2014057983 컴퓨터공학과 지소현

1. Summary of Algorithm

-Discovers clusters of arbitrary shape : Density-Based-Spatial clustering of Applications with Noise.

-A cluster is defined as a maximal set of density connected points.

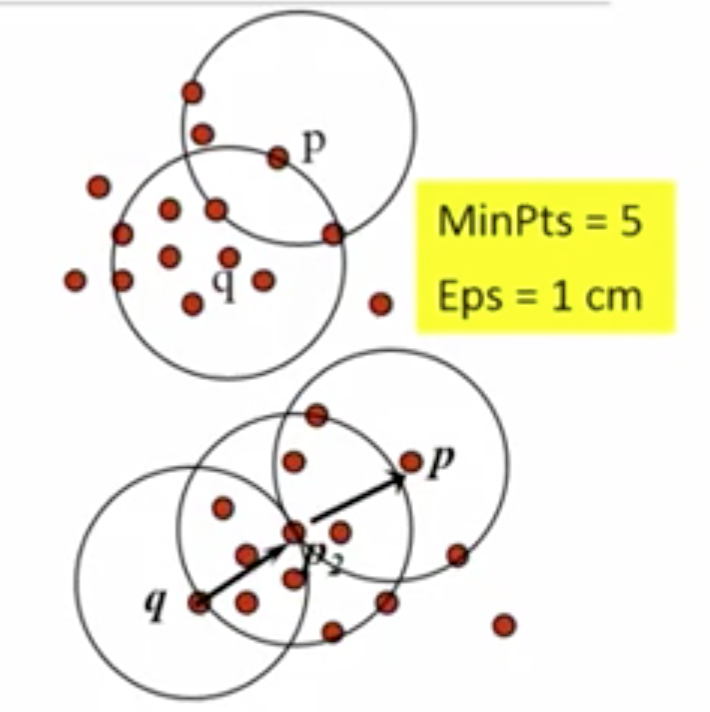
(1) Arbitrarily select a point P.

(2) Retrieve all point density-reachable.

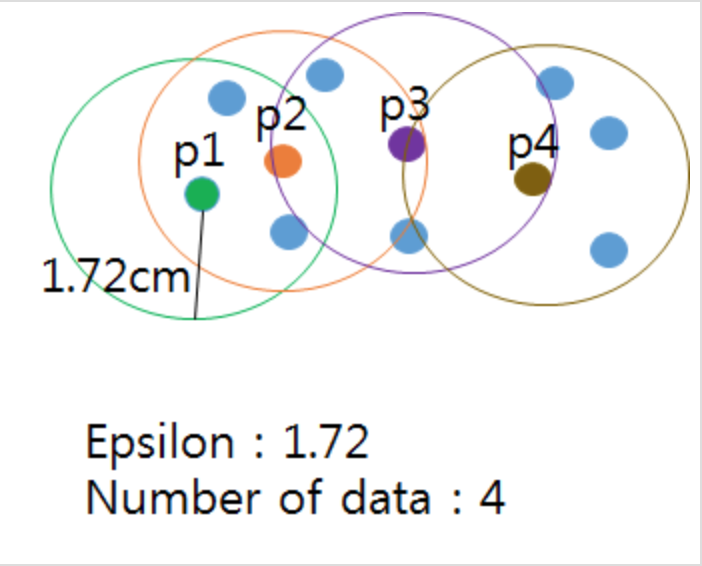
-If p is core point, cluster is formed.

-If p is a border point, no points are density-reachable from p, and DBSCAN visits the next point of database.

(3) Continue the process until all of the points have been processed.



