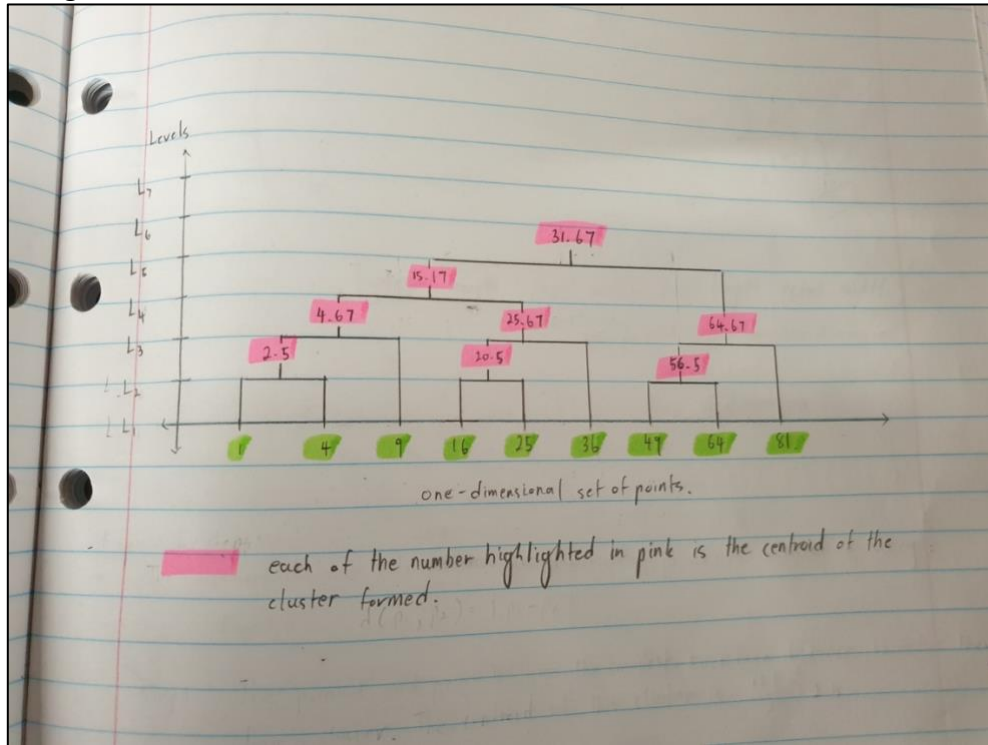


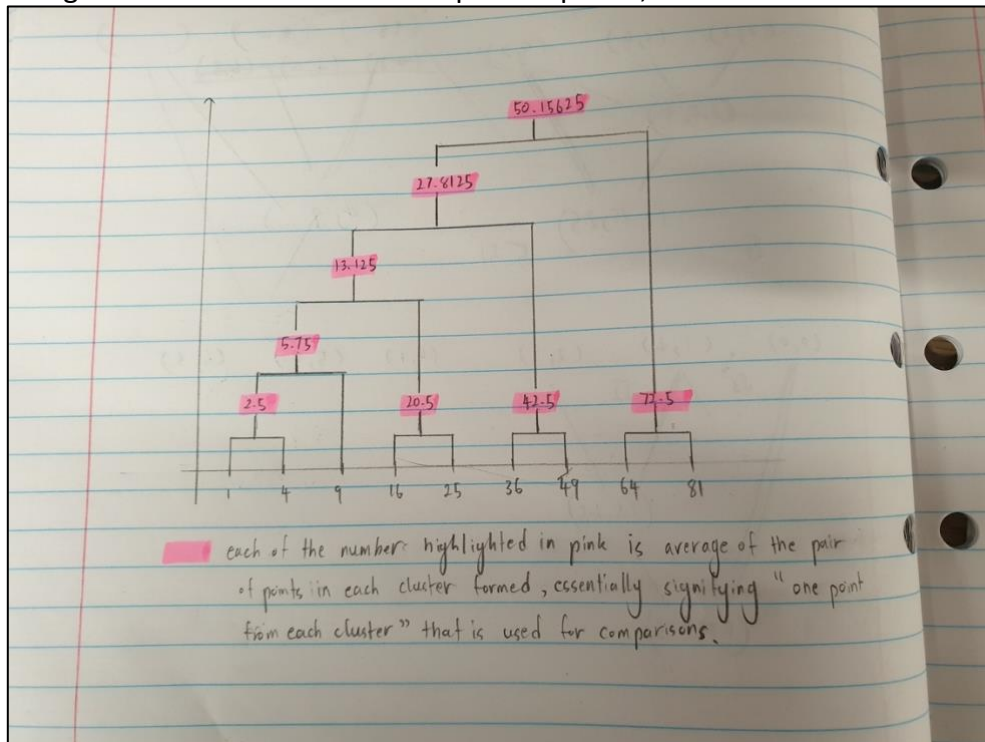
Exercise 2:

Question 1:

- a. Using the minimum the distance-of-centroid rules:



- b. Using the minimum distance of all pairs of points, one from each cluster:



Question 2:

- a. Comparing the before and after K-means clustering plots of the first 2 dimensions of the iris dataset:

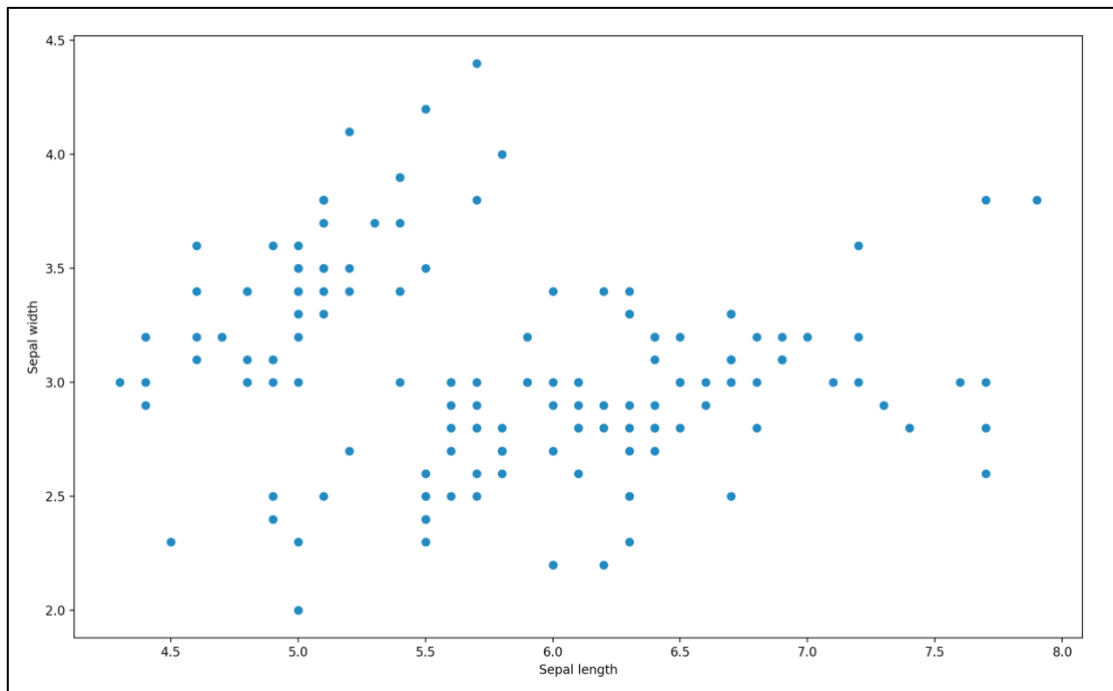


Figure 1: 2-dimensional plot of the first 2 dimensions of the iris dataset where x-axis is the sepal length and y-axis is the sepal width before K-means Clustering is applied.

