

Santosh S R
ME20B157

CS5691:
Pattern Recognition and Machine Learning

Assignment 1

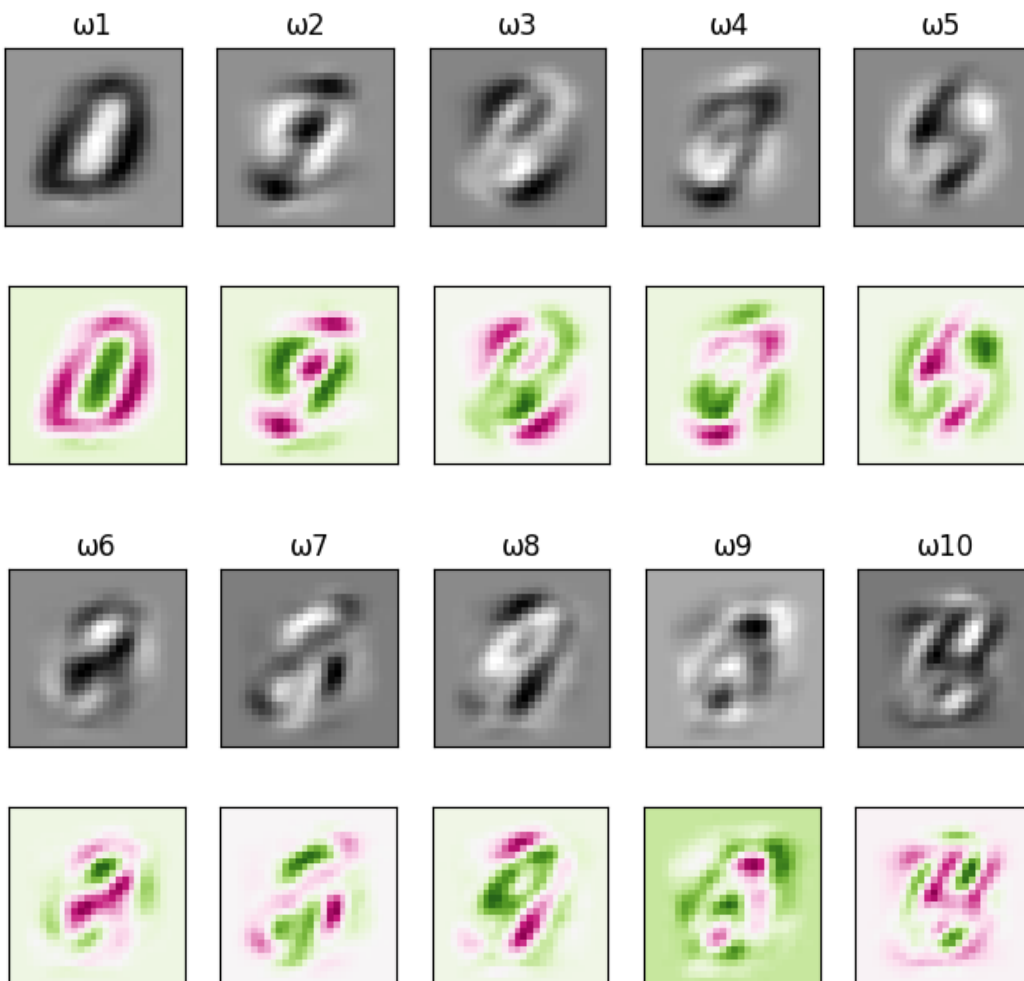
PCA
KPCA
K-Means
Kernel K-means

Question 1:

(i) The first 10 Eigenvectors or the first ten principal components are the below images. The first principal component explains about ~10% of the variance in the data, followed by the second component explaining ~8% and so on.

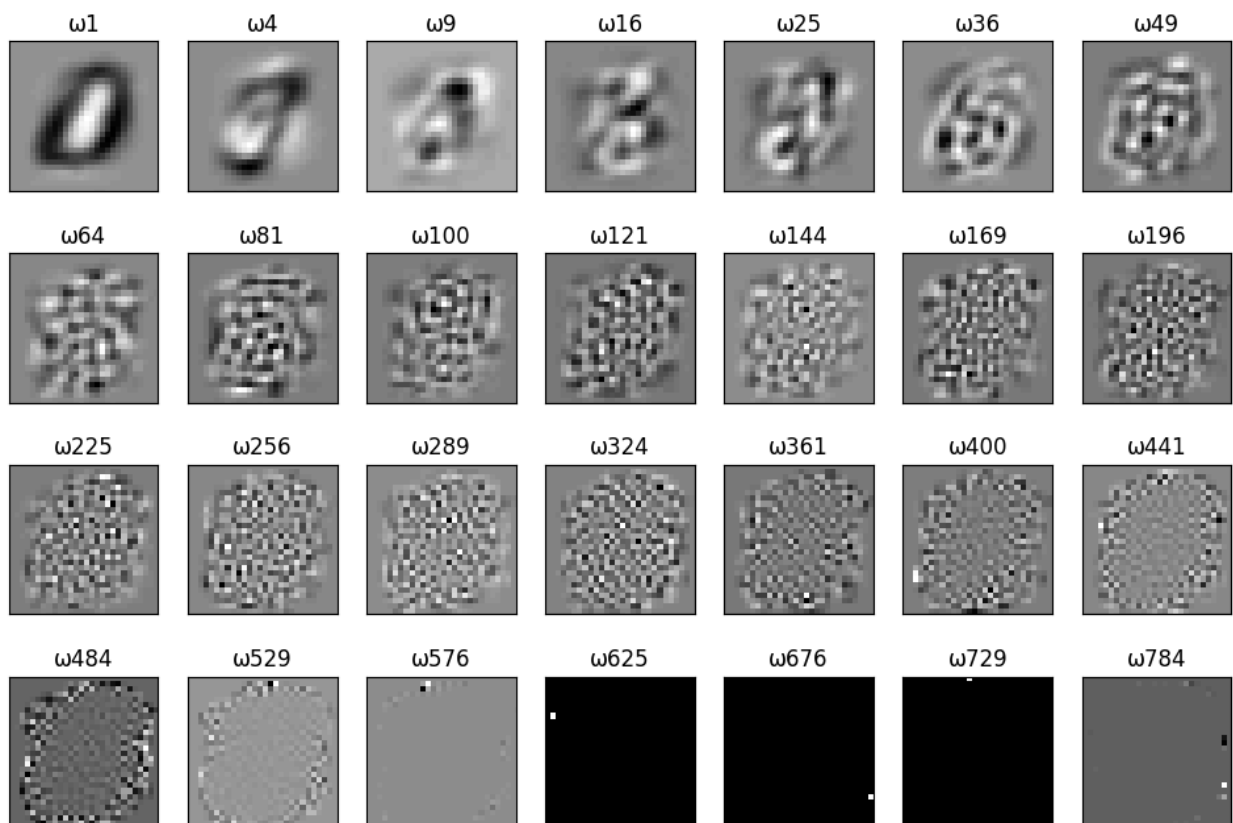
Clearly the Initial vectors explain information in accordance to their EigenValues, that are higher the more primary components.

