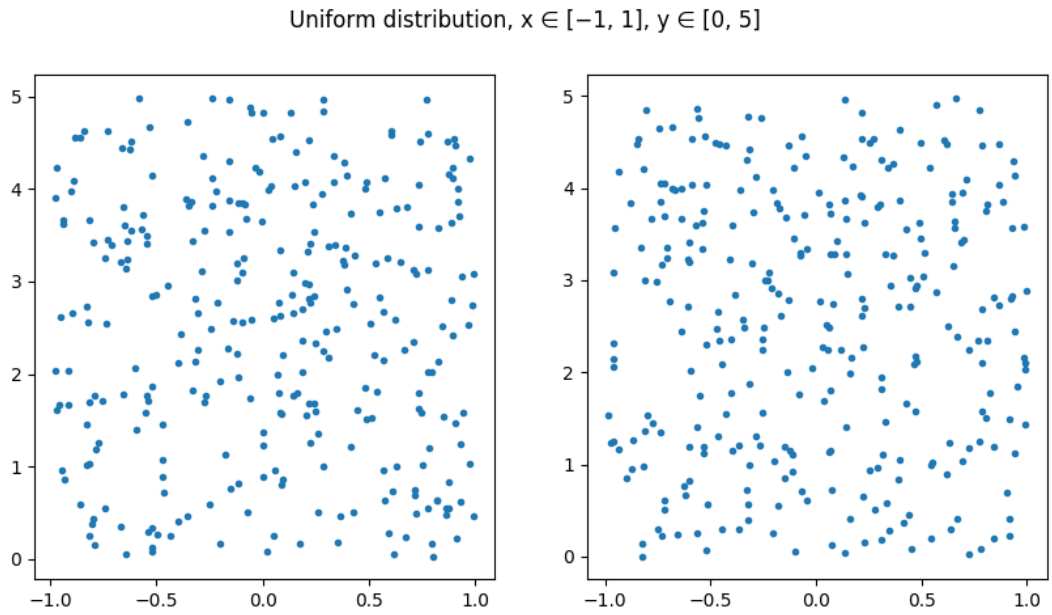
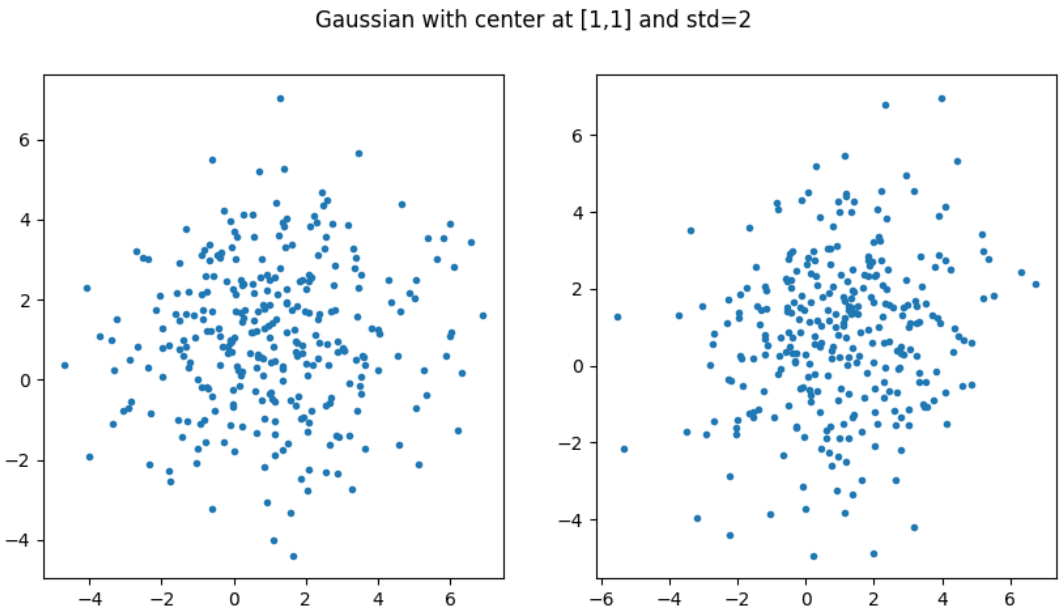
Homework #2: Similar items, Clustering, Community Detection

