# Assignment - 01
# CS5691 Pattern Recognition and Machine Learning

Ramasamy Kandasamy
CS22M068

September 12, 2022

## Question - 1

### Introduction

#### Organization of code

There are four files:

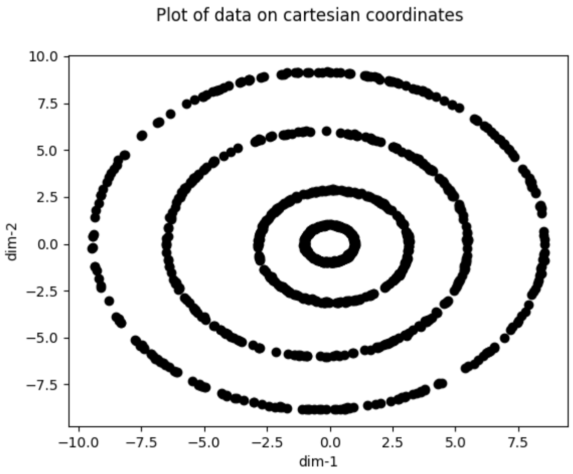
- `main.py`: This code is organized in the format of the assignment questions.
- `data.py`: This file contains all the functions to do PCA and clustering. Each of these functions are methods in the object data. The object data contains the original dataset and other derived information such the PC components, cluster idenitity etc.
- `plot.py`: This file contains all the code need to generate the plots.
- `utils.py`: Contains `custom_eigh` function which returns Eigen values and vector in the decreasing order of the eigen value.

**How to reproduce the plots?**
Unzip the file `Solutions_CS22M068`. On a Linux machine run the following commands.

```
$ mkdir plots
$ python3 main.py
```

#### Exploratory data analysis

A plot of the input data on a cartesian coordinate, indicates that the the datapoints are not linearly independent.



## Q1-(i) Running PCA with centering

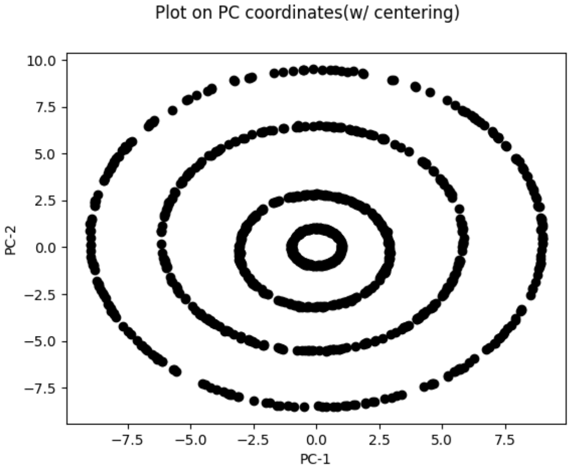
The code for running PCA in `data.py` file in the function `pca`.

The result of running PCA are as follows. The following information would be output into the standard output when `main.py` is run.

```
PCA with centering
==================
PC - 1  :  [0.323516  0.9462227]
PC - 2  :  [-0.9462227  0.323516 ]


Variance along each of the principal components
-----------------------------------------------
Variance along PC- 1  is  52.092671607646466 %.
Variance along PC- 2  is  47.90732839235353 %.
```

The plot of data points along the principal component axes are as follows.



## Q1-(ii) Running PCA w/o centering

The code for running PCA in `data.py` file in the function `pca`.

The result of running PCA without centering are as follows:

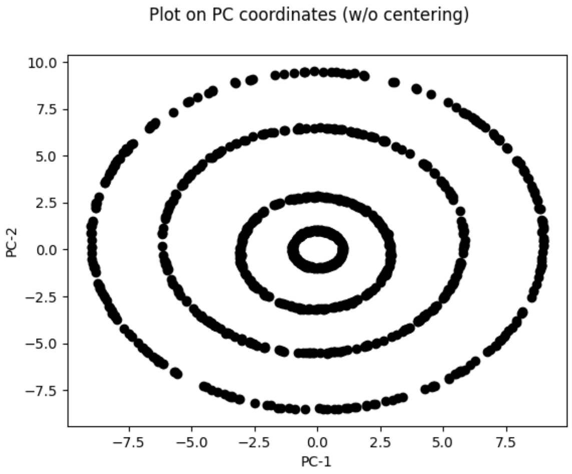
```
PCA without centering
=====================
PC - 1  :  [0.323516  0.9462227]
PC - 2  :  [-0.9462227  0.323516 ]


Variance along each of the principal components
-----------------------------------------------
Variance along PC- 1  is  52.092671607646466 %.
Variance along PC- 2  is  47.90732839235353 %.
```

The plot of data points along the principal component axes are as follows.
There is no difference in the value of principal components and variance of data along the principal components. This is also reflected in the PCA plot. They look identical for both PCA with and without centering.