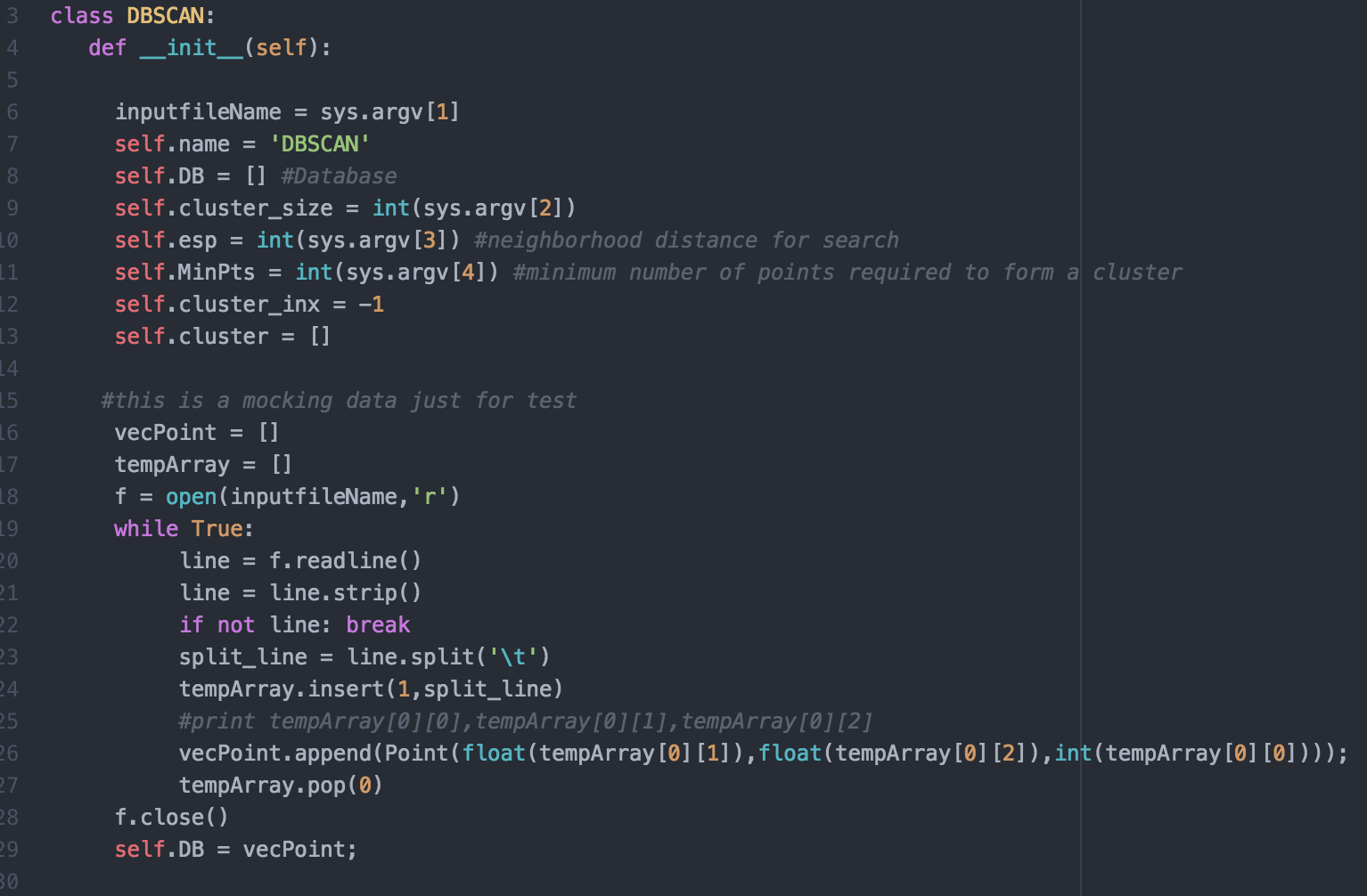
🡪 A point p is density-reachable from a point q w.r.t Eps,MinPts if there is a chain of points p1,p2,….,pn,p1=q,pn=p such that p(i+1) is directly density-rechable from p(i).



