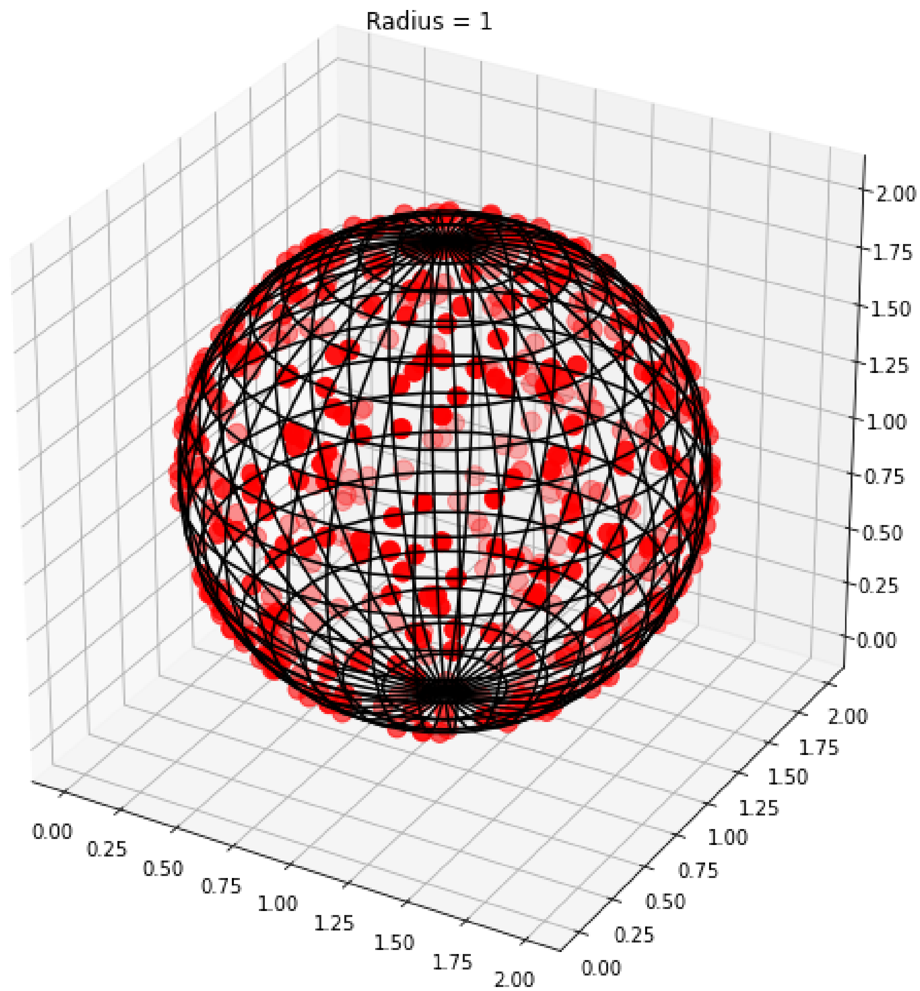
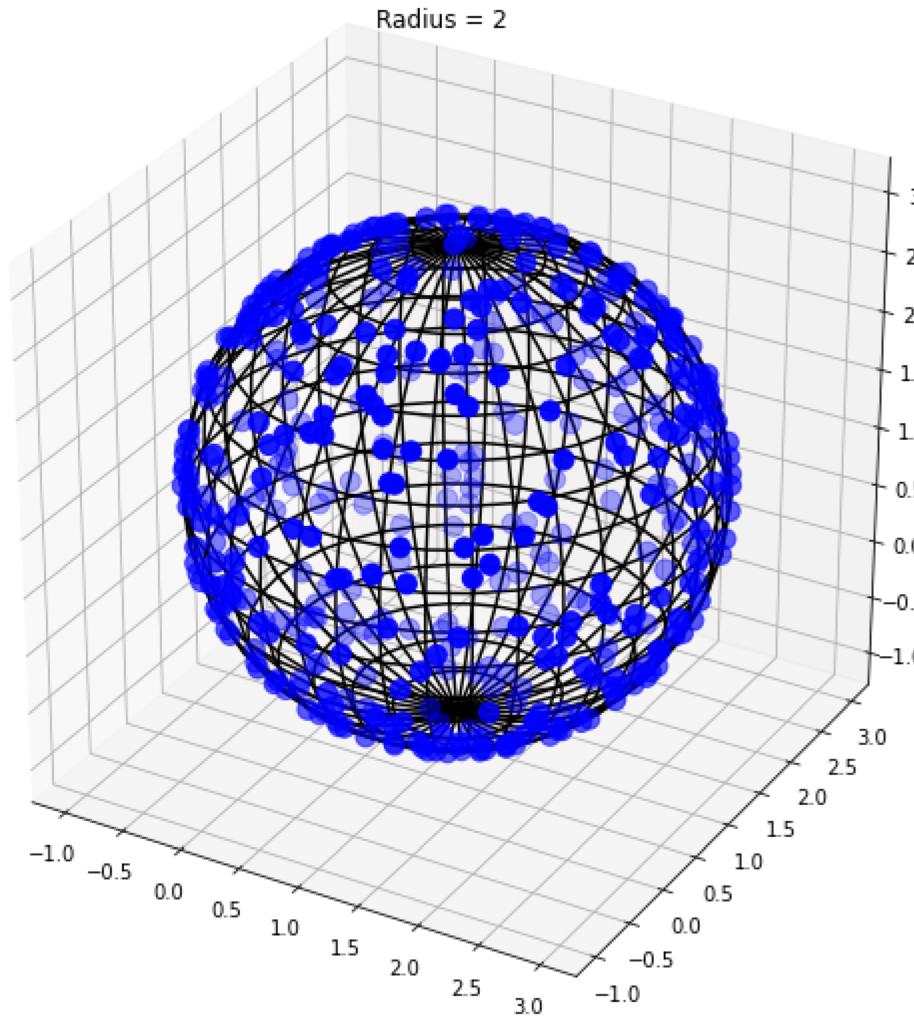
Sampling of the Data

