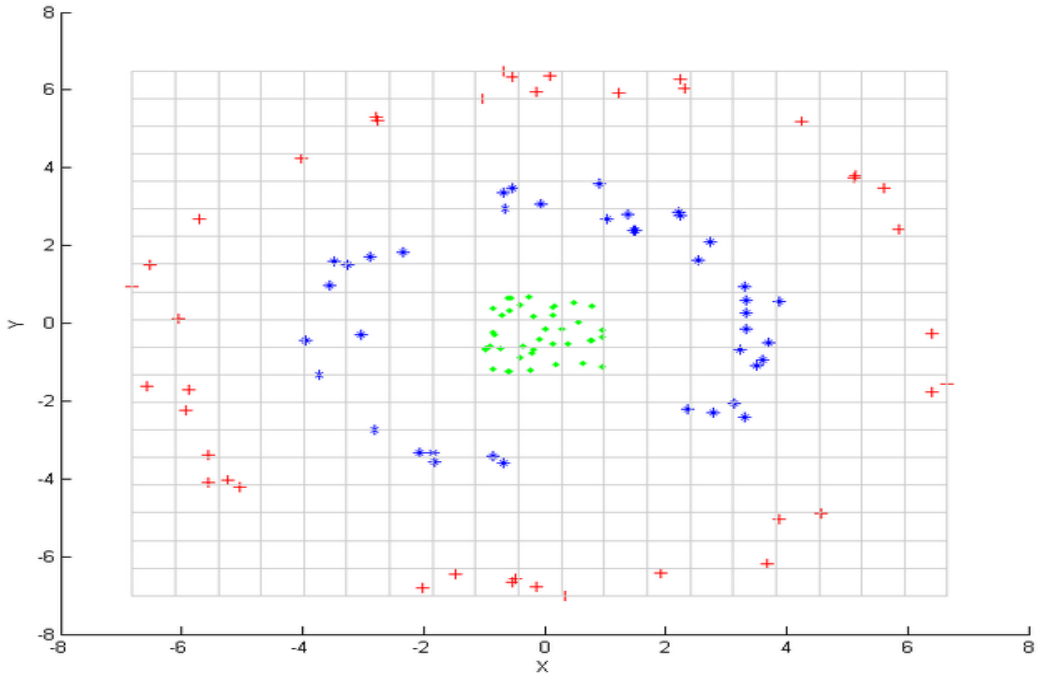
Name: Shrinivas Acharya

Roll No: 10010164

Difference between Linear and Kernelised version:

Kernelised version has extra features that it can separate out even the non-linear data sets. N points cannot in general be linearly separated in d < N dimensions, they can almost always be linearly separated in d >= N dimensions. That is, given N points, X_i, if we map them to an N-dimensional space, then we can always separate them with a line. This is the key idea behind the use of kernels. We construct a hyperplane that divides the points into arbitrary clusters. Of course, this feature mapping creates linearly independent vectors, so there is no covariance on which to perform eigendecomposition explicitly as we would in linear PCA. Instead, in kernel PCA, a non-trivial, arbitrary feature function is 'chosen' that is never calculated explicitly, allowing the possibility to use very high dimensional feature mapping if we never have to actually evaluate the data in that space. Since we generally try to avoid working in the feature-space, which we will call the 'feature space', we can create the N-by-N kernel and go up to N dimensions.

Data Points before we apply the Kernel Function:



Data Points after we apply the kernel function:

Output after kernel PCA with. The three groups are distinguishable using the first component only.

$$k(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} + 1)^2$$