Problem 3

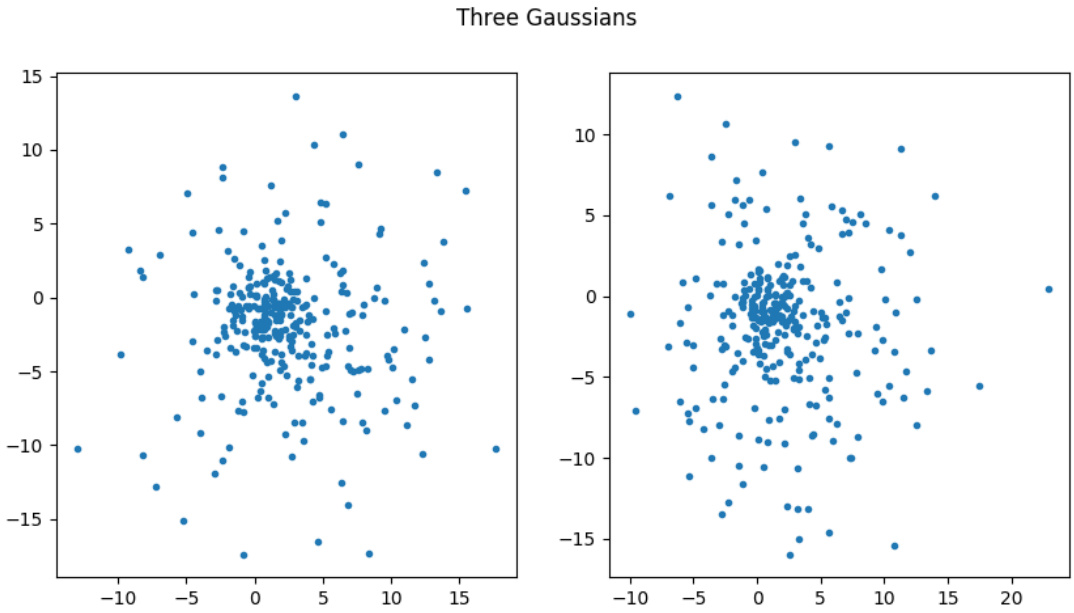
(a)



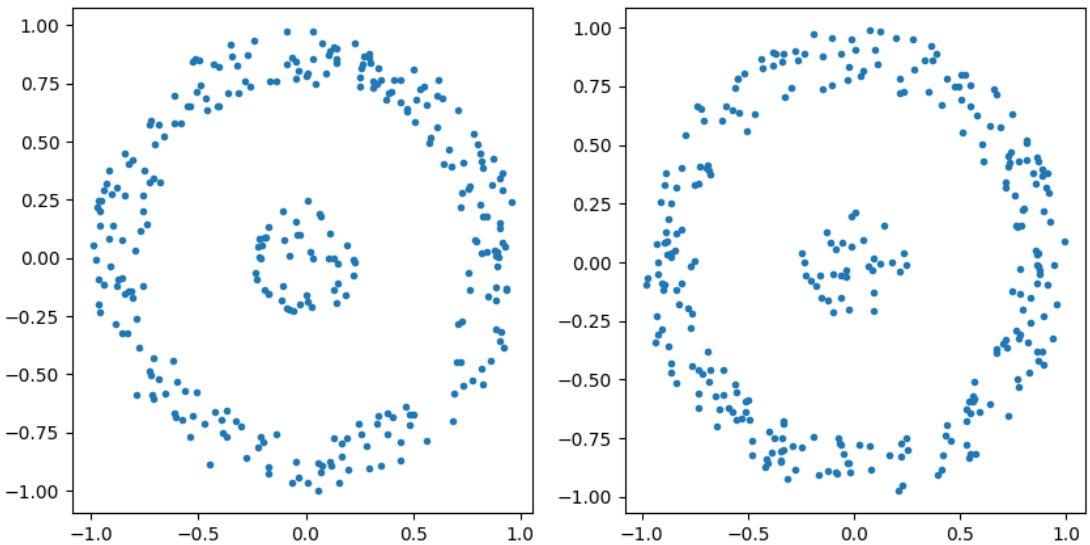
(b)



(c)



(d)

