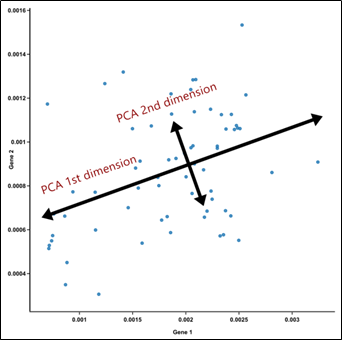
* *Principal Component Analysis*

When training different machine learning models, we come across many high dimensional datasets. In order to train models efficiently with such types of datasets, the dimensions should be reduced. Principal Component Analysis is one of the most widely used techniques for this purpose. One main constraint to dimensionality reduction is that, while reducing the dimensions – the variability or the statistical information contained by the dataset should be preserved. While reducing the dimensions, sometimes a small trade-off is done with accuracy in order to maintain simplicity. There are some practices that should be followed while performing PCA. Data Standardization plays a crucial role before doing PCA. If it’s not done, PCA would consider the larger values to be dominant over the small values. While doing dimensionality reduction, one has to make sure that the contribution of the variables is not compromised. Standardizing data prevents the result to be biased. The covariance matrix is computed in order to remove redundant data as sometimes the variables are strongly correlated. Computation of Eigenvectors and Eigenvalues: Eigenvectors represent the direction of axes where most of the information or variance lies. Each eigenvector has an eigenvalue which is the amount of variance. Now, we have the corresponding Eigenvectors and their eigenvalues, principal components are ready to be computed.

Principal Component one is computed such that the variance contained by it is maximum in the dataset. Similarly, the second principal component is computed in a way that it has the least correlation with PC1 and the second highest variance. Creation of the Feature Vector: This is one the main steps in the process of principal component analysis as after this we have the components which we decide to keep. i.e., a subset of components is chosen and the ones with less significance are removed. Finally, the data is reoriented using this feature vector and now, the axes are redefined using principal components.

Principal Component analysis reduces the chances of model-overfitting, also increases the algorithm performance and data visualization is made easier since the dataset is low dimensional. Some disadvantages that come along with PCA are: it is important to perform data standardization, the data becomes less interpretable since the data is manipulated in PCA. The below diagram shows a brief depiction of Principal Component Analysis:



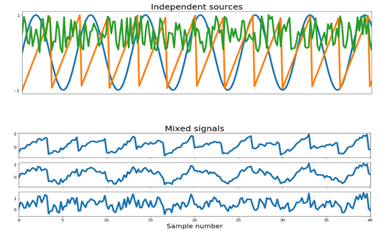
*figure 19 : basic representation of PCA*

* *Independent Component Analysis*

ICA, unlike PCA that focuses on maximizing the data points, is another feature extraction technique that concentrates more on independent components. The main aim is to find linear representation of non-gaussian data in order for the components to be either statistically independent or as independent as possible. It can be said that ICA is somewhere related to Principal Component Analysis, but it is more powerful as it is used as a signal separation and dimensionality reduction method to identify hidden underlying independent features from the mixed signals. ICA can be explained better with an example such as, a person in a party is talking to two of his friends who are standing at different distances from him. Although one of them is closer to him, he can still hear both but the sound signal of both of his friends’ voices reaches him as a mixed signal. His brain further un-mixes these signals and perceives their voices in a way such that the closer one’s voice signal will be louder. Now it can be said that each voice signal is a sine wave with some constant frequency. All the signals combine to form the mixed signal wherein the closer the signal, the more dominant it is in the mixed signal. This can be represented as a vector-matrix notation, where matrix X (n x p) is said to be a product of matrices, A (n x k) and S (k x p) where A is the matrix for acoustic source signals and S represents the mixed signal source.

*X = AS*

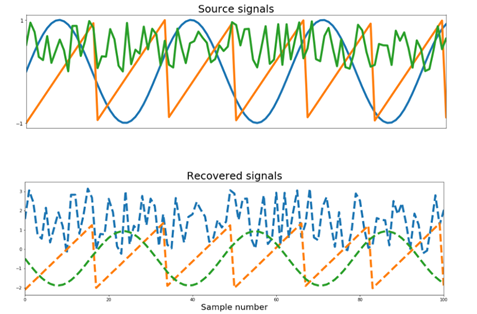
The main objective of ICA is to separate A and S from X with some following preconditions for it to work. The mixed signals should be a linear combination of any number of sources. The extracted features should not be related to each other in any way, i.e., they are independent of each other. The extracted independent components should be non-gaussian.

**

*figure 20 : original signals, independent and mixed*

In this figure we can see the independent signals represented in the first plot graph are found as a part of mixed signals in the second plot graph where one signal can be seen similar to the sine wave and one signal dominating others.

After application of PCA the independent components can look like:

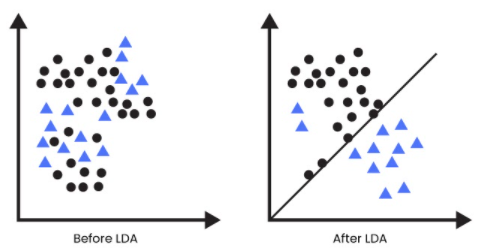


*figure 21 : components after applying PCA*

* *Linear Discriminant Analysis*

LDA is a methodology for reducing dimensionality in supervised classification problems. It's used to represent gaps between groups, such as separating two or more classes. It's a technique for projecting features from a higher-dimensional space to a lower-dimensional one. Image recognition and other marketing applications use this method of dimensionality reduction. This dimensionality reduction technique assists in the exploration of two areas: parameters that can be used to describe the relationship between a group and an entity, and a classification preceptor model that can be used to assist in group separation. LDA is commonly used to model various types of varieties because of this. This method can be used to spread a variable between two or more groups.

LDA can classify objects into binary and multi-class categories. Bayes' Theorem is used in LDA models to estimate probabilities. They make predictions based on the likelihood that a new input dataset would fall into one of the groups. The output class is the one with the highest likelihood, after which the LDA makes a prediction. The prediction is made using Bayes' Theorem, which calculates the likelihood of the output class given the input.



*figure 22 : application of LDA*

LDA is used in a variety of applications. It is applicable to any problem that can be converted into a classification problem. Speed recognition, face recognition etc. LDA is used to reduce the number of facial features before applying the classification method. You can use LDA to collect customer features if you want to identify customers based on their likelihood of purchasing a product. LDA can be used to classify diseases into severe, mild, or moderate categories. There are several patient parameters that will be considered when performing this classification task. After performing the above three dimensionality reduction techniques, a considerable tradeoff is seen in the time taken by the  neural network to train itself and the accuracy. Principal component analysis gives the best result among the three when the number of components are chosen as 17. Hence, the Gender prediction is done using Artificial Neural Networks after applying Principal Component Analysis. This prediction is done on the BVC Age and Gender estimation dataset with an accuracy of 85.51%.