

Word	Description
Factor Analysis	<p>Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors. Factor analysis aims to find independent latent variables. Factor analysis also assumes several assumptions:</p> <ul style="list-style-type: none"> • There is linear relationship • There is no multicollinearity • It includes relevant variables into analysis • There is true correlation between variables and factors <p>There are different types of methods used to extract the factor from the data set:</p> <ol style="list-style-type: none"> 1. Principal Component Analysis 2. Common factor analysis 3. Image factoring 4. Maximum likelihood method
False Negative	<p>Points which are actually true but are incorrectly predicted as false. For example, if the problem is to predict the loan status. (Y-loan approved, N-loan not approved). False negative in this case will be the samples for which loan was approved but the model predicted the status as not approved.</p>
False Positive	<p>Points which are actually false but are incorrectly predicted as true. For example, if the problem is to predict the loan status. (Y-loan approved, N-loan not approved). False positive in this case will be the samples for which loan was not approved but the model predicted the status as approved.</p>

It is a method to transform features to vector. Without looking up the indices in an associative array, it applies a hash function to the features and uses their hash values as indices directly. Simple example of feature hashing:

Suppose we have three documents:

- John likes to watch movies.
- Mary likes movies too.
- John also likes football.

Now we can convert this to vector using hashing.

Feature Hashing

Term	Index
John	1
likes	2
to	3
watch	4
movies	5
Mary	6
too	7
also	8
football	9

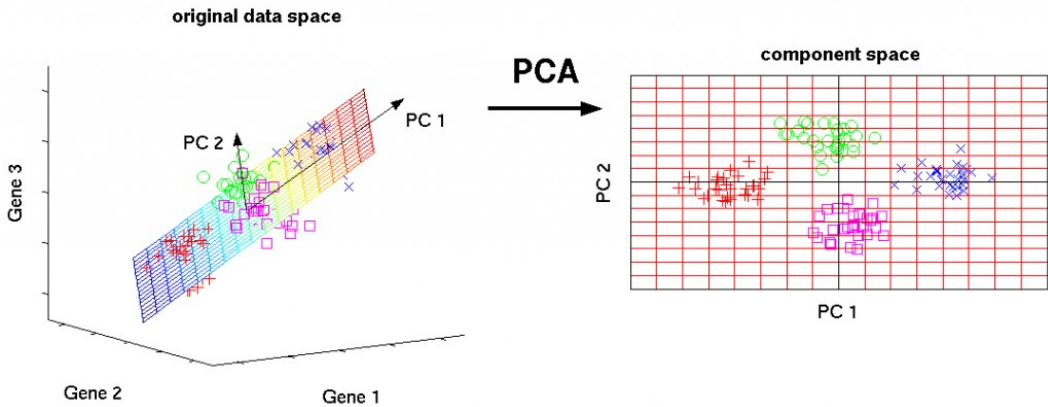
The array form for the same will be:

$$\begin{pmatrix} \text{John} & \text{likes} & \text{to} & \text{watch} & \text{movies} & \text{Mary} & \text{too} & \text{also} & \text{football} \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Feature Reduction

Feature reduction is the process of reducing the number of features to work on a computation intensive task without losing a lot of information.

PCA is one of the most popular feature reduction techniques, where we combine correlated variables to reduce the features.



**Feature
Selection**

Feature Selection is a process of choosing those features which are required to explain the predictive power of a statistical model and dropping out irrelevant features.

This can be done by either filtering out less useful features or by combining features to make a new one.

