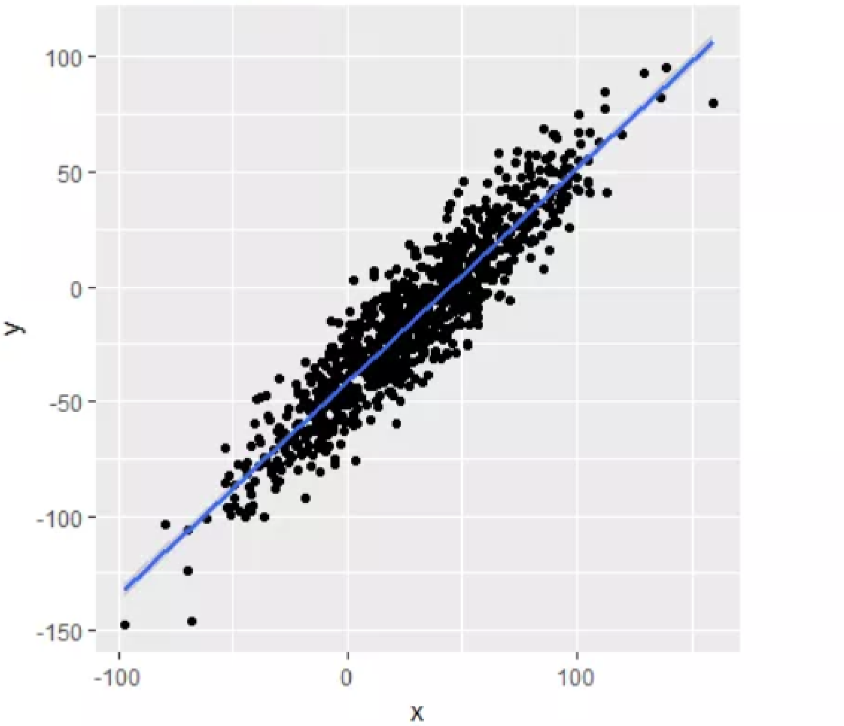
Principal Component Analysis Implementation Issues

Principal Component Analysis (PCA) is a statistical procedure used to convert a set of observations of correlated variables into a set of values of uncorrelated variables called principal components. For instance, Max is building a regression model to predict some component of interest rate trading. Max's model draws upon hundreds of distinct values, and in this high dimensional space it is hard to find sufficient real world data to make accurate predictions. Thus, Max uses PCA to perform dimensionality reduction to group variables into logical collections, combining them all into one factor score. Let us say Max's model has 43 variables. Then, using PCA, his model can be reduced to 6 or 7 variables whilst still being able to explain a significant amount of variance. There is sufficient real world data to populate six dimensions instead of looking highly sparse with 43 variables.

The following are the implementation issues associated with PCA:

- The applicability of PCA is limited by the assumptions made in its derivation. It is limited by the assumption of linearity, on the statistical importance of mean, and that large variances have important dynamics (Luo, 2009). These limitations are explained below.
- PCA uses variance as a measure of importance for a particular dimension. Hence, it always considers the low variance components in the data as noise and recommends us to get rid of these components. However, sometimes those components play a major role in the supervised learning problem. For instance, let us say we have the grades of 100 students for Math and French and Math scores have a higher variance. Then, Math will be a

better discriminating factor since in data analysis the direction along which data varies the most is of high importance.