Figures: Top 10 principal components



Here the Principle Components of a square sequenced ordering has been given, we can see that the principal components are getting more and more spreaded as the component number increases. As the index increases the points of importance are very scattered unlike the first principal component which is more recognisable as a number with respect to the components showing scarcer points as we go on.

Figures: Principal components:

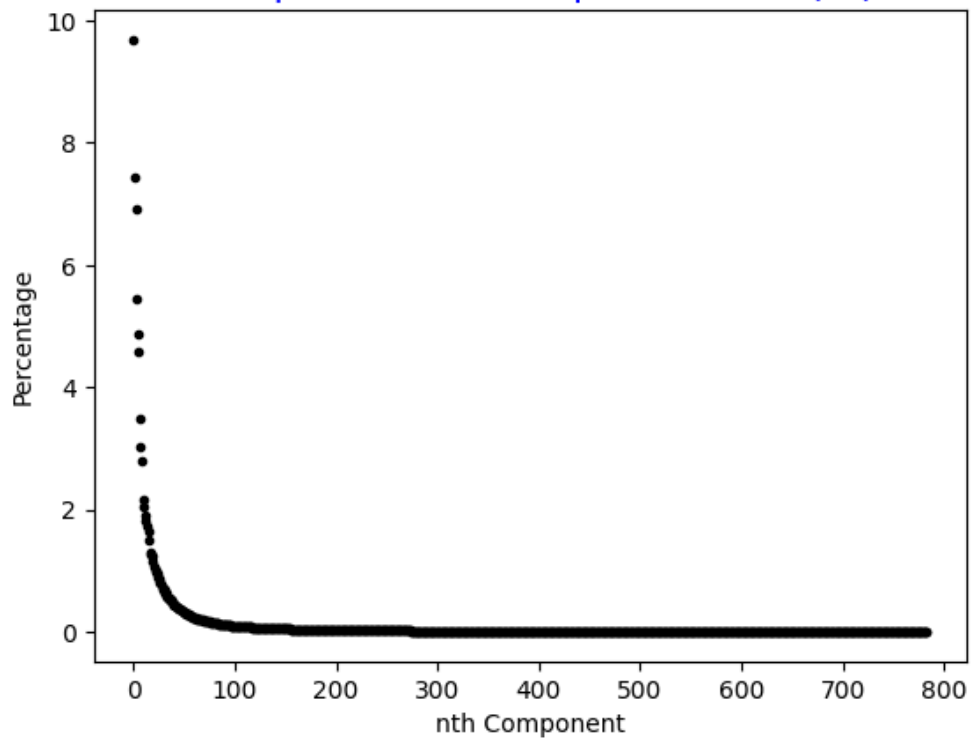


Figures: Reconstruction of Data Points



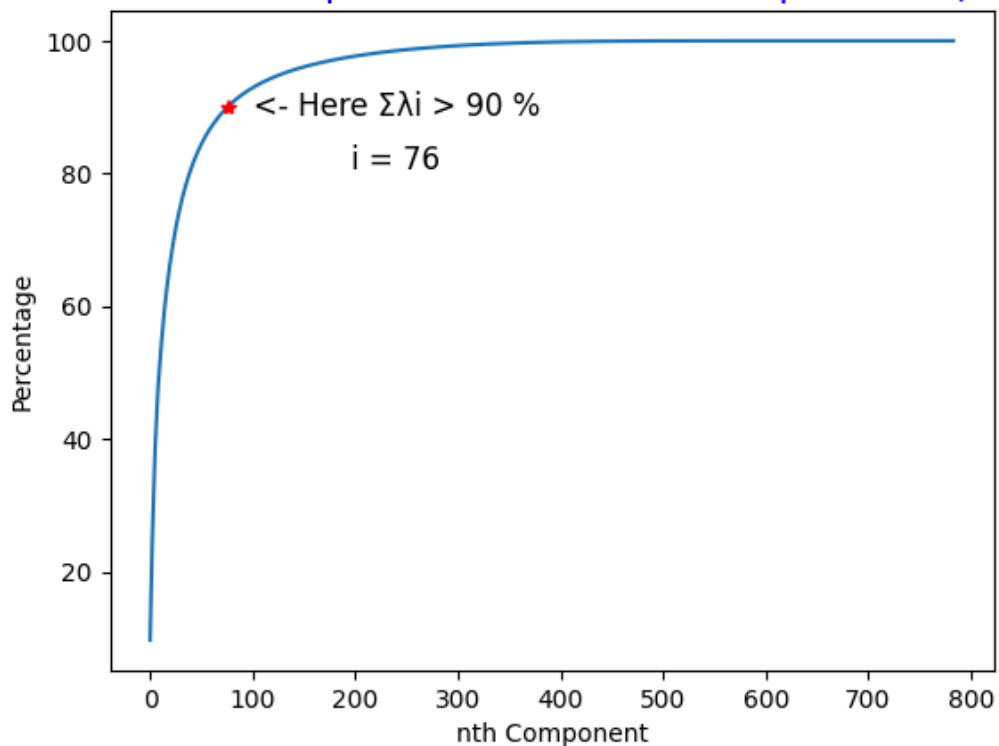
(ii) Plots: The EigenValue wise importance of Information captured

Component wise importance in (%)



Plots: Cumulative EigenValue wise importance of Information captured

Cumulative importance of first n components (in %)

