and Xn are the original signals present in the mixed signal and Y1, Y2, …, and Yn are the new features and are independent components which are independent of each other.

Restrictions on ICA

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.