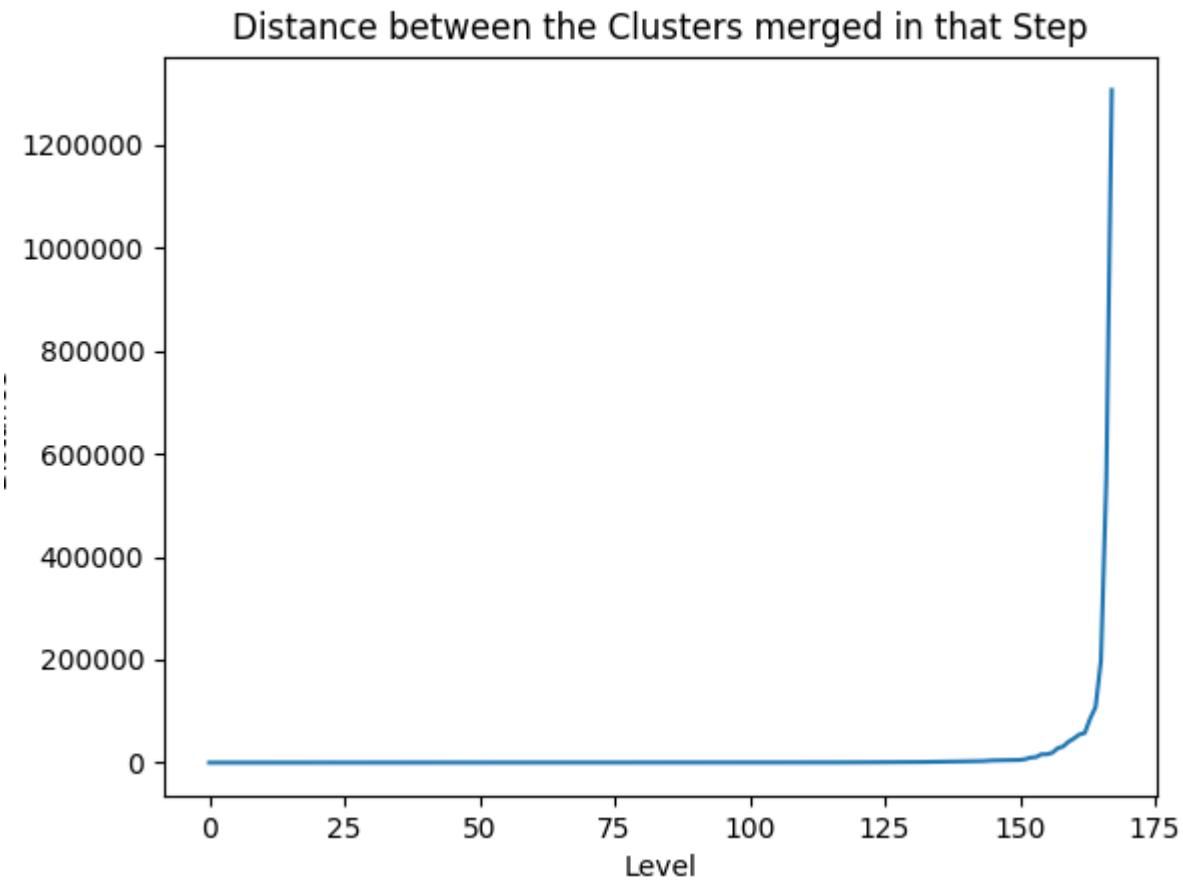


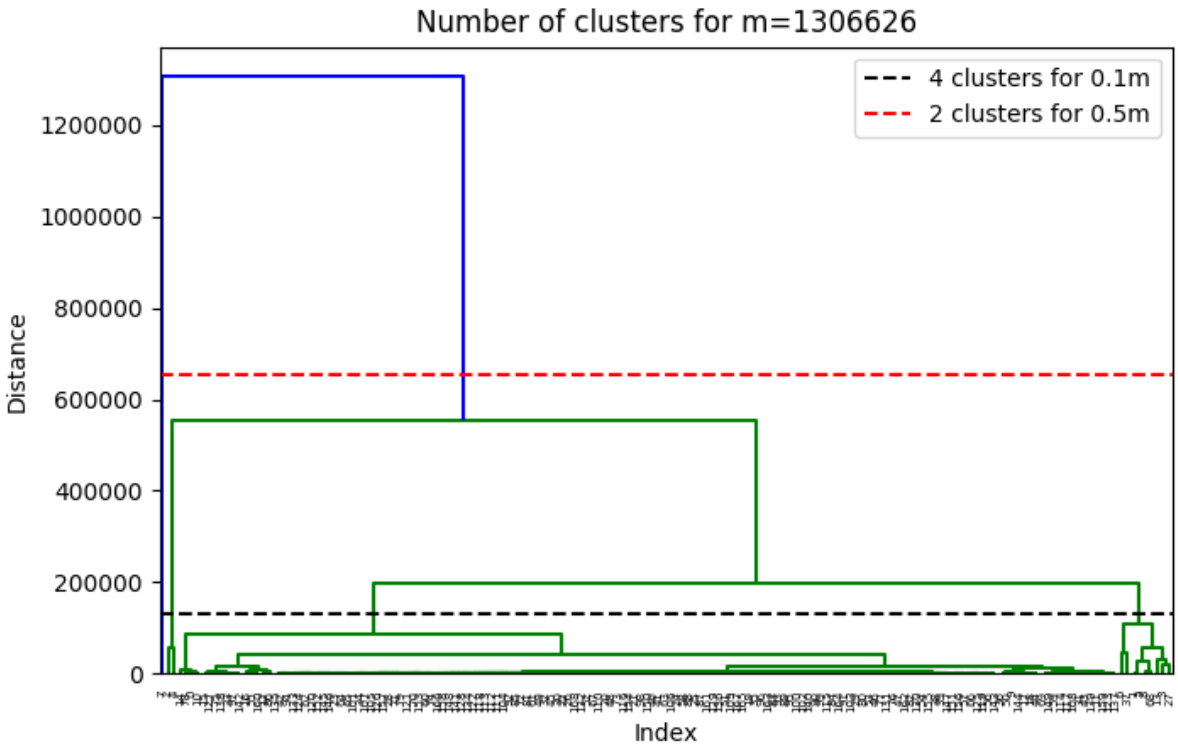
(e)