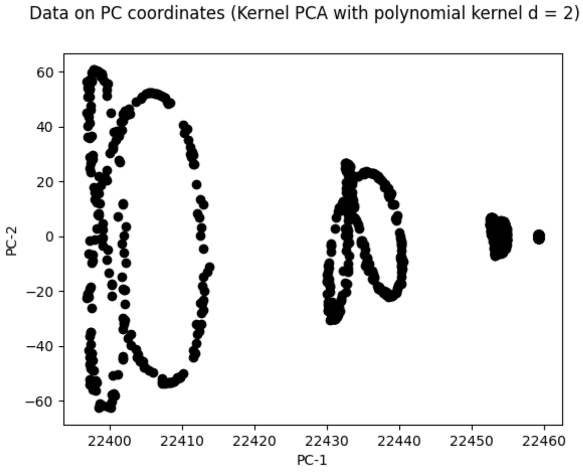
Here centering does not change anything because the data is already nearly centered. The center of the data points is as follows:

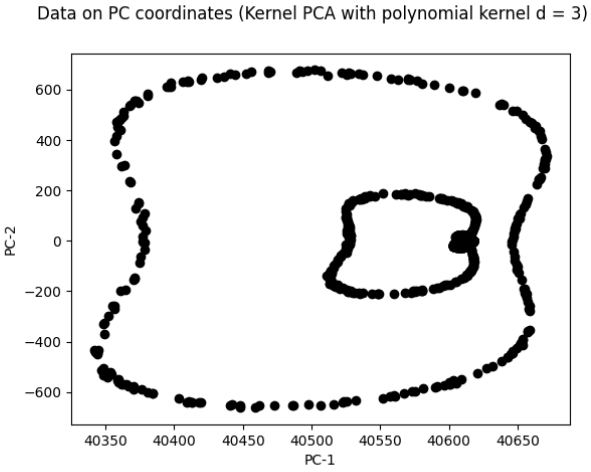
```
Print mean of the data points
=============================
Center of the data:  [2.398081733190338e-17, -5.1514348342607266e-17]
```



## Q1-(iii). A.

Kernel PCA was performed with polynomial kernel given in the assignment with $d = 2$ and $d = 3$. Plots of data point onto the top 2 principal components for these two kernels are as follows:

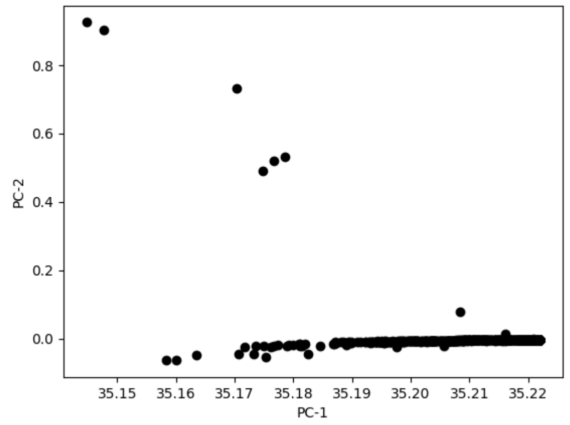




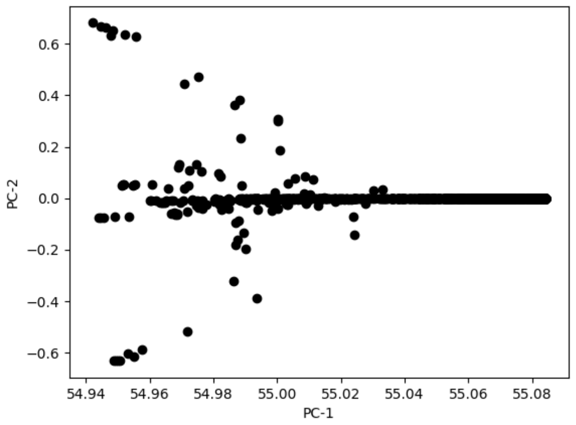
## Q1-(iii). B.

