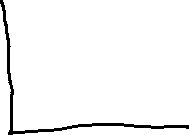
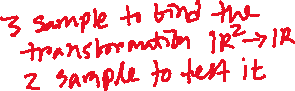
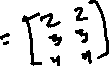
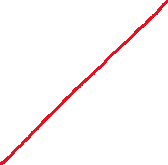
This Document will explain the real mathematical reason to do each step of PCA:

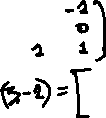
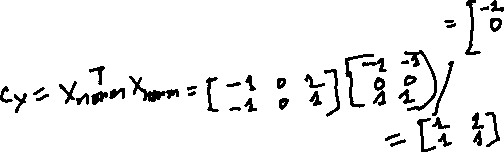
There are 3 data points in a 2D plane can be represented by the coordinate system (x1, x2).



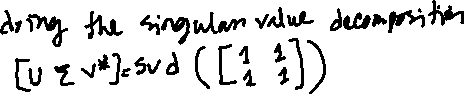
Here, we need two dimensions or two features x1 and x2 to present the data points. Our goal is to develop a technique where we can remove the redundancy between the features. For example, we look the data in the following direction shown in the figure below to represent the data where we will need only one dimension to represent.



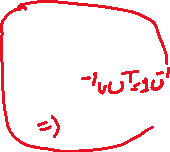
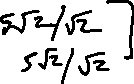
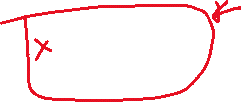
In order to find the redundancy of the data, we need to find the correlation between the features which can be calculated as the following equation:



This matrix has both diagonal and off-diagonal components. Diagonal components are presenting variance and off-diagonal covariance for each data point pairs. Covariance are presenting the correlation between the pairs. But we would like to make the correlation to zero in order to remove the redundancy between x1 and x2. The only way to do this is to diagonalize this covariance matrix. In other way, we convert Cx to eigen value and eigen vectors [ U] where is a diagonal matrix with the variances in descending order.



We have successfully diagonalized Cx. Where, we can see the variance in only first column which is able to capture all the data variation using only the new basis U[:,1] or first principal component



Here is the python code to implement the whole process:

