CLUSTERING : DBSCAN

ESRA NUR ERSAN

HANYANG UNIVERSITY

Department of Computer Science &

Software Engineering

9325020182

DBSCAN Algorithm

Consider a set of points in some space to be clustered. Let ε be a parameter specifying the radius of a neighborhood with respect to some point. For the purpose of DBSCAN clustering, the points are classified as core points, (density-)reachable points and outliers.

A point p is a core point if at least minPts points are within distance ε of it (including p).

A point q is directly reachable from p if point q is within distance ε from core point p. Points are only said to be directly reachable from core points.

A point q is reachable from p if there is a path p_1, ..., p_n with p_1 = p and p_n = q, where each p_{i+1} is directly reachable from p_i. Note that this implies that all points on the path must be core points, with the possible exception of q.

All points not reachable from any other point are outliers or noise points.

Now if p is a core point, then it forms a cluster together with all points (core or non-core) that are reachable from it. Each cluster contains at least one core point; non-core points can be part of a cluster, but they form its "edge", since they cannot be used to reach more points.

A cluster then satisfies two properties which is all points within the cluster are mutually density-connected and if a point is density-reachable from any point of the cluster, it is part of the cluster as well.

DETAILED DESCRIPTION OF MY CODE

(Note: Actually i code the algorithm in python language first but i had so many problems so i decided to try using c++.)

Struct Point : For each point in dataset. Includes distance function to calculate distance between two points.

Class Cluster –> parameters are our dataset,n,eps and minpts values

FindNeighbors() : To find the nearest points according to results of comparison between their distance and eps value

CorePoint() : To decide whether the point we are checking is core or not.

ExpandCluster() : Check situation among the points in our dataset and to decide if the point we start with is core point or not so we know if we have to extend cluster or not.

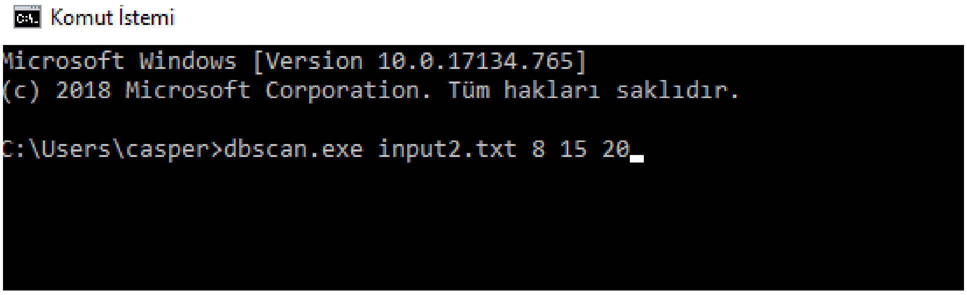
dbscan(): If p is a core point a cluster is formed. If point is a border point, no points are density reachable from point and dbscan visits the next point of database.Continue the process until all of the points have been processed.

ReadFile()

ShowResults()

How to execute my program

#/[dbscan.exe] [inputfile.txt] [n] [eps] [minpts]



```
Microsoft Windows [Version 10.0.17134.765]
(c) 2018 Microsoft Corporation. Tüm hakları saklıdır.

C:\Users\casper>dbscan.exe input2.txt 8 15 20
```