

Introduction to Machine Learning

Assignment 3

Group 31

Stijn Kammer (s4986296) & Ramon Kits (s5440769)

October 5, 2022

INTRODUCTION

METHODS

The method used in this assignment is the use of the Density-based spatial clustering of applications with noise algorithm. This algorithm is a density based clustering algorithm. It is a density based clustering algorithm because it uses the density of points in a region to determine if they belong to the same cluster. The algorithm is based on two parameters, the minimum number of points in a region and the maximum distance between points in a region. The algorithm works as follows:

1. For each point in the dataset, determine the number of points in a region around it.
2. If the number of points in the region is larger than the minimum number of points, the point is a core point.
3. If the number of points in the region is smaller than the minimum number of points and it belongs to the neighbors of a core point, it is a border point.
4. If the number of points in the region is smaller than the minimum number of points and it does not belong to the neighbors of a core point, it is a noise point.
5. For each core point, determine the cluster it belongs to.
6. For each border point, determine the cluster it belongs to.

this algorithm has a few advantages. you can use it to find clusters of any shape and size. There is no need to specify the number of clusters beforehand. And is also very robust to noise and outliers. This algorithm works best when you choose the minimum number of points based on the size and noise of the dataset. For two-dimensional data, the use of 4 as minimum number of points is recommended. for higher dimensional data, the use of two times the dimensionality of the data is recommended.

EXPERIMENTAL RESULTS

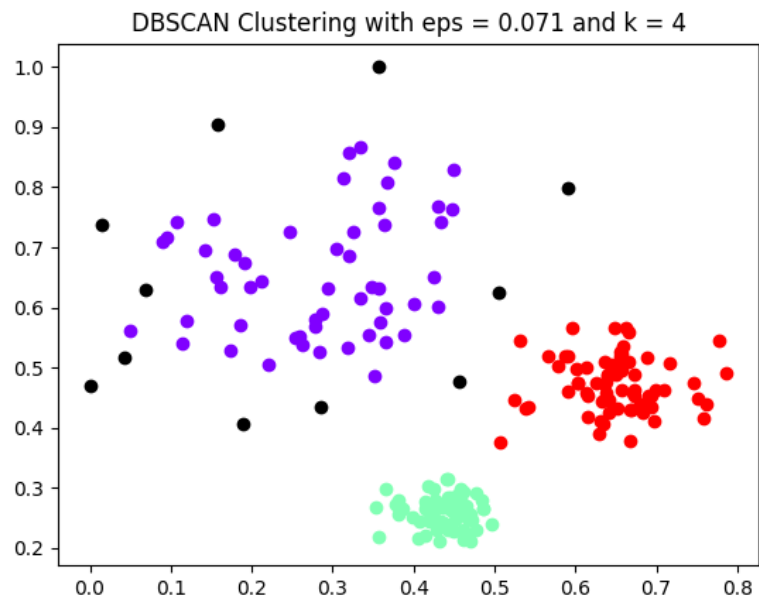


Figure 1: DBSCAN with $\text{eps}=0.071$ and $k=4$

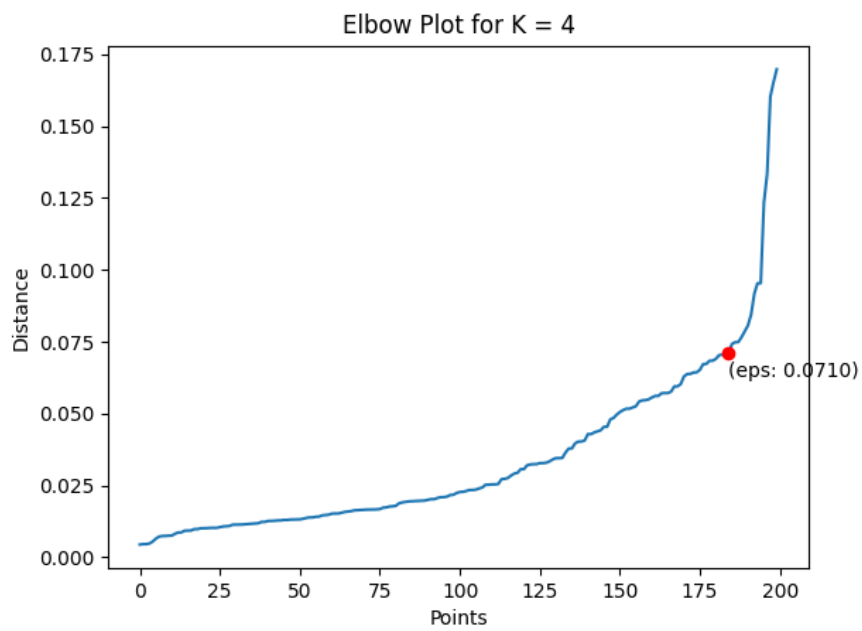


Figure 2: Elbow plot with $k=4$

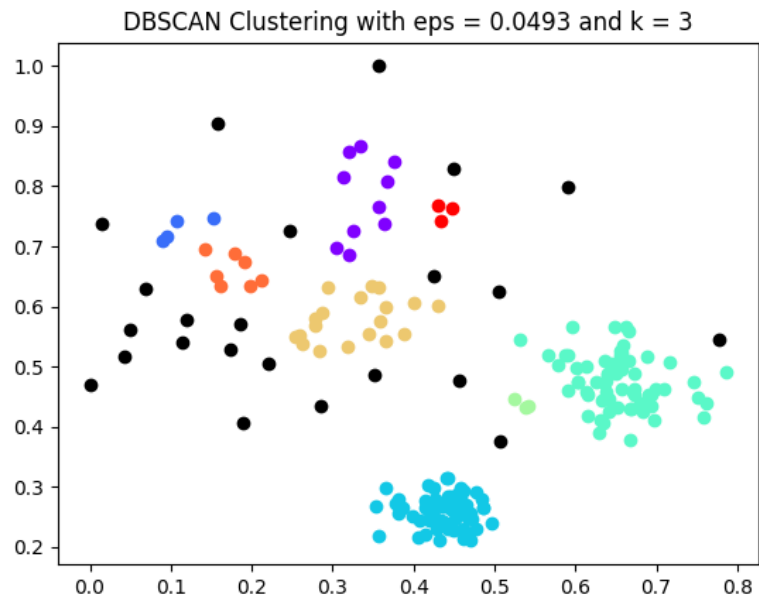


Figure 3: *DBSCAN with $\text{eps}=0.0493$ and $k=3$*

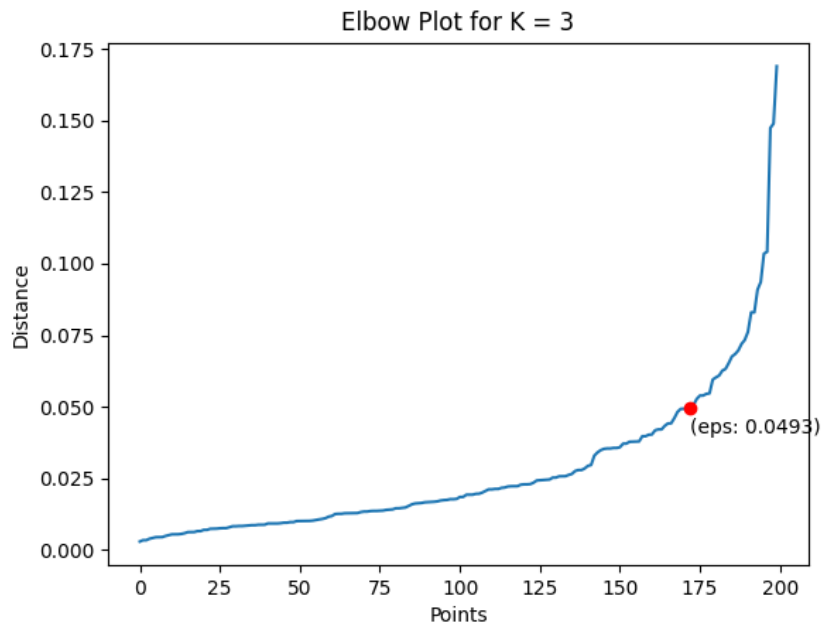


Figure 4: *Elbow plot with $k=3$*

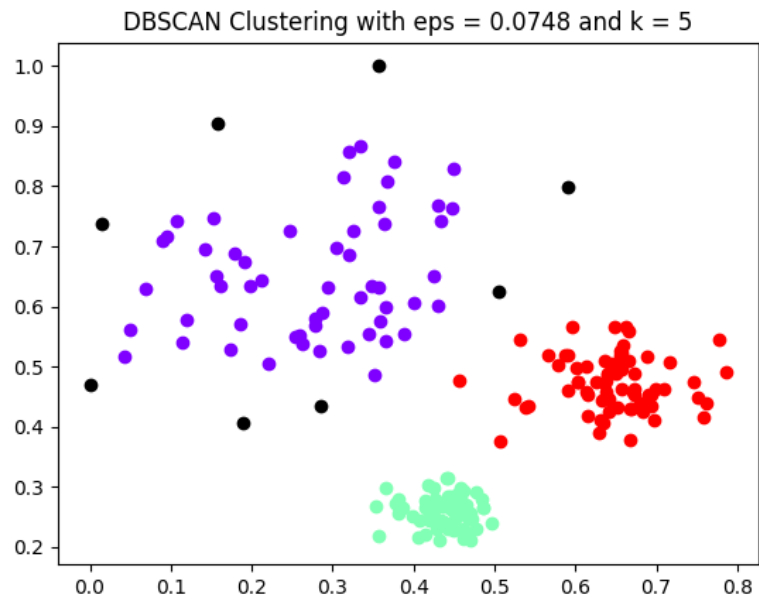


Figure 5: DBSCAN with $\text{eps}=0.0748$ and $k=5$

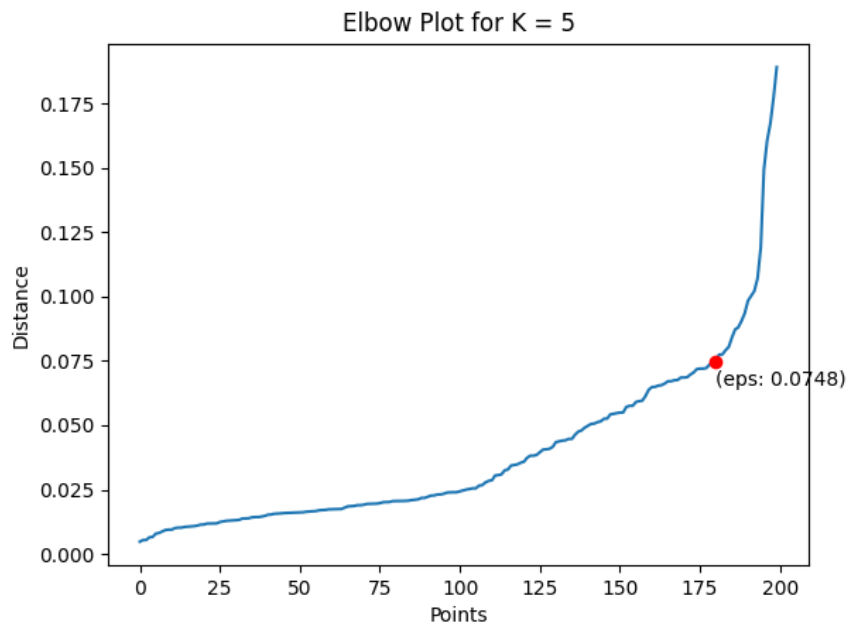


Figure 6: Elbow plot with $k=5$

DISCUSSION