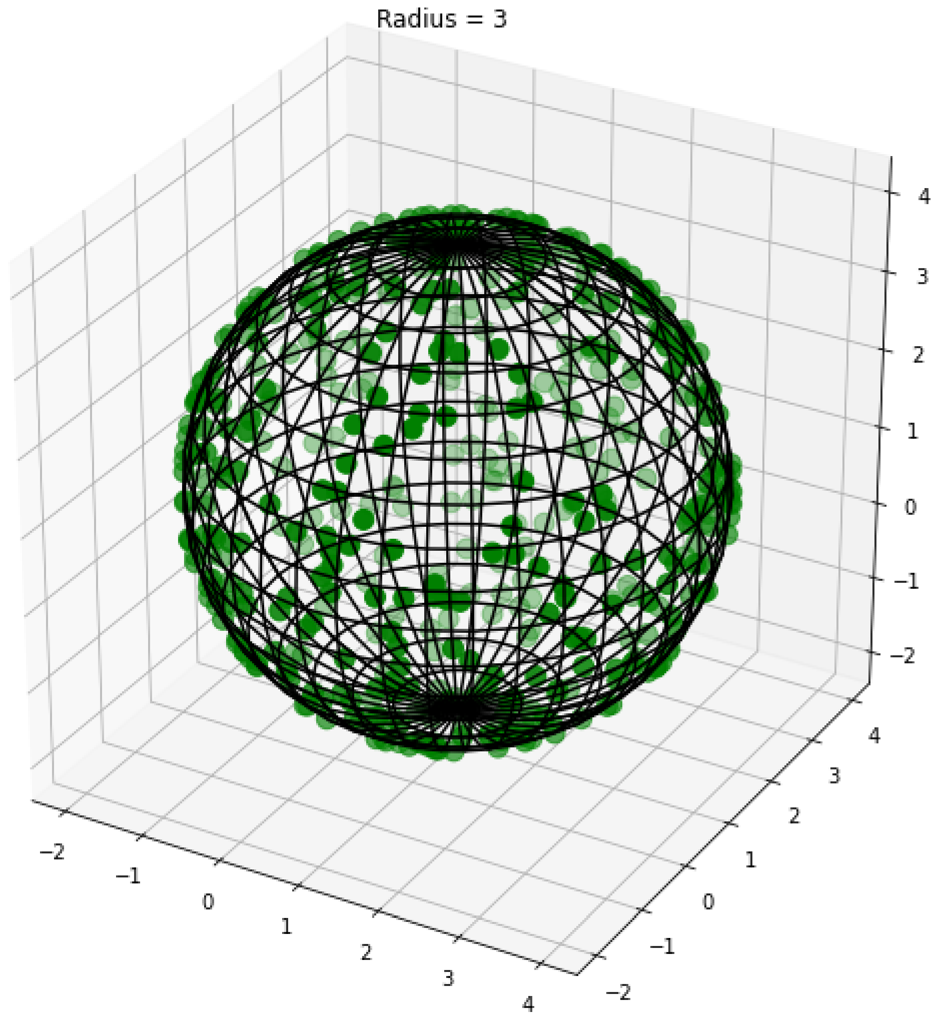
I have sampled the 3 dimensional point from a standard normal distribution. Then shifting the origin to 1,1,1 and modifying the vector so that it satisfies the following constraints