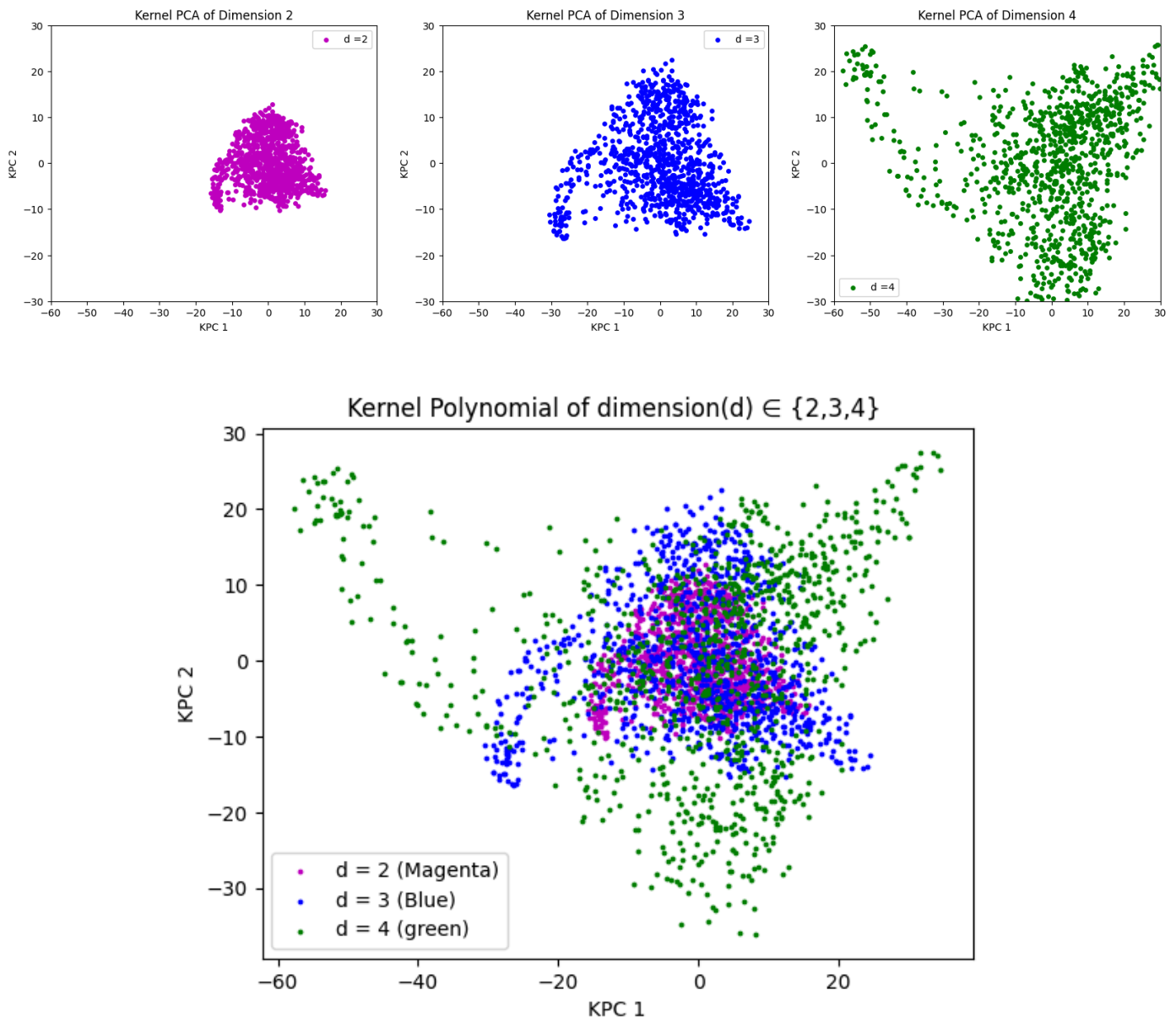


Looking at the Plotted Graphs and the Reconstructed images I would prefer 76 components to be considered as the information value post 76 components would be redundant considering the increase in storage needs.

(iii) $\kappa(x, y) = (1 + x^T y)^d$ for $d = \{2, 3, 4\}$

$$\kappa(x, y) = \exp \frac{-(x-y)^T(x-y)}{2\sigma^2}$$

Figures: Polynomial Kernel Regression



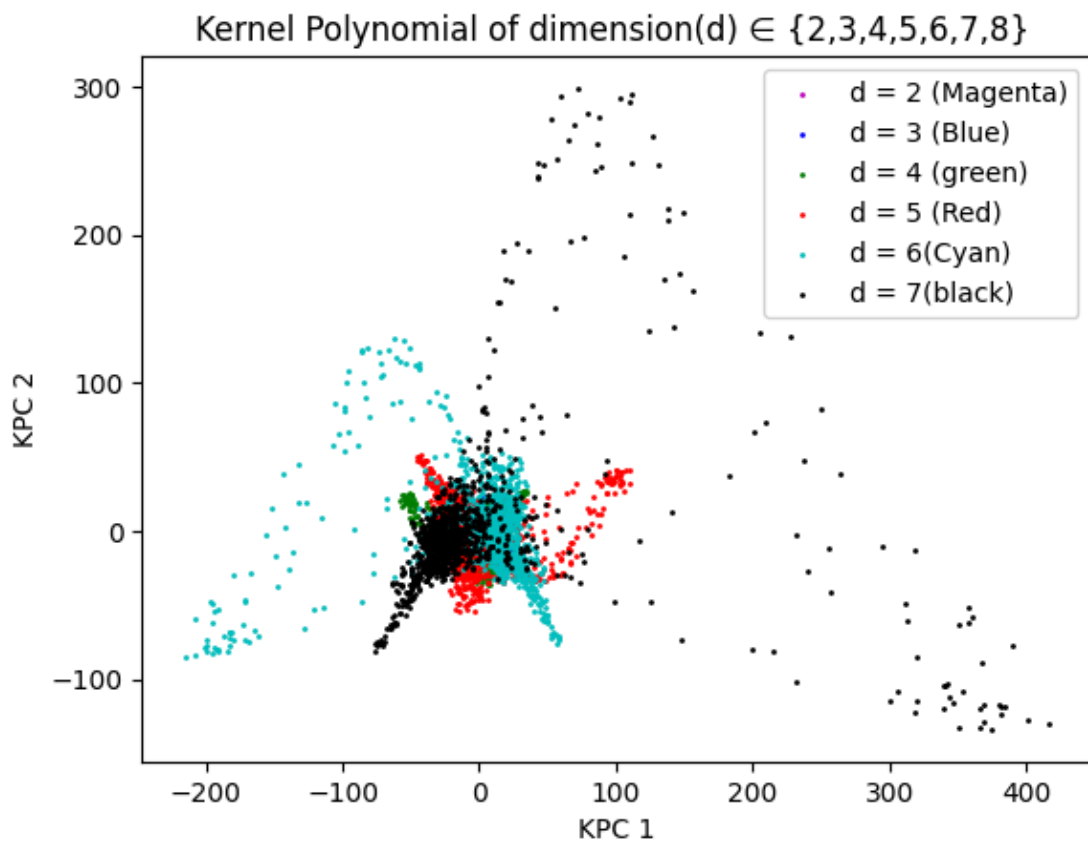
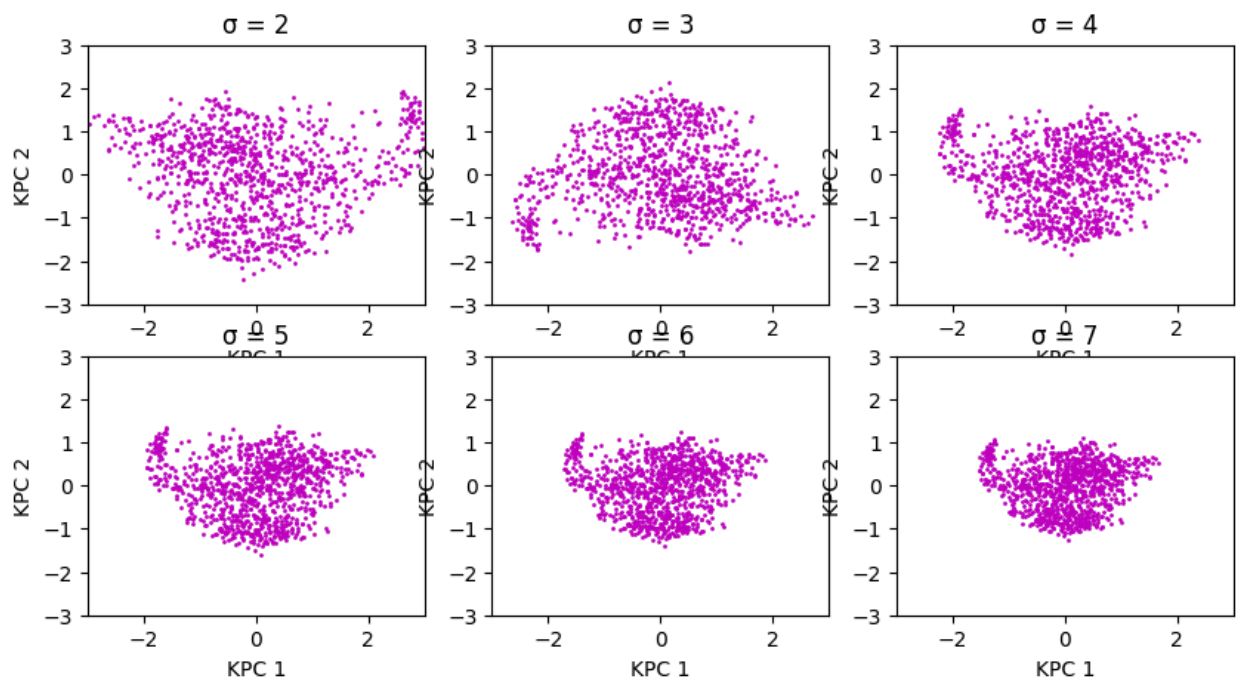