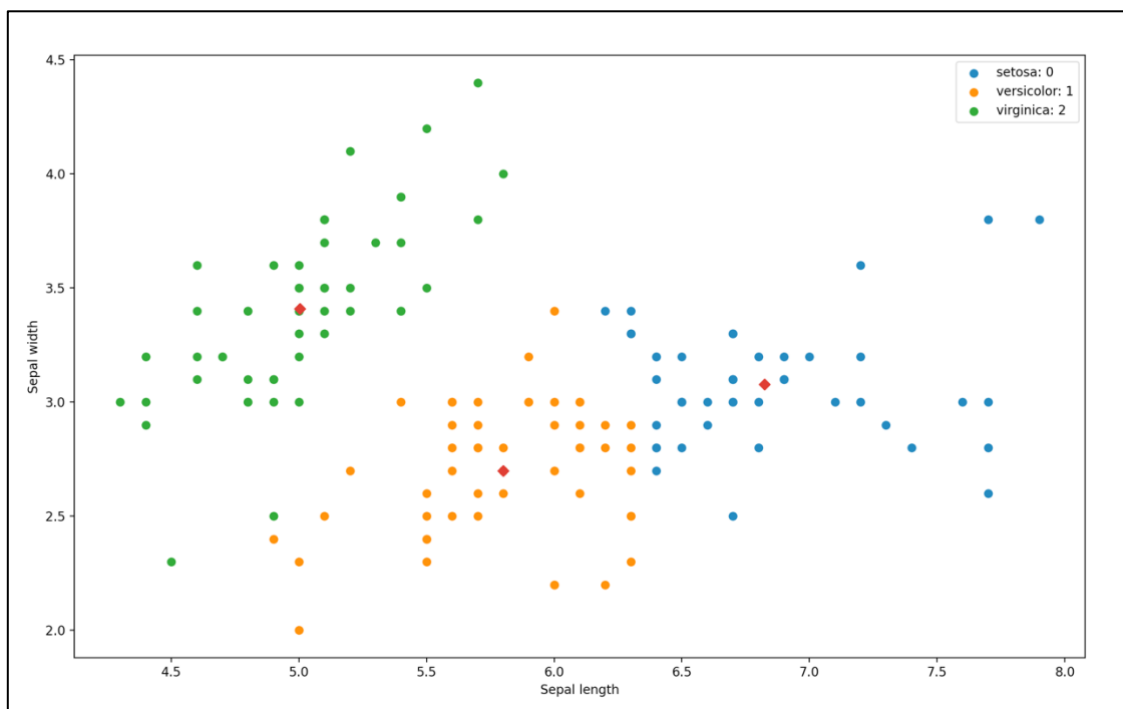


Figure 2: 2-dimensional plot of the first 2 dimensions of the iris dataset where x-axis is the sepal length and y-axis is the sepal width after K-means Clustering is applied. (Converged K-means)

- The red diamonds are the centroids of each cluster.
- **Description about the number of rounds taken for the centroids to converged:**

- In the code, I implemented a while loop that is conditioned to the “treshold of the difference between the old centroid in previous round and the new centroid in the new round”, once the treshold is breached, the centroids will converge with an arbitrary number of rounds due to the nature of the random intialisation of the first centroids.
- Example of the centroids in each round:

Number of Rounds: 6		
Centroid 1	Centroid 2	Centroid 3
[4.4 3.]	[6.2 3.4]	[5.7 2.8]
[4.84210526 3.24736842]	[6.54098361 3.22131148]	[5.75490196 2.71960784]
[4.98627451 3.37843137]	[6.75490196 3.09019608]	[5.78541667 2.68125]
[5.00392157 3.40980392]	[6.81276596 3.07446809]	[5.79038462 2.69615385]
[5.006 3.428]	[6.81276596 3.07446809]	[5.77358491 2.69245283]
[5.006 3.428]	[6.81276596 3.07446809]	[5.77358491 2.69245283]

- b. The k value or better known as the number of clusters that I’ve picked is 3 due to prior knowledge of the iris dataset. This prior is obtained from scikit-learn datasets library where it specifies that there are 3 classes that were being classified using the

sklearn.datasets.load_iris

`sklearn.datasets.load_iris(*, return_X_y=False, as_frame=False)` [\[source\]](#)

Load and return the iris dataset (classification).

The iris dataset is a classic and very easy multi-class classification dataset.

Classes	3
Samples per class	50
Samples total	150
Dimensionality	4
Features	real, positive

[Read more in the User Guide.](#)

iris dataset.

- Class 0 is the classification for iris setosa.
- Class 1 is the classification for iris versicolor.
- Class 2 is the classification for iris virginica.

Exercise 3:

- The worst-case that can happen for the 3 advertisers A,B and C on the query stream of xxyzz for the greedy algorithm are as follows:

- BCC_ _
- CCBB_ _
- BCBC_ _
- CBCB_ _
- CBBC_ _
- BCCB_ _

By exhausting the budget of the advertiser B or C first will guarantee that the greedy algorithm will assign at least 4 out of the 6 queries. Therefore, the worst-case of the greedy algorithm is proven.

- Another sequence of queries that will guarantee that the greedy algorithm will assign as few as half the queries of an optimal algorithm would be as follow:
 - xxzz or yyzz

- i. If advertiser C were to be assigned as the first 2 advertisers for the sequence, then advertisers B and A will not be selected for the rest of the queries because query z is not bid by both of them.
 - 1. CC__