$$(x-1)^2 + (y-1)^2 + (z-1)^2 = r^2 \text{ where } r \in \{1,2,3\}$$

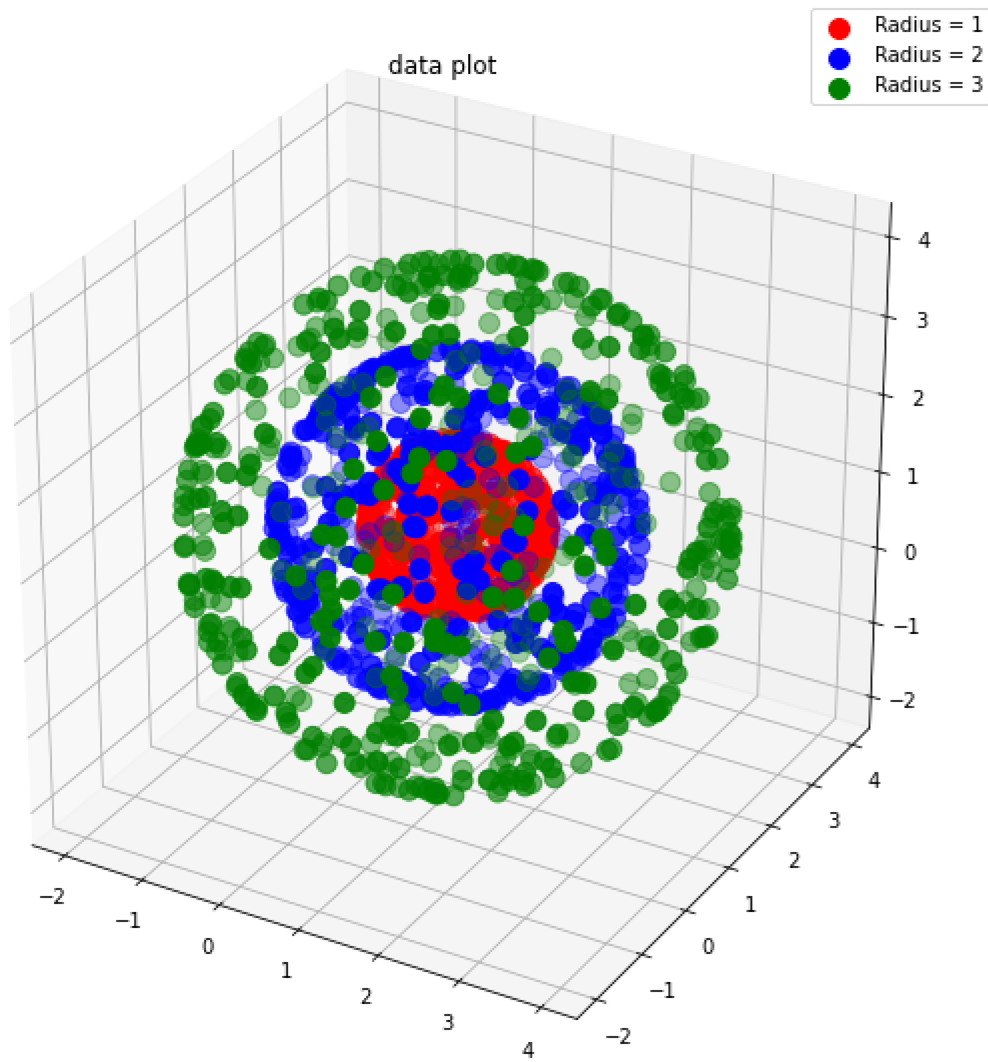
Total number of data = 1500 (500 each)