Figure: Radial Basis Function Kernel PCA



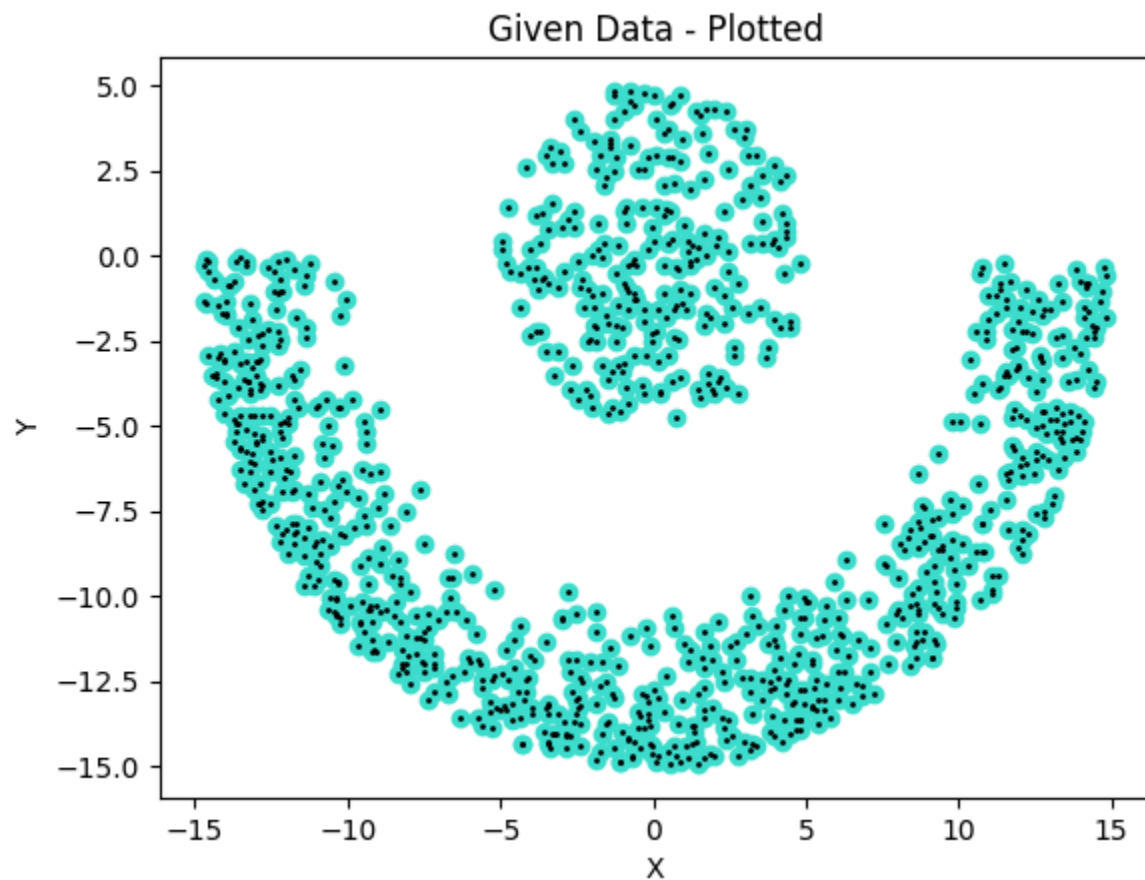
(iv)

Conclusions:

1. For MNIST's low-dimensional, digit-shaped data, complex dimensionality reduction isn't necessary.
2. Standard PCA effectively captures key variations and reduces features, making it ideal.
3. Kernel PCA or high-degree Polynomial Kernels for MNIST - they add complexity without much benefit, but capture intricate details in a small number of components

Question 2:

Given Dataset:



The "Crescent Moon" dataset, containing a collection of Cartesian points, is well-suited for unsupervised learning tasks, particularly K-Means clustering and its successors, which are demonstrated below.

(i), (ii)

Plot: Explaining Error reduction Vs Iteration for the K means / Lloyd's Algorithm ($k = 2$)

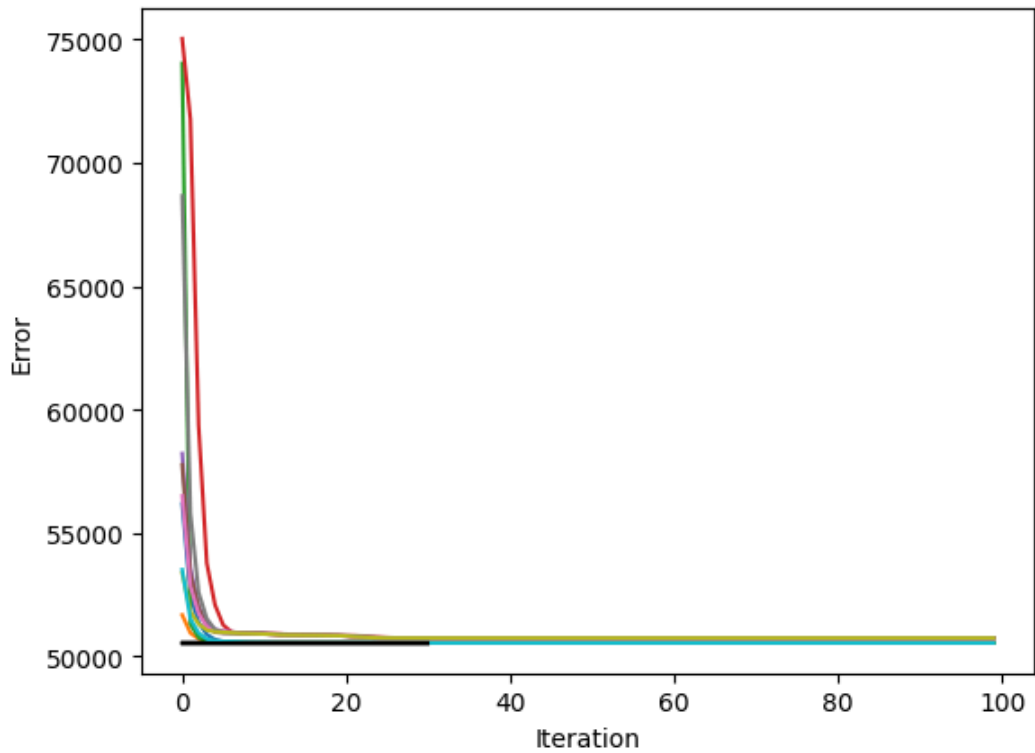


Figure: Bi colored Depiction of K means clustered Dataset ($k = 2$)