Problem 4

For hierarchical clustering and the plot of the dendrogram we use the package `scipy.cluster.hierarchy`. Function `linkage` provides the hierarchical clustering algorithm for different methods. We use `method='centroid'` in part (a) and `method='complete'` in part (b). Function `dendrogram` is used to plot the dendrogram.

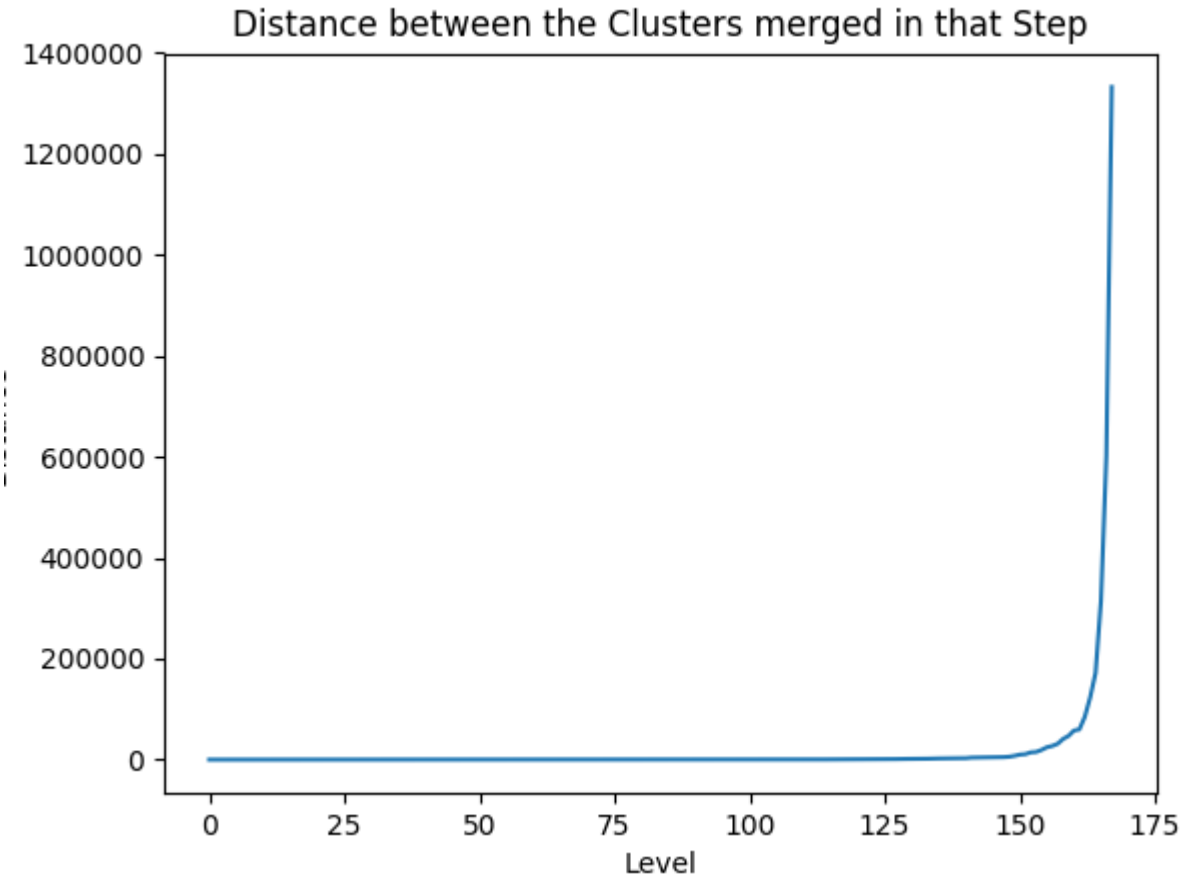
(a) merge clusters with the closest centroids