Radius = 2
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0

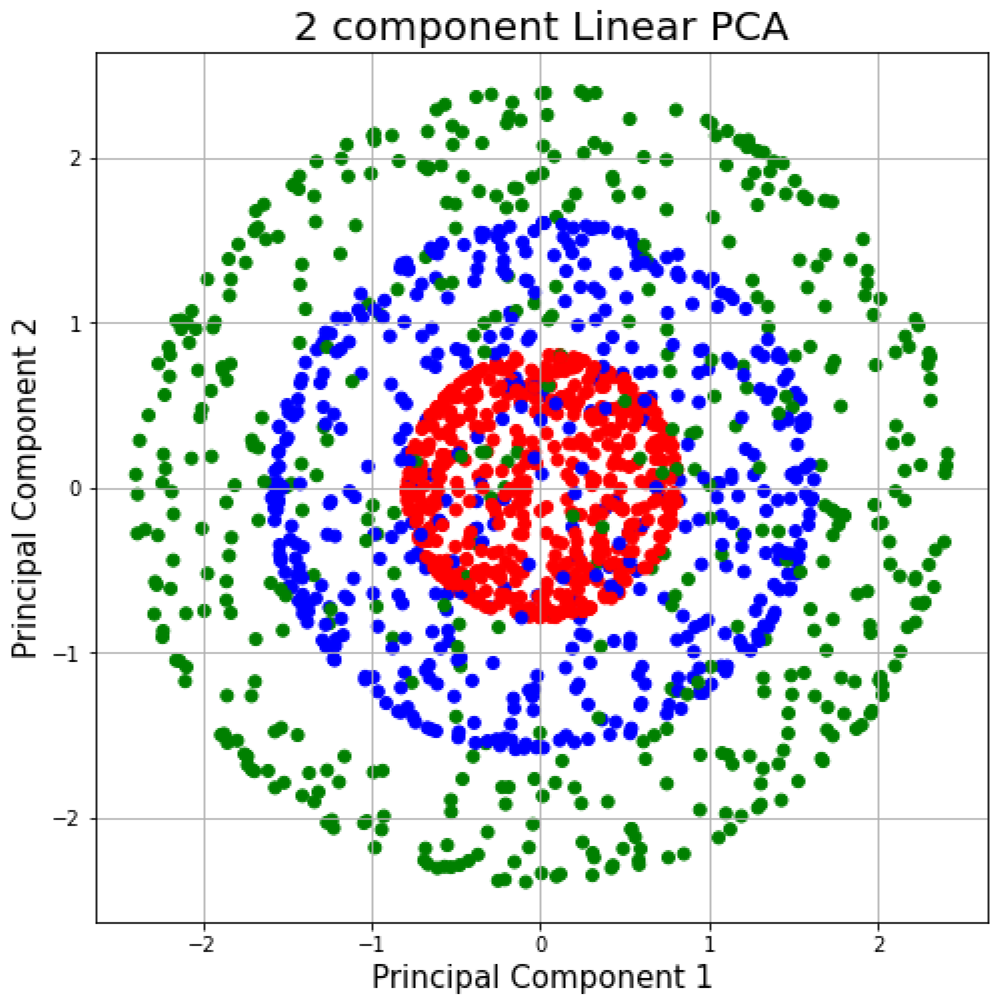

Radius = 3
4
3
2
1
0
−1
−2
−2
−1
0
1
2
3
4
4
3
2
1
0
−1
−2

Below is the plot of the data



Now next step is to normalize the data and applying the Linear PCA

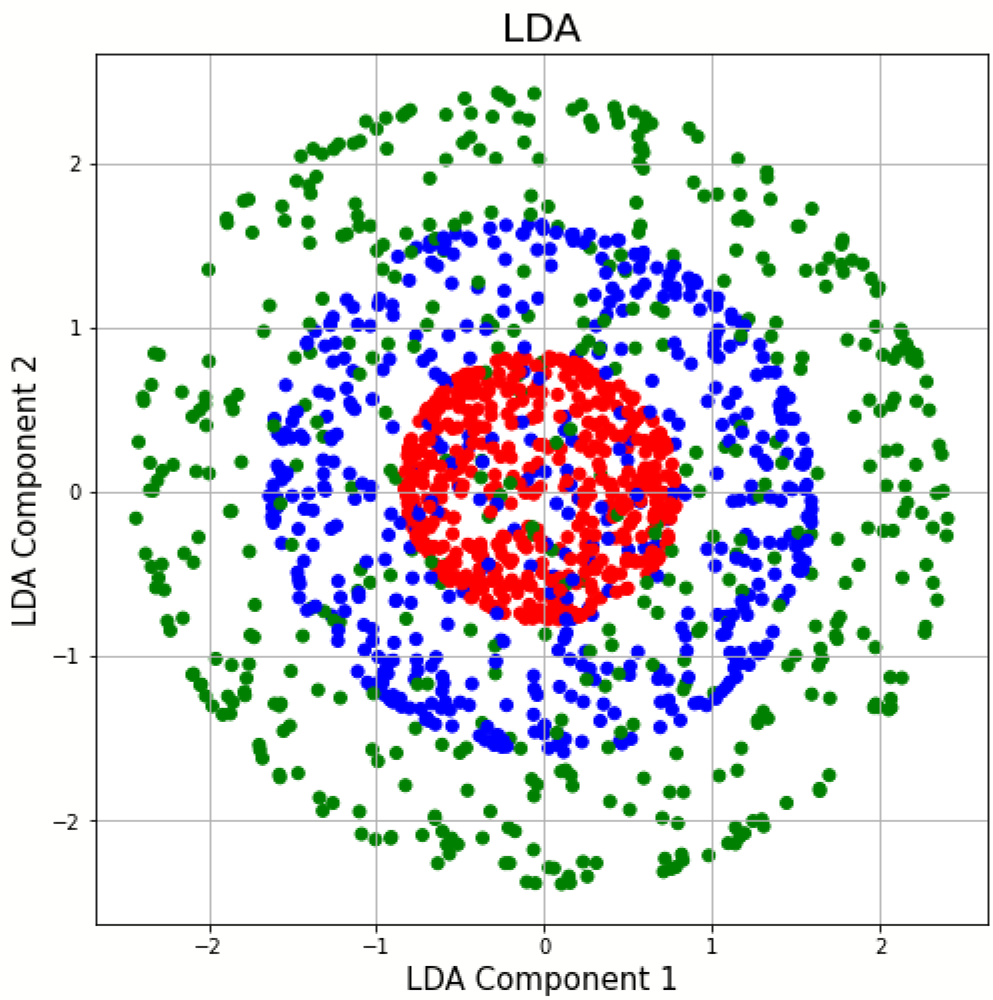
Results of Linear PCA



Since we know that the data is not linearly separable , hence the linear won’t give us this the appropriate result i.e data is not seperated. We can also have a look at the result of the command : **pca.explained_variance_ratio_** which gives the percentage of variance explained by each of the selected components are : 0.34498655, 0.33884903.