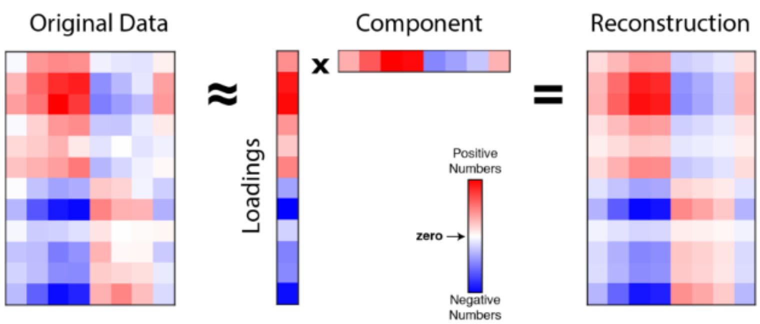


The above plot depicts that most of the variance in the dataset can be explained by projecting points on the line since x and y are strongly correlated i.e. PCA seeks linear combinations of the axis where dispersion among points is maximum

- The mean removal process before constructing the covariance matrix for PCA is another problem. Consider a field like astronomy, where all the signals are non-negative, and the mean removal process forces the mean of some astrophysical exposures to be zero, which consequently creates negative fluxes. An additional step, forward modeling, needs to be performed to recover he true magnitude of signals.
- PCA's main aim is to represent data in lower dimensions by getting rid of redundant features. It does this by finding the

orthogonal principal components. But, the above property is not valid if the joint distribution of data follows a distribution that is not a multivariate normal distribution.

- The results of the PCA depend on the scaling of variables. The PCA results are sensitive to the shape of the data cloud. For instance, if you center the variables, leaving the variances as they are, it is called 'PCA based on covariances'. If you standardize the variables to variances = 1, it is called 'PCA based on correlations', and this is very distinct from the former.



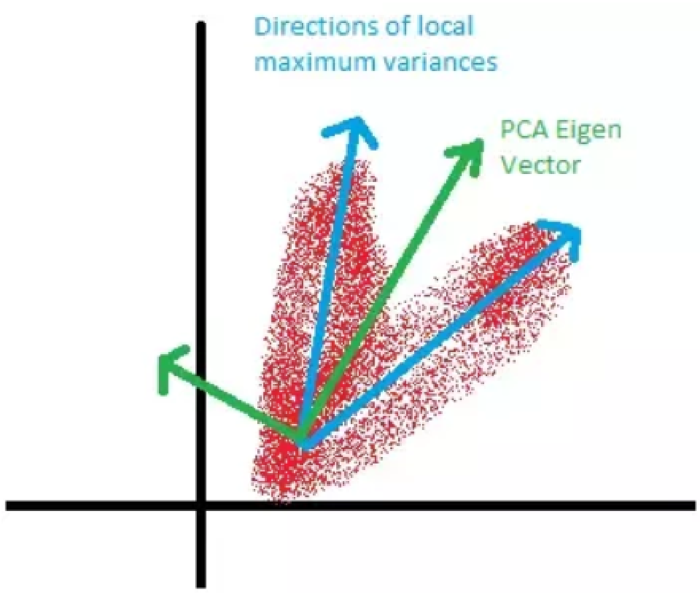
The above figure depicts reconstruction of the original dataset with 1 principal component. Most data are computed with multiple principal components

- PCA cannot be used in automated target recognition (ATR) since its main goal is to minimize mean square error between original and reduced spaces, and not to minimize or maximize any metric related to ATR.

- PCA requires normalization to mitigate the effects of scale, but normalizing results in the spreading the influence across most of the other principal components i.e. more PCs will be needed to explain the same amount of variance in the data.

- PCA always finds linear principal components to represent data in a lower dimension, but sometimes we need non-linear principal components. Performing a PCA computation on non-linear data will result in the results being meaningless. It works optimally only in the situation where correlations are linear.

- PCA has a limitation in the presence of heavy bias in the sampled data. PCA's components explain most of the variation, so if a good amount of the samples is part of the population then it will come up with a component that mostly spans this population.
- PCA always finds orthogonal principal components. Sometimes, our data demands non-orthogonal principal components to represent data. As seen in the figure below, even in presence of the actual maximum variance directions, PCA fails to find those vectors.



The figure above shows the principal components in green and the actual maximum variance in blue

Luo, Q. Advanced Computing, Communication, Control and Management. 2009. Retrieved from: https://books.google.co.in