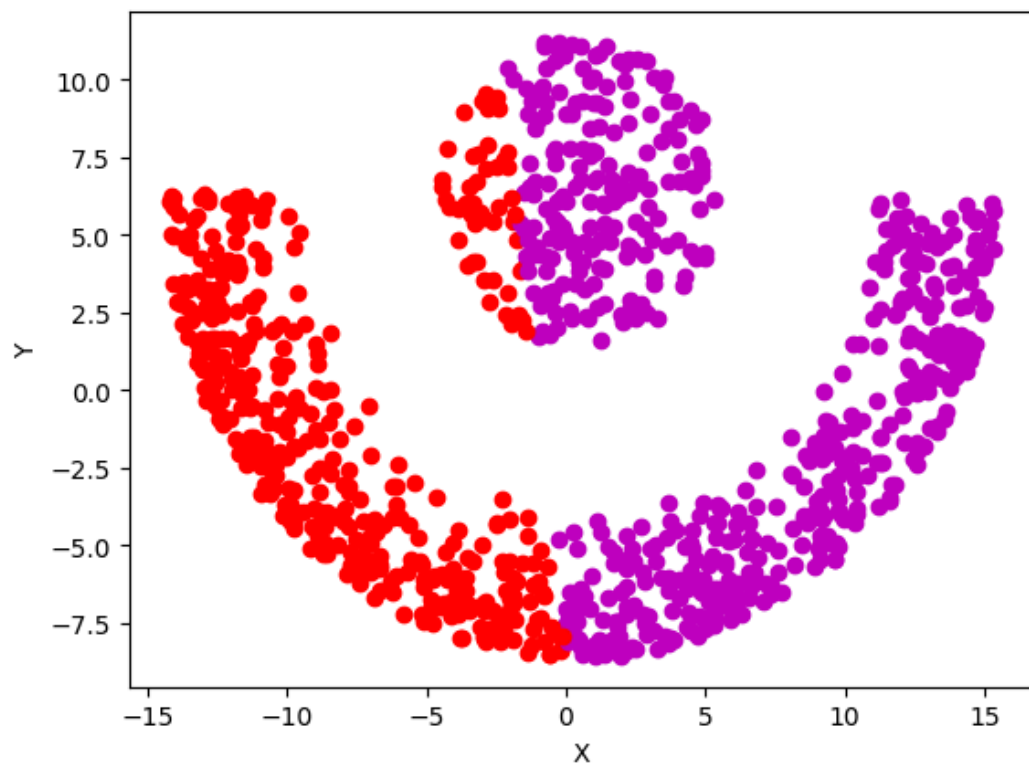
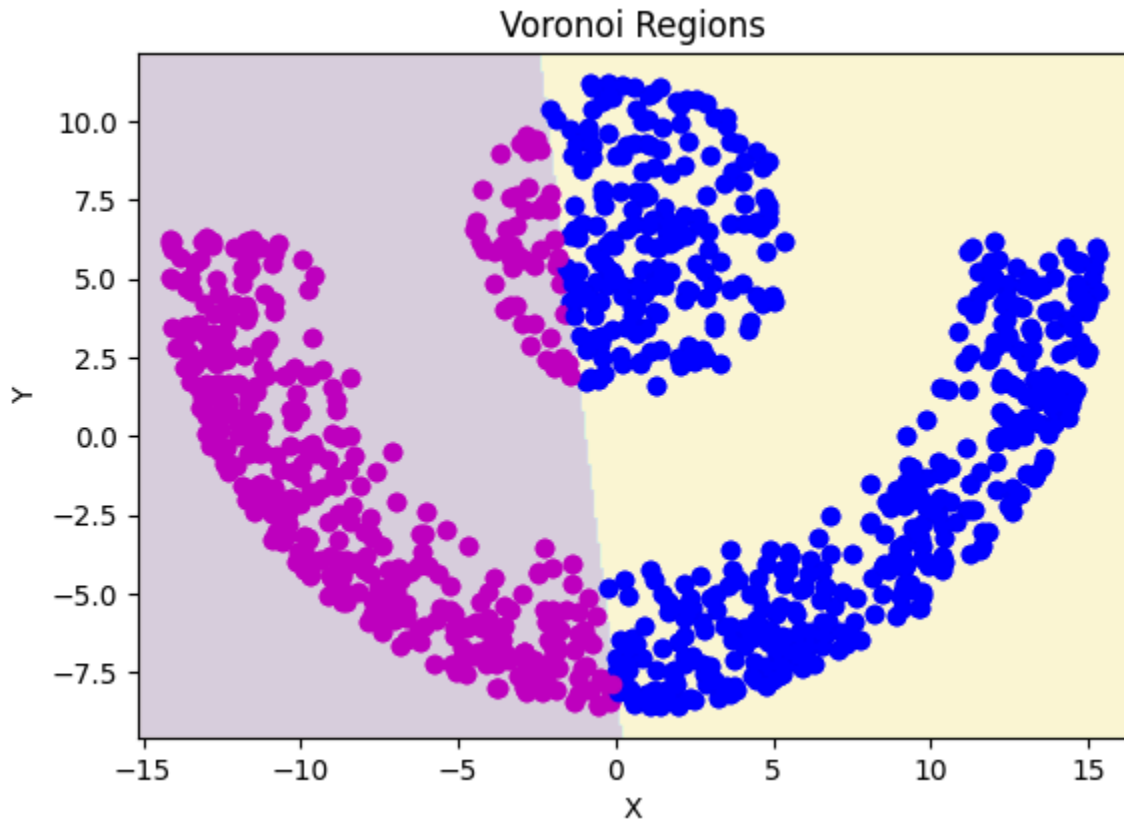


Figure: Voronoi Region of K Means (k =2) Cluster



PTO

Explaining Error reduction Vs Iteration for the Lloyd's Algorithm

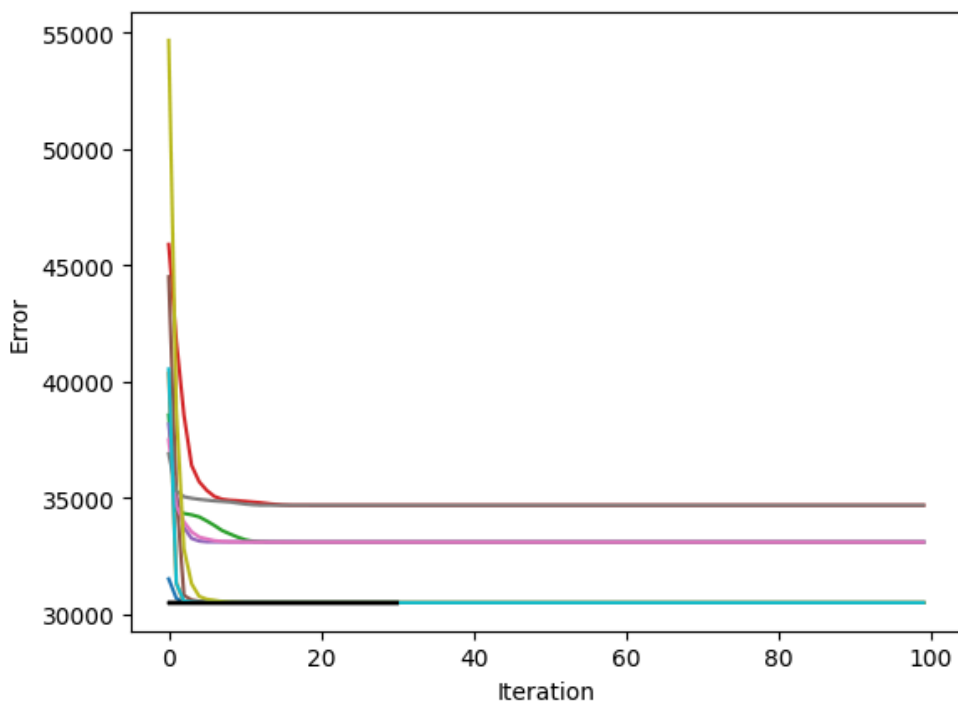


Figure: Tri colored Depiction of K means clustered Dataset ($k = 3$)

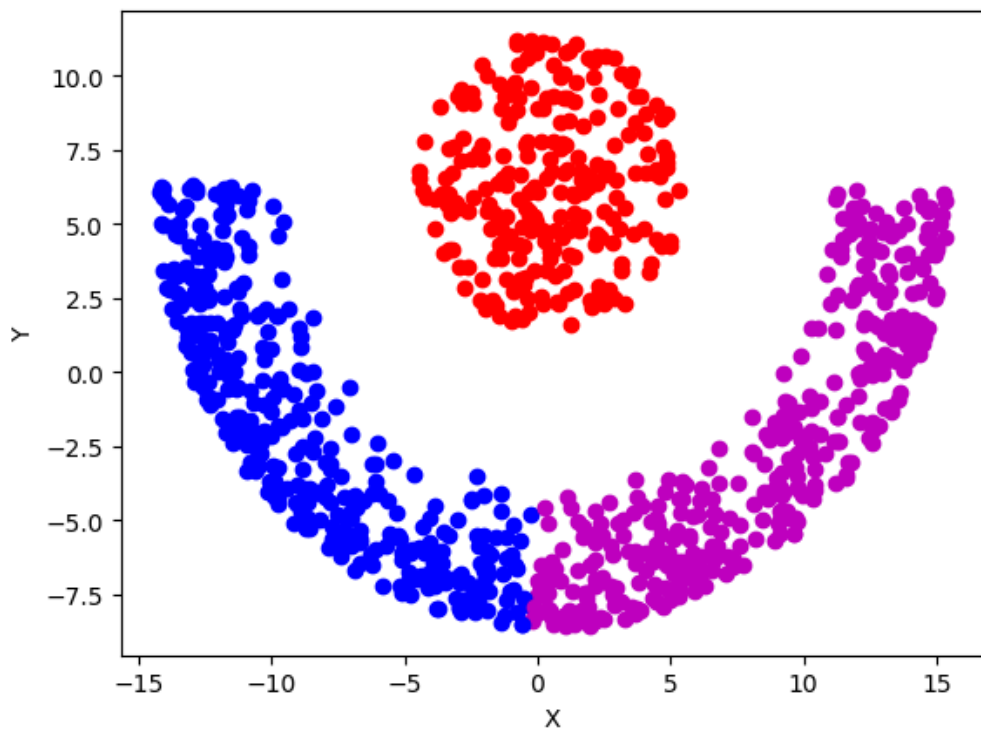
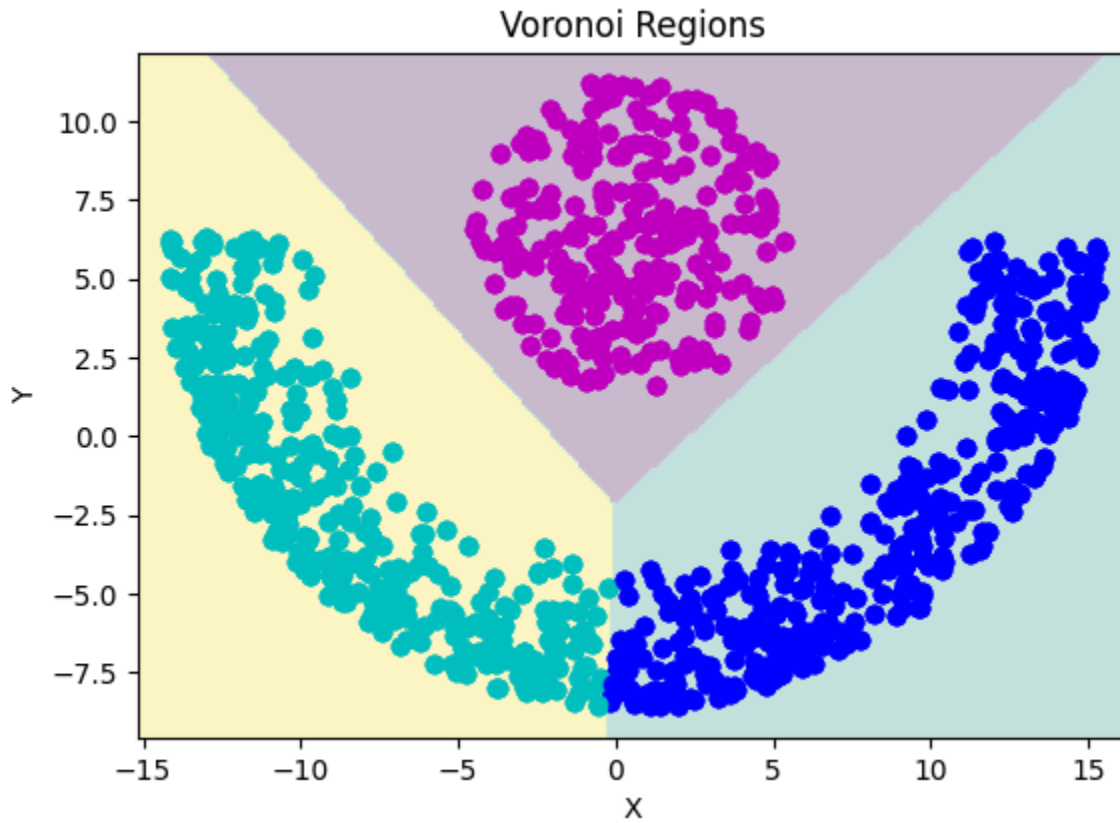


Figure: Voronoi Region of K Means (k =2) Cluster



PTO

Explaining Error reduction Vs Iteration for the Lloyd's Algorithm / K Means ($k = 4$)

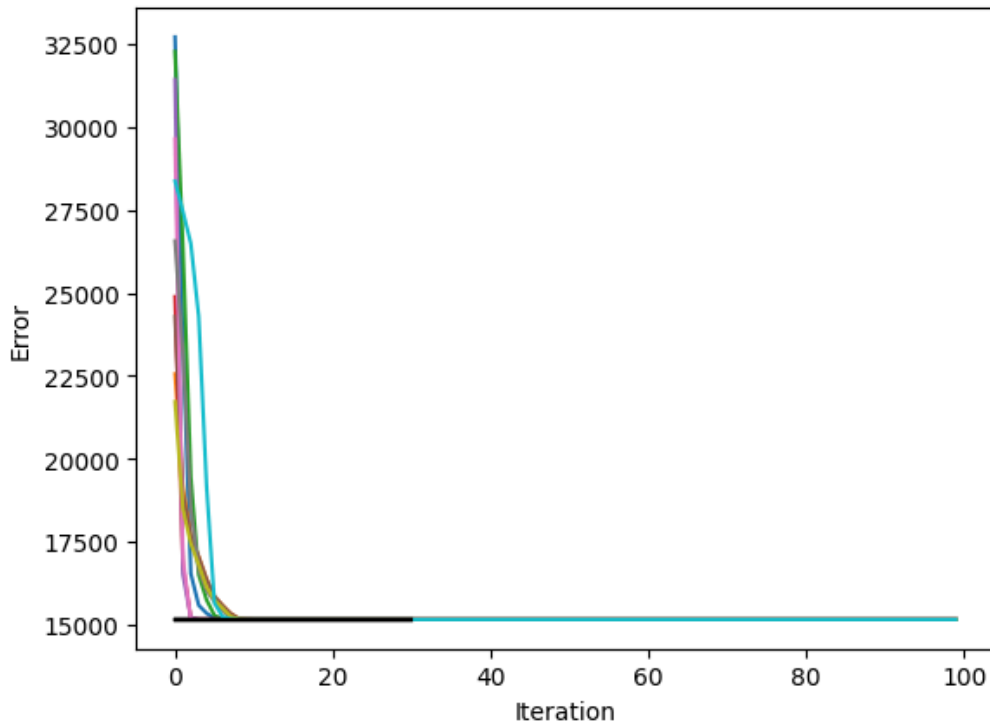


Figure:quad colored Depiction of Kmeans clustered Dataset ($k=4$)

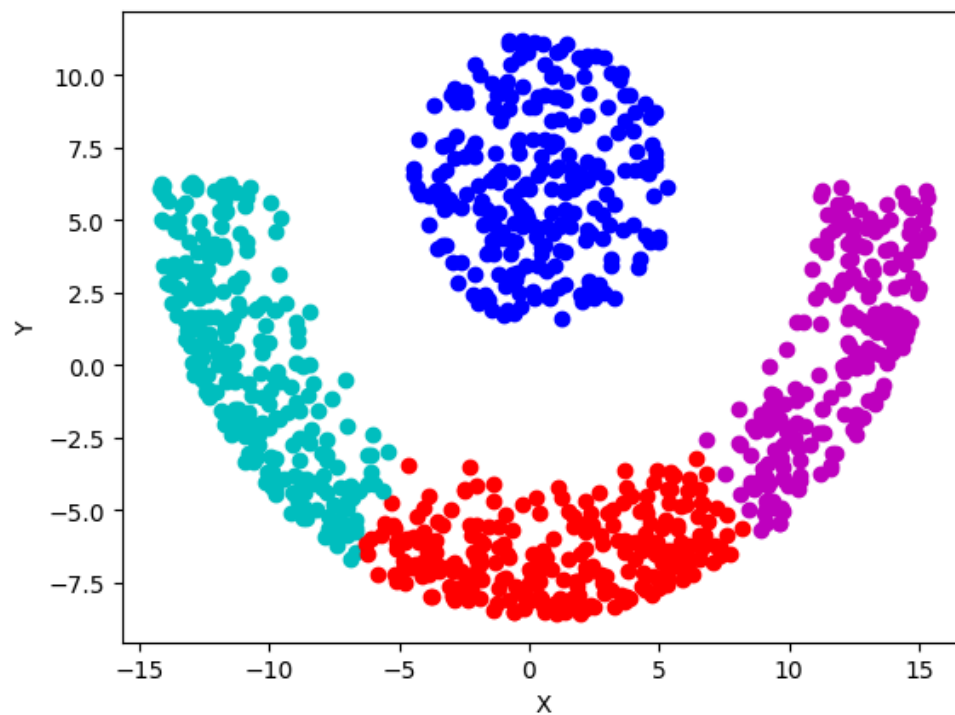
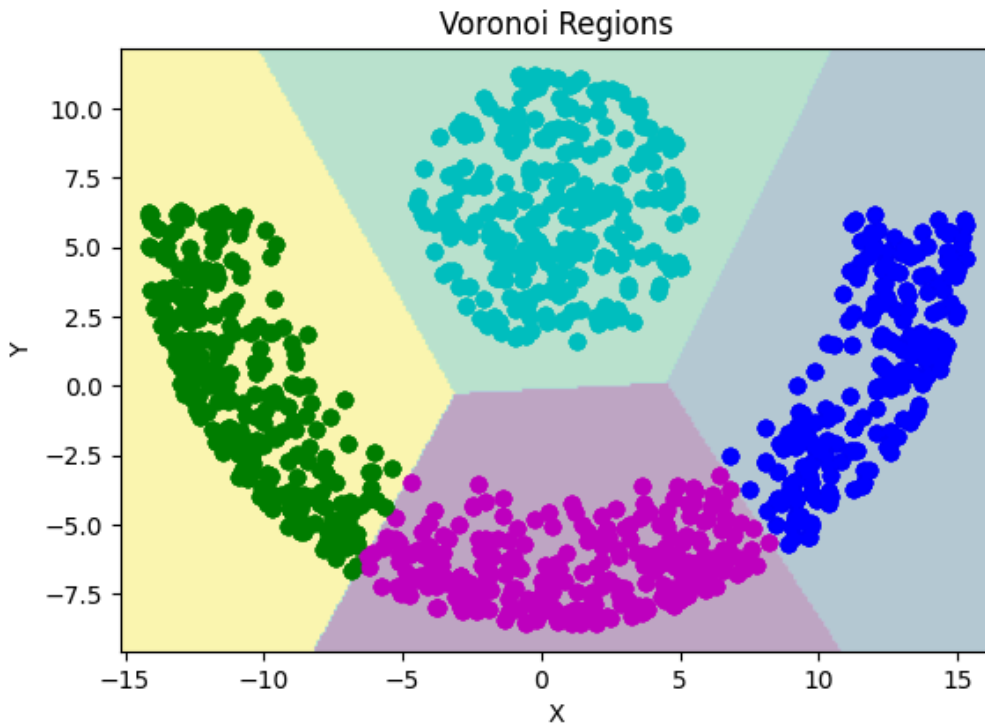