Hence total information preserved = 0.34498655+0.33884903 = 0.68383558 i.e 68% of information is preserved.

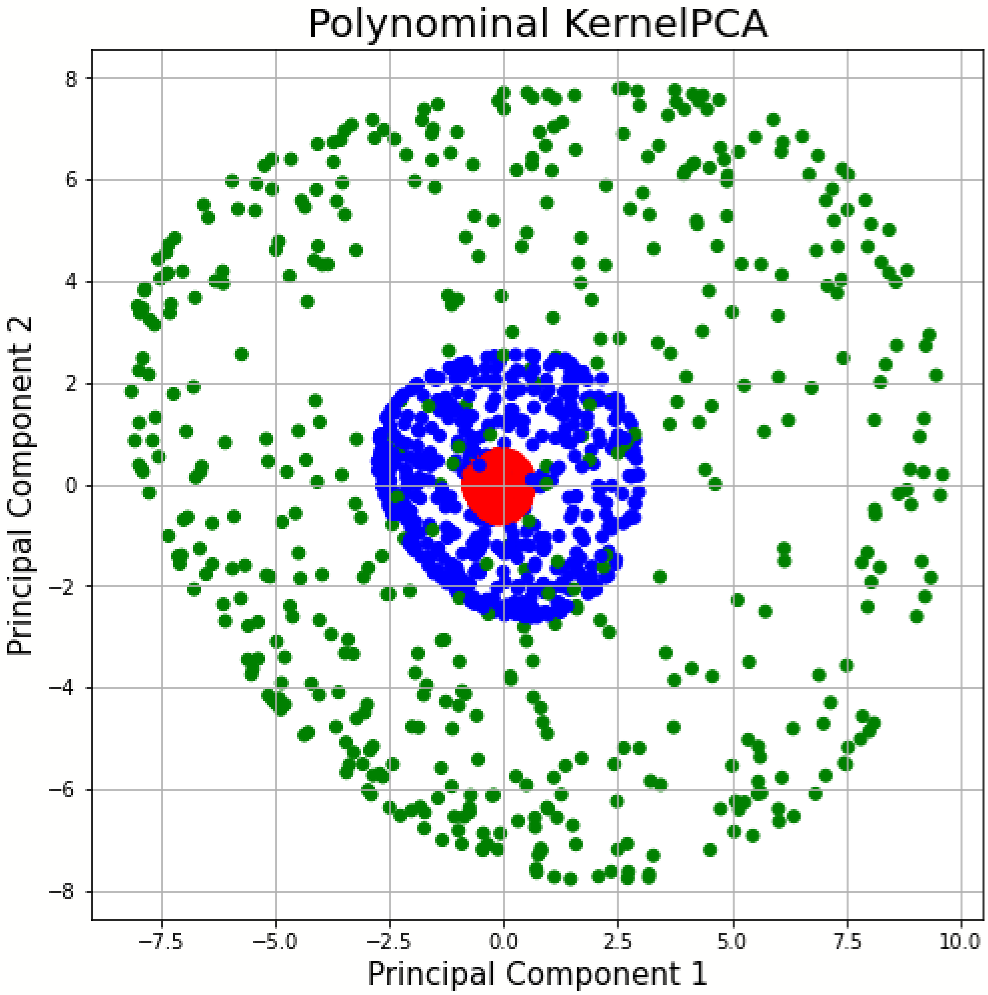
Now applying LDA :



So LDA tries to reduce the number of dimensions (i.e. variables) in a dataset while retaining as much information as possible. But here since the data is not linearly separable, the obtained result doesn’t have separated data.

Applying the KernelPCA

Polynomial Kernel of degree = 5



So we can see that Polynomial Kernel has done a better job than Linear PCA. It is able to separate the data except the few points but still missing a clear cut separating boundary.