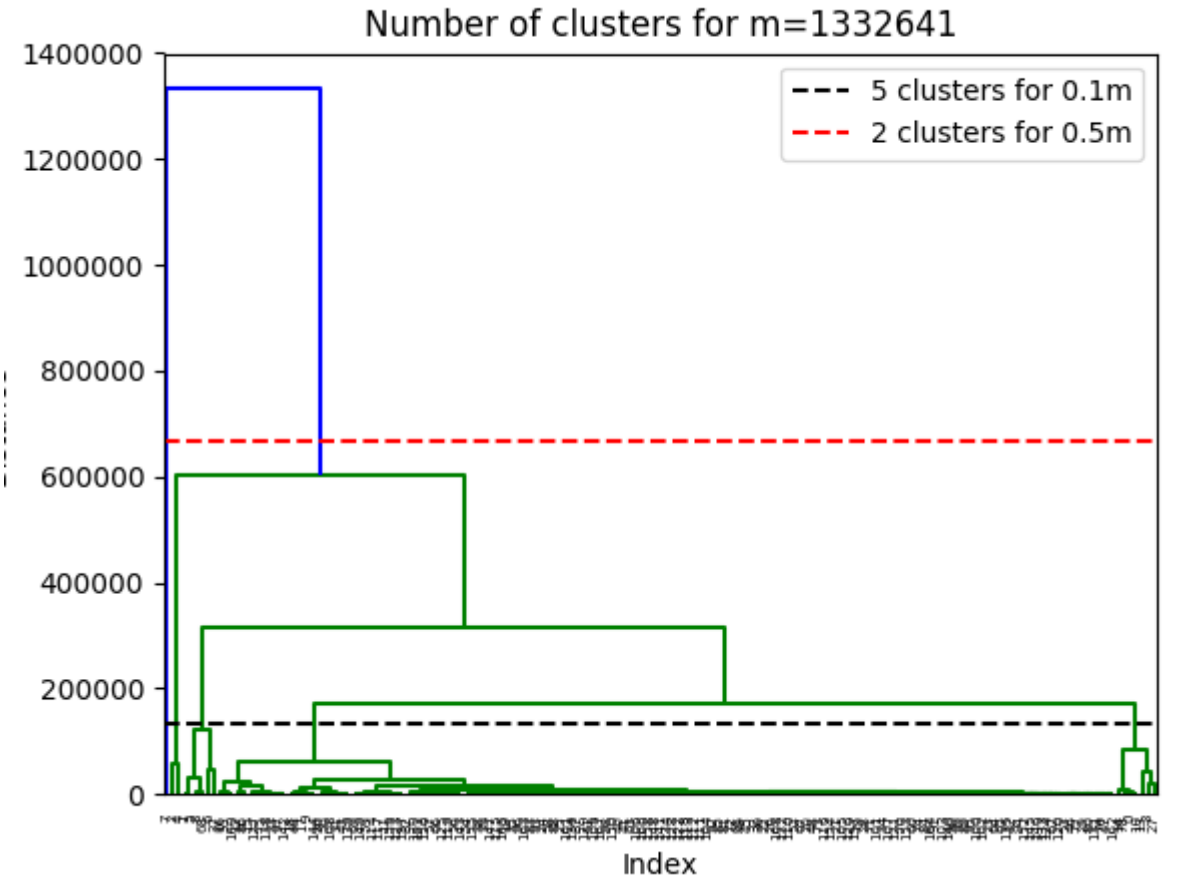




The Dendrogram shows that we would get 4 clusters if we decide to stop clustering at a distange of 0.1m and 2 clusters if we decide to stop at 0.5m.

(b) merge clusters so that the new diameter is the smallest among all options





The Dendrogram shows that with this method we would get 5 clusters if we decide to stop clustering at a distange of 0.1m and 2 clusters if we decide to stop at 0.5m.