Kernel PCA was performed with gaussian kernel given in the assignment with $\sigma = \{0.1, 0.2, \cdots, 1.0\}$ .Plots of data point onto the top 2 principal components for these two kernels are as follows:

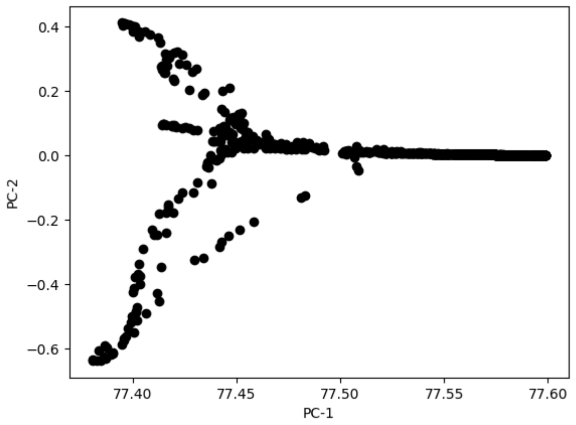
Data on PC coordinates (Kernel PCA with gaussian kernel s = 0.1)



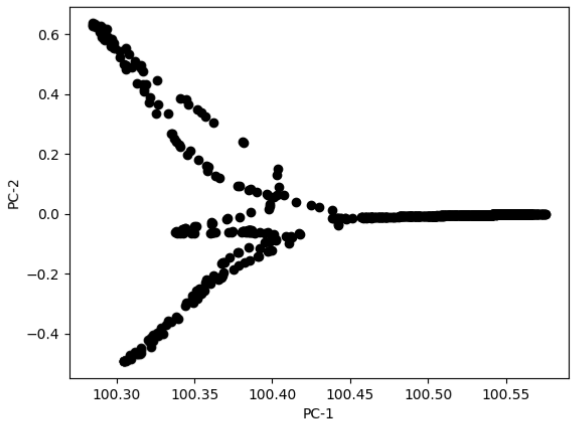
Data on PC coordinates (Kernel PCA with gaussian kernel s = 0.2)



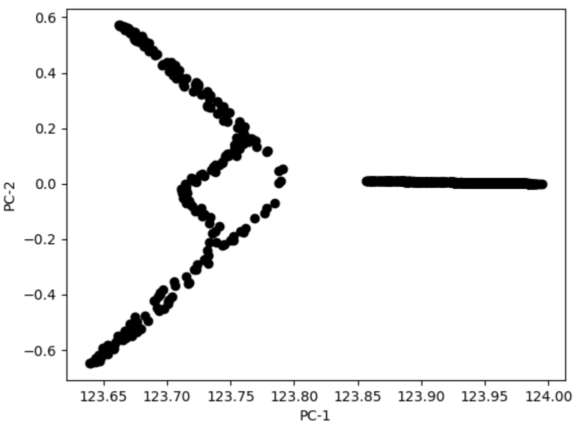
Data on PC coordinates (Kernel PCA with gaussian kernel s = 0.3)



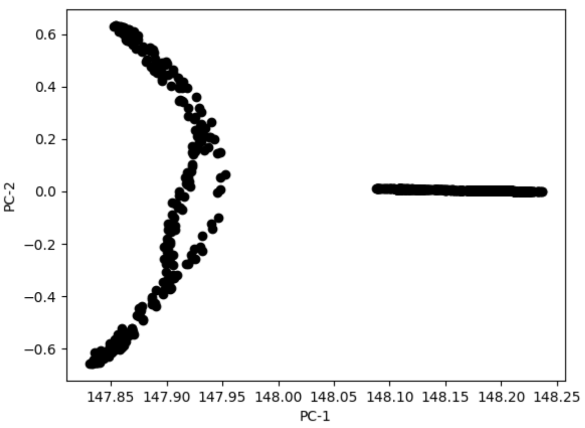
Data on PC coordinates (Kernel PCA with gaussian kernel s = 0.4)



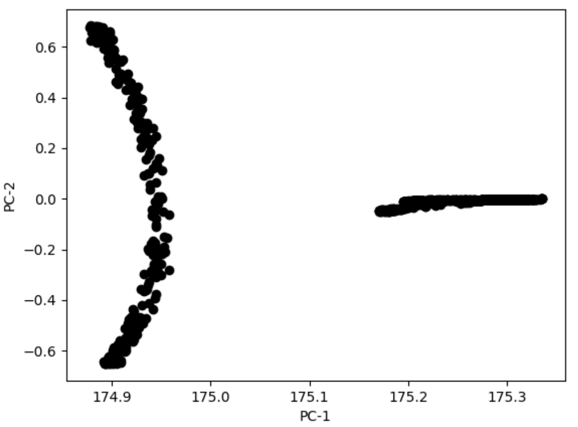
Data on PC coordinates (Kernel PCA with gaussian kernel s = 0.5)



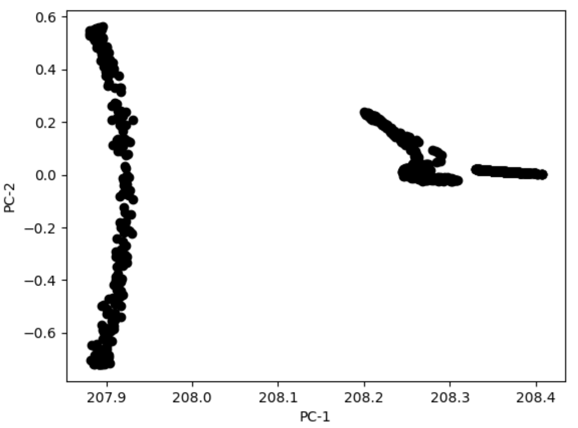
Data on PC coordinates (Kernel PCA with gaussian kernel s = 0.6)

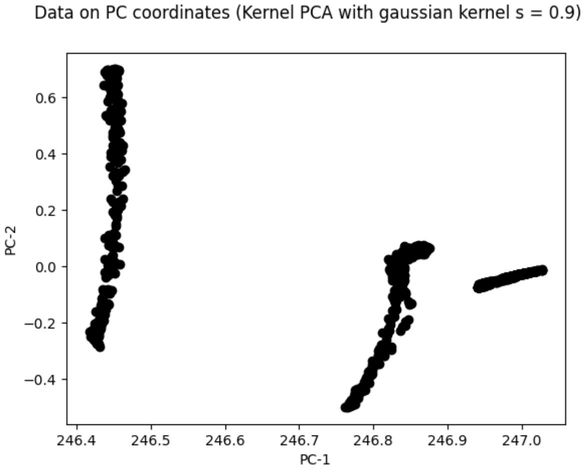


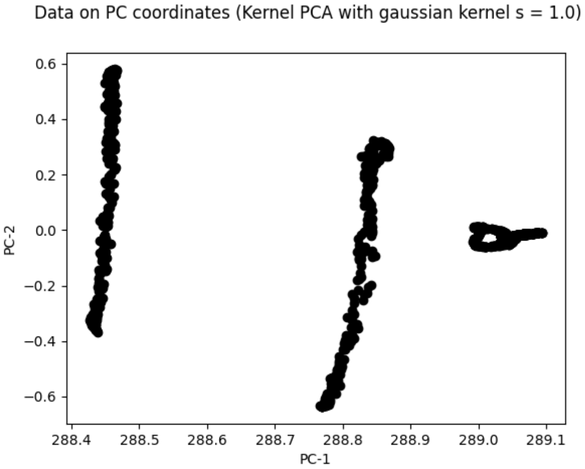
Data on PC coordinates (Kernel PCA with gaussian kernel s = 0.7)



Data on PC coordinates (Kernel PCA with gaussian kernel s = 0.8)

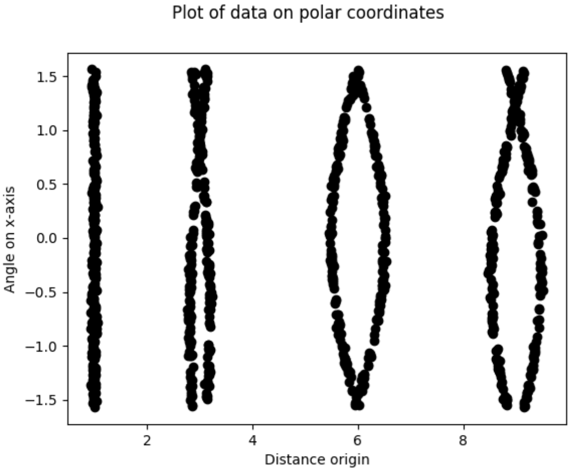




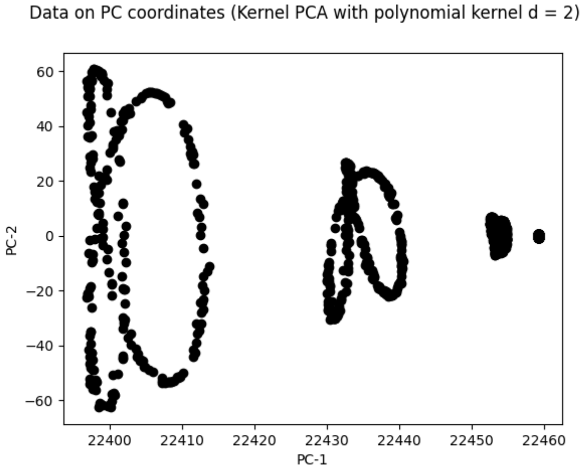


## Q1-(iv)

I think the kernel best suited for this data set is polynomial kernel in assignment with $d = 2$. The justification is as follows.
The data is in the form of concentric circles. The data are closely related based on radius. Plotting the data in polar coordinated gives:



In this coordinate the datapoints are linearly independent. This plot looks simillar to the plot obtained by polynomial kernel with $d = 2$

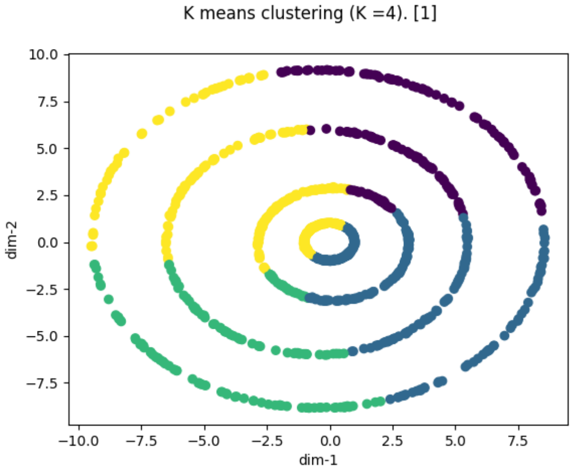


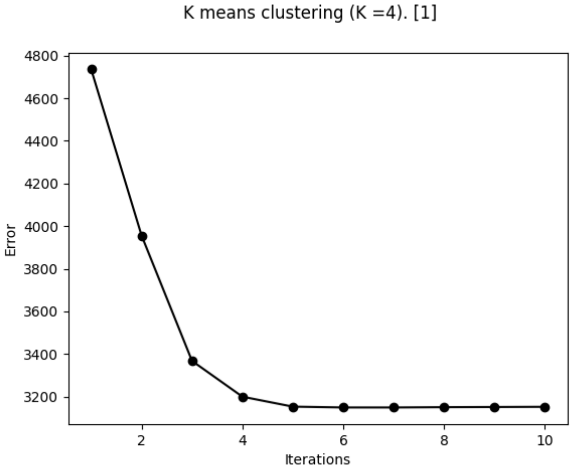
Data on PC coordinates (Kernel PCA with polynomial kernel d = 2)

## Q2

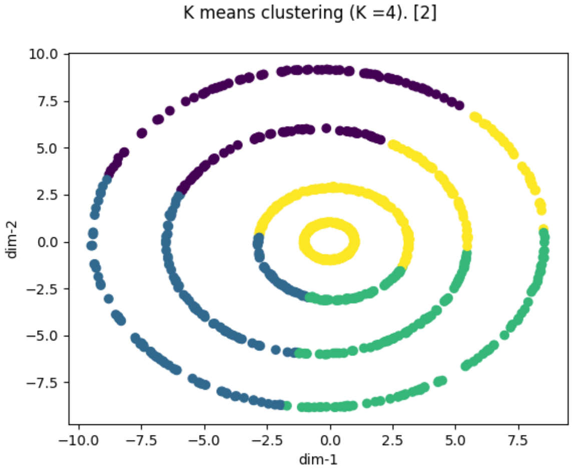
### Q2-i K - means clustering for 5 different initilizations

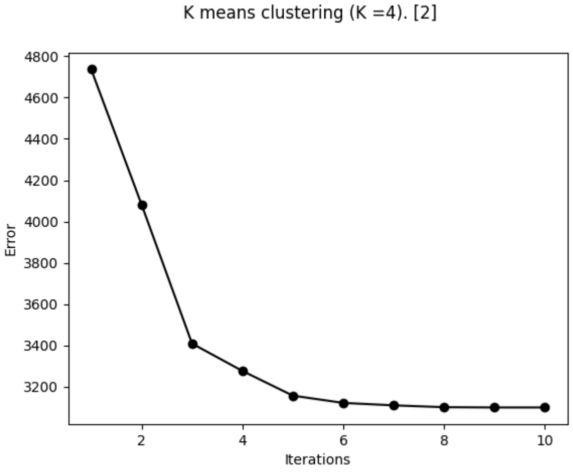
**Initialization-1**



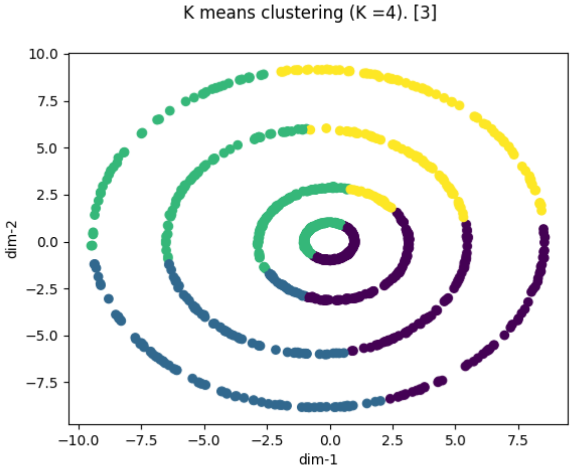


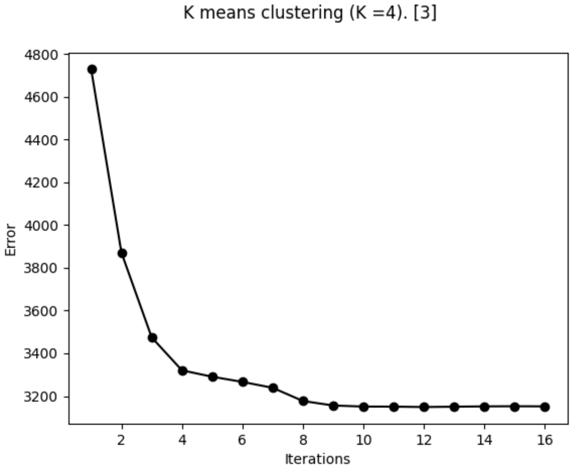
**Initialization-2**



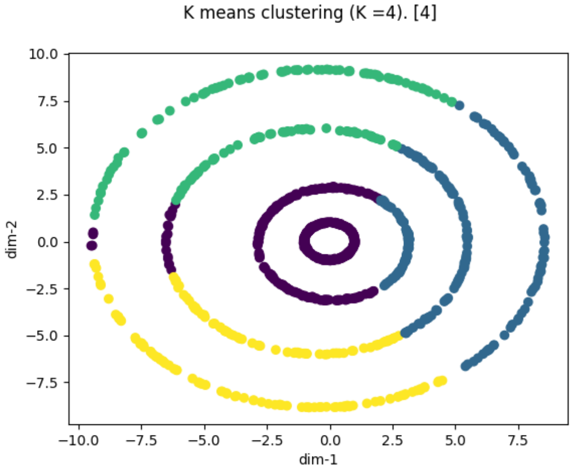


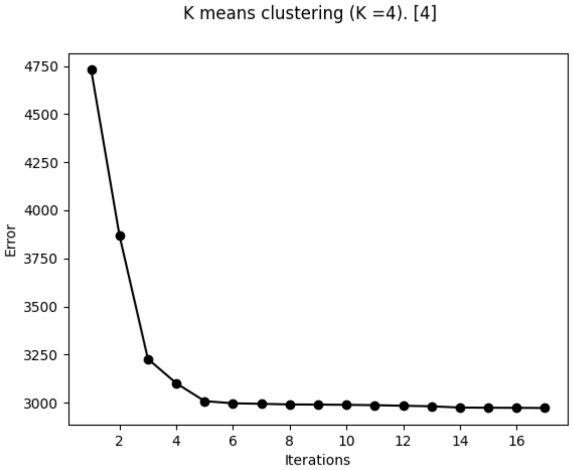
## Initialization-3



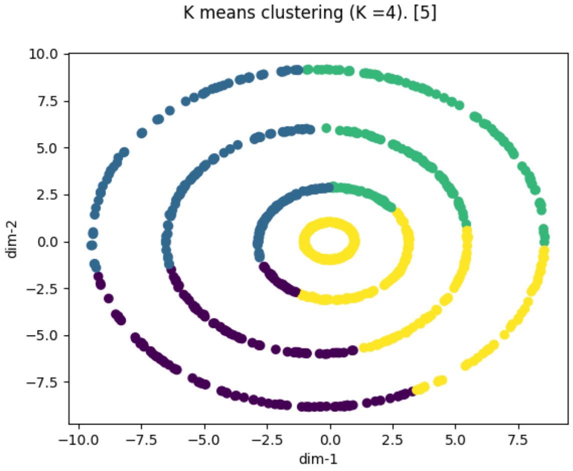


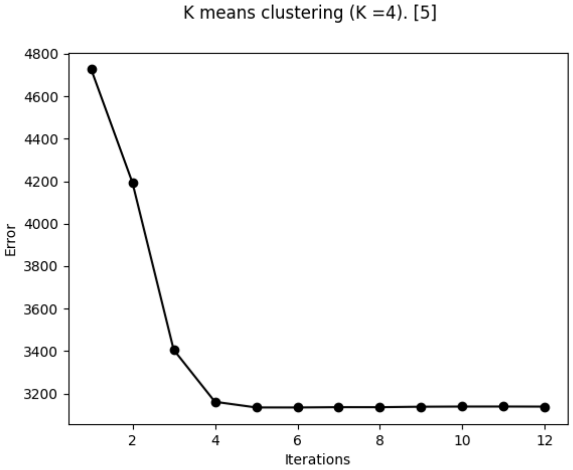
## Initialization-4





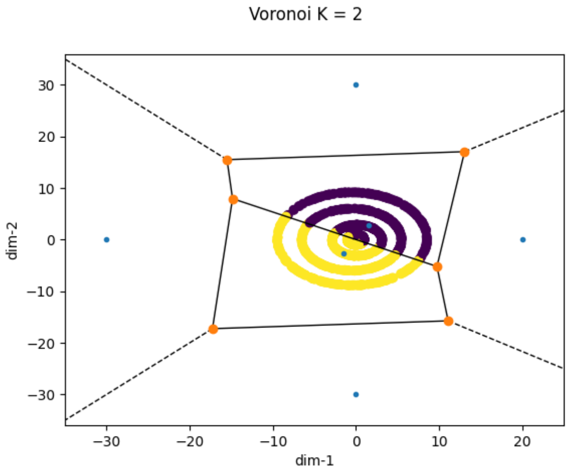
### Initialization-5

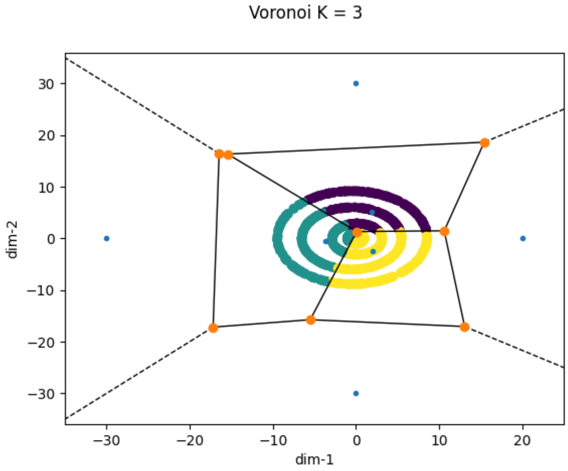



K means clustering (K =4). [5]

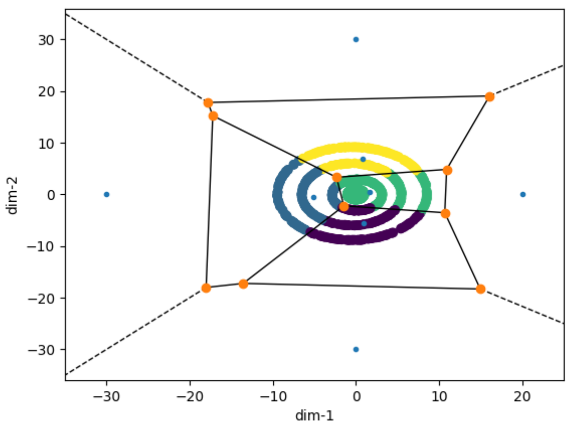
## Q2-(ii)

K-means clustering were performed for $K = \{2, 3, 4, 5\}$ and the datapoint were plotted on top two principal components along with voronoi regions.

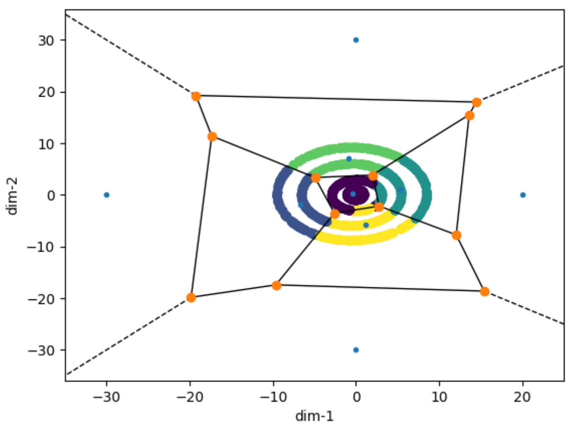

Voronoi K = 2


Voronoi K = 3

Voronoi K = 4



Voronoi K = 5

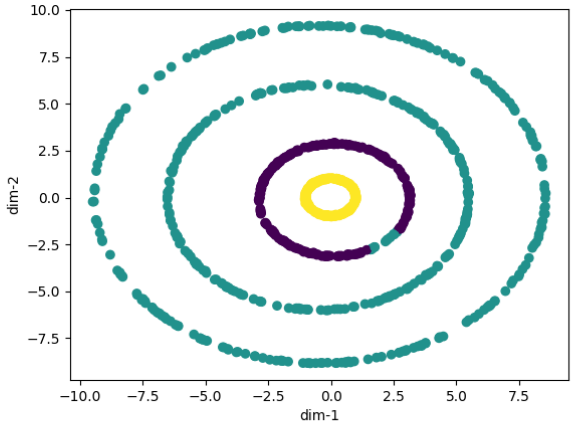


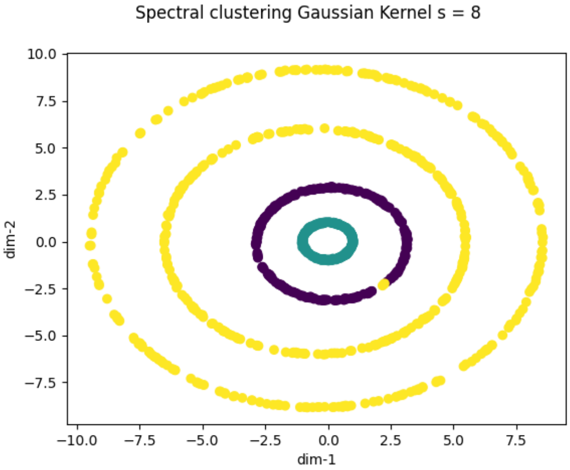
## Q2-iii

**Choose appropriate kernel and plot the data.**

For this data I would choose Gaussian kernel for Gaussian kernels with $\sigma = 0.7$ or $\sigma = 0.8$.

Spectral clustering Gaussian Kernel s = 7





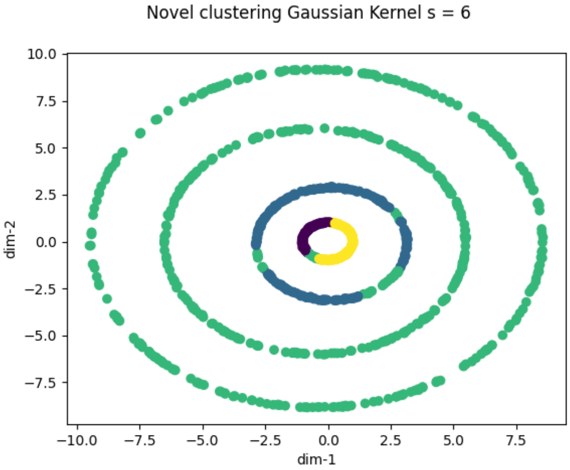
**Explain your choice of kernel based on the output you obtained.**

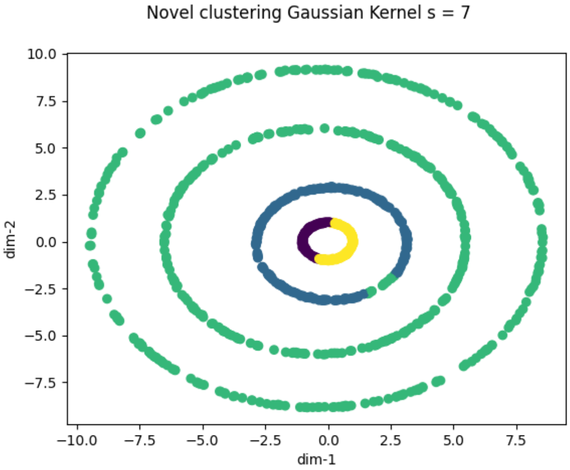
The clusters are plotted below. A good kernel should be able to cluster point in one circle togather. While this does not cluster all the four circles of points, it performs better than the others.

## Q2-(iv)

**How does this mapping perform?**

Clustering was performed using the novel approach described in the assginment Q2-iv. This clustering is comparable to spectral clustering but not as good as spectral clustering. The best clustering results are achieved for Gaussian kernels with $\sigma = 6$ and $\sigma = 7$, as shown in the figures below.





Note that here the inner-most circle is divided into two clusters eventhough there are no such obvious two clusters. Therefore I think that this novel clustering is inferior to spectral clustering.

**Explain your insights.**

I think the reason spectral clustering performs is that it undergoes several rounds of reassignment and therefore achieves refinement of existing cluster allocation. However this takes time. The novel method here achive quite good results with just the first round of allocation. I think that these two methods could be combined to produce high quality of clustering in short time. That is, I think the novel method could be used to initialize the clustering and then spectral clustering can used to refine it. Since we are starting with an already good cluster allocation, the Lloyd's algorithm might converge faster.