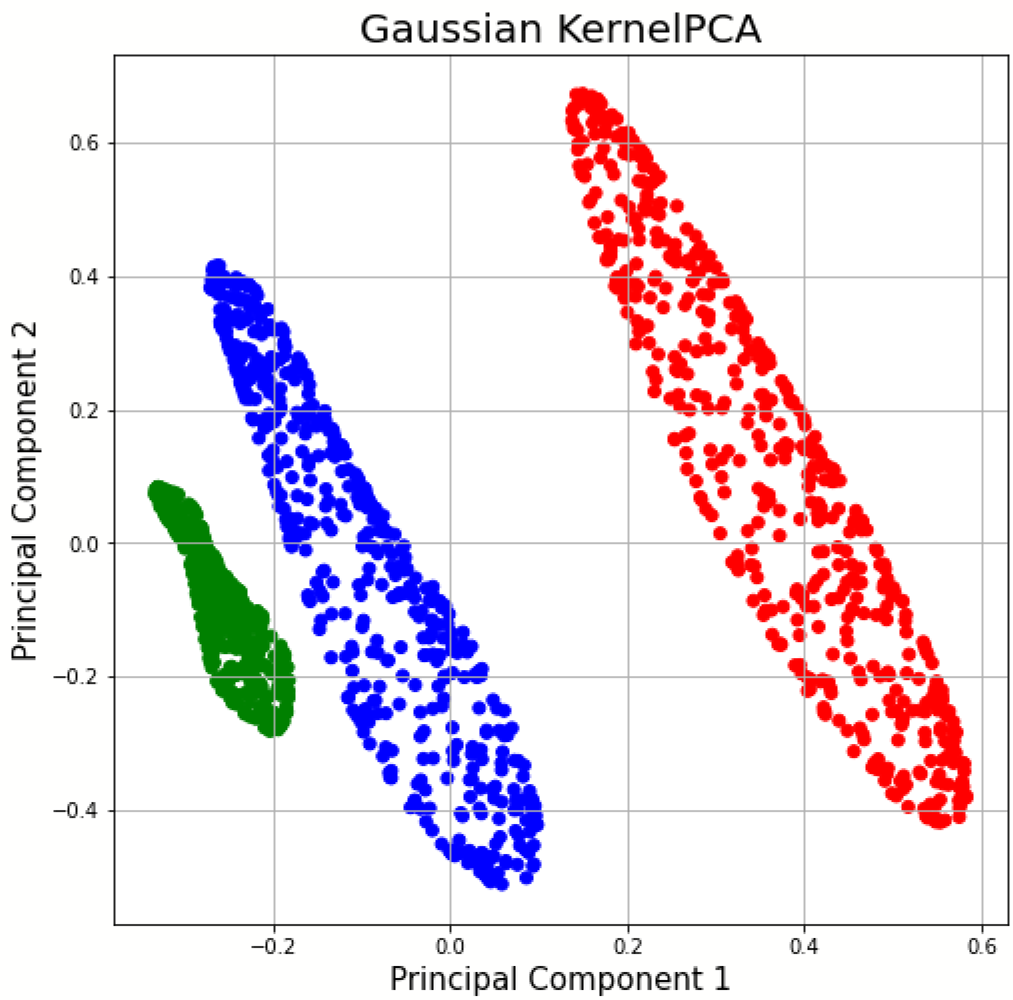
Gaussian Kernel of gamma = 0.8

$\gamma = 1/2\sigma^2$, So,

$K(x, x') = \exp(-\gamma \|x - x'\|^2)$

which is the radial basis function kernel in scikit learn

Hence the sigma $\sigma = 0.79$



Here the gaussian has done an excellent job , separating the data and showing a clear cut separating boundary. Recall a kernel expresses a measure of similarity between vectors. The RBF kernel represents this similarity as a decaying function of the distance between the vectors (i.e.the squared-norm of their distance). That is, if the two vectors are close together then, $\|x-x'\|$ will be small. Then, so long as $\gamma >$

0, it follows that $-\gamma\|x-x'\|^2$ will be larger.Thus, closer vectors have a larger RBF kernel value than farther vectors. So points on sphere or radius 1 will have low rbf and will have high values for points at s radius = 2 and 3. Hence it is able to separate the data well.