Figure: Voronoi Region of K Means ($k=4$) Cluster



(iii)

Figure: Cluster appearance of Relaxed Spectral Clustering Plot

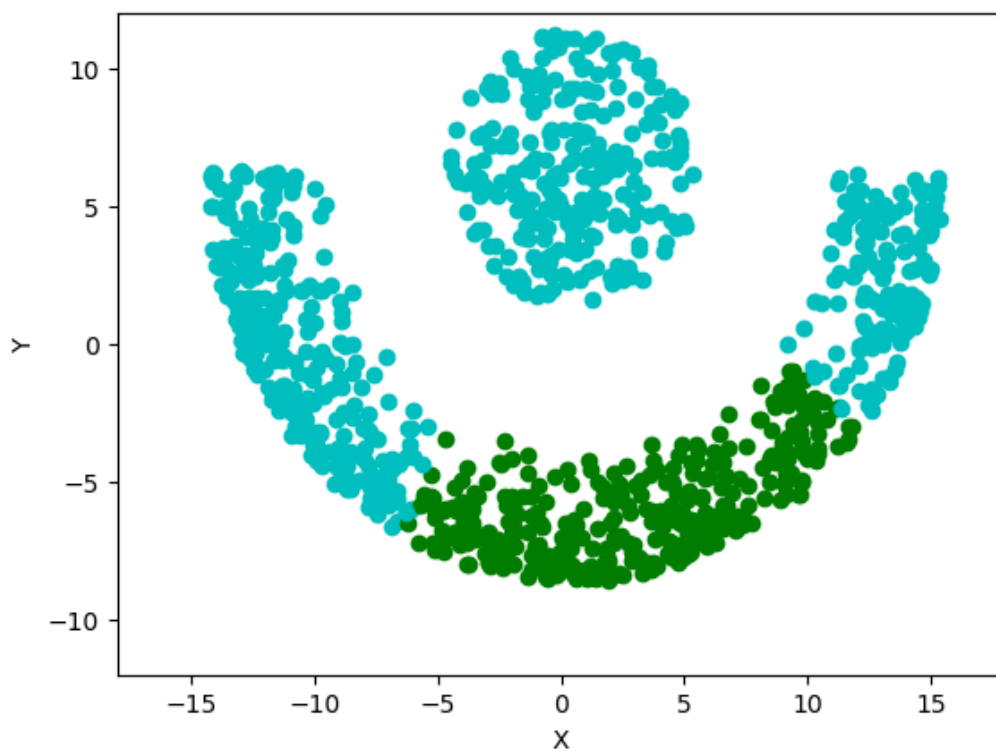
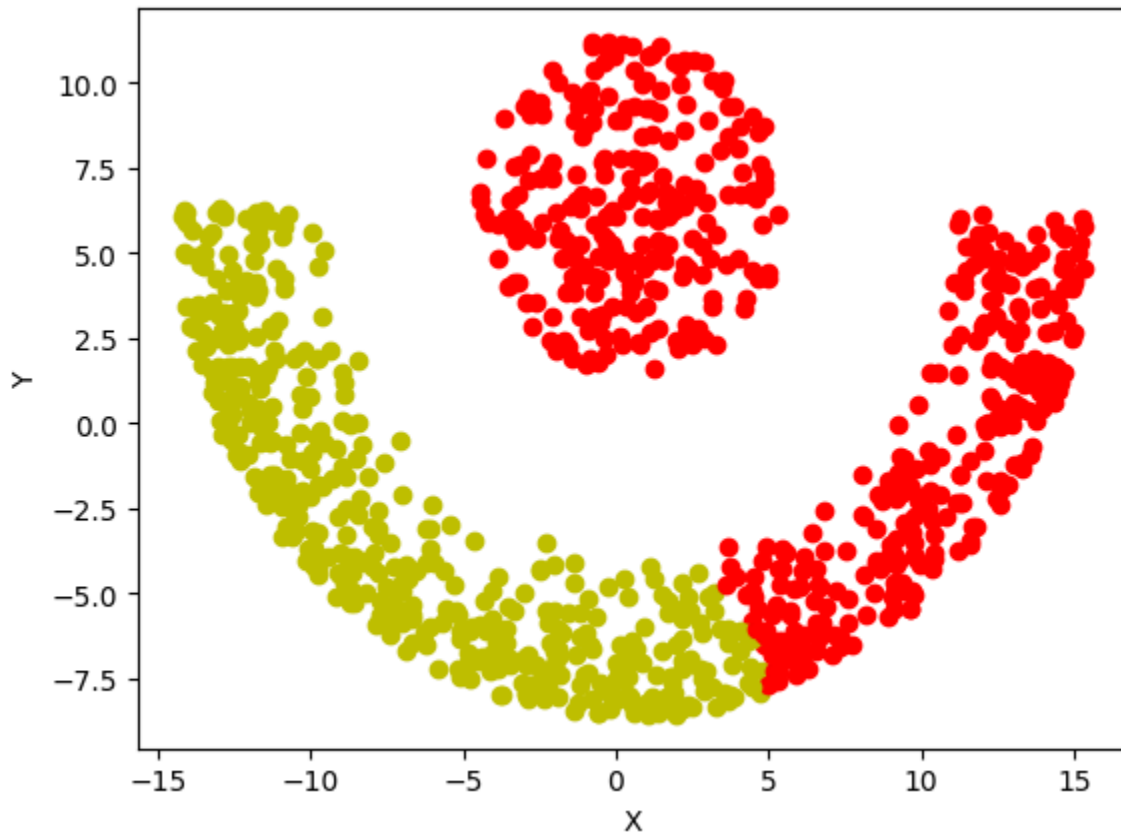


Figure: Cluster appearance of Custom Argmax Clustering Function



(iv)

Key features of each Clustering type:

- K-Means (Lloyd's Algorithm):
 - Simple and efficient: Easy to implement and understand, making it a good first choice for most clustering tasks.
 - Centroid initialization matters: The initial placement of centroids (cluster centers) significantly impacts the final

clusters. In the crescent moon case, carefully chosen centroids (e.g., one on each "horn") can lead to better separation.

- Kernel K-Means (Spectral Clustering):
 - Handling non-linearity: Can handle non-linear data structures, potentially useful if the crescent moon data has some underlying non-linearity (though likely not a major advantage for this simple shape).
 - Kernel function selection: The choice of kernel function (e.g., Gaussian) affects the clustering behavior. Experimentation might be needed to find the most suitable kernel for the crescent moon data.
- Custom Kernel Argmax Function:
 - Uses the already used kernel function and relaxes to approximate on the likelihood of H being $ZL^{.05}$

Overall the assignment was very hands on and simple enough for me to start and learn a lot. This has made me solidify my understanding of Unsupervised Learning