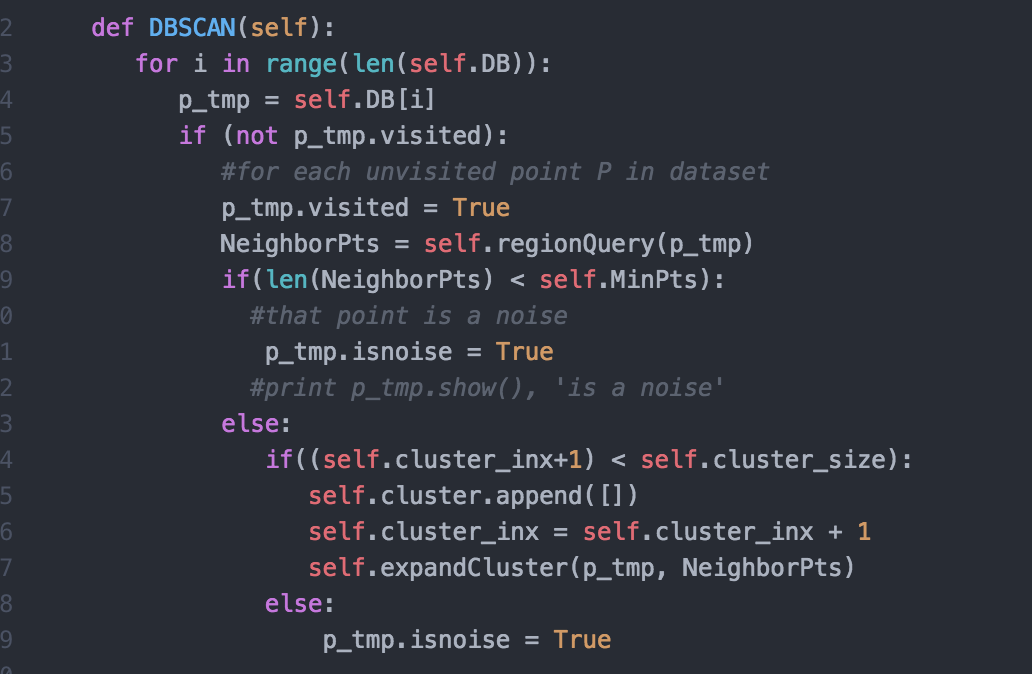
🡪 This picture is one of cluster that is ‘density-connected’ . We can expand cluster using this property.

1. Detailed description of codes
2. Class DBSCAN :

It has DB(input file dataset), name, cluster\_size(number of clusters for the corresponding input data…That is,cluster Maximum Size),eps(epsilon), MinPts(Minimum number of points in an Eps-neighborhood of a given input),cluster index, and cluster list.It must get and save DB from the input file data.



1. DBSCAN Function.. It check one by one from the DB list and save into ‘p\_tmp’. If ‘p\_tmp’ is not visited, get the Neighbor points of ‘p\_tmp’. When getting the Neighbor points of ‘p\_tmp’, its number of points is less than MinPts, then It change ‘isnoise’ into True. Otherwise, add new cluster and expand cluster from ‘p\_tmp’ and Neghborhood of ‘p\_tmp’.

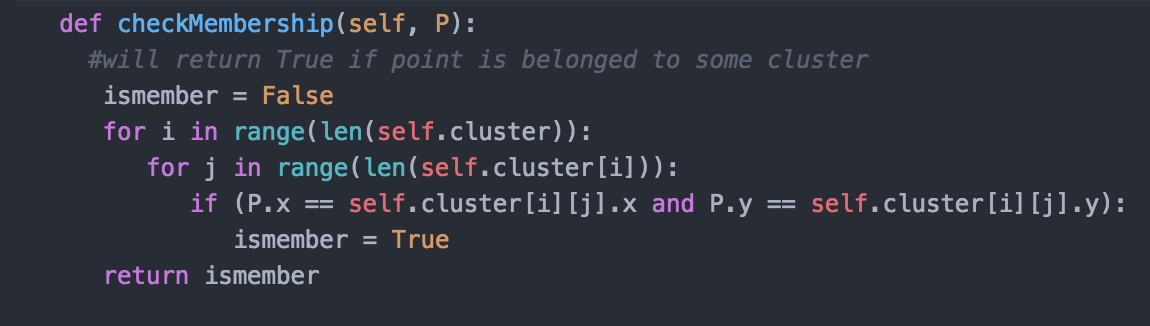


1. ExpandCluster Function iterates all neighbor points.

Each negibor point is checked ‘isVisited’. If it is not visited, then get its neighbor points using ‘regionQuery’ function. Also, it is check number of neighbor points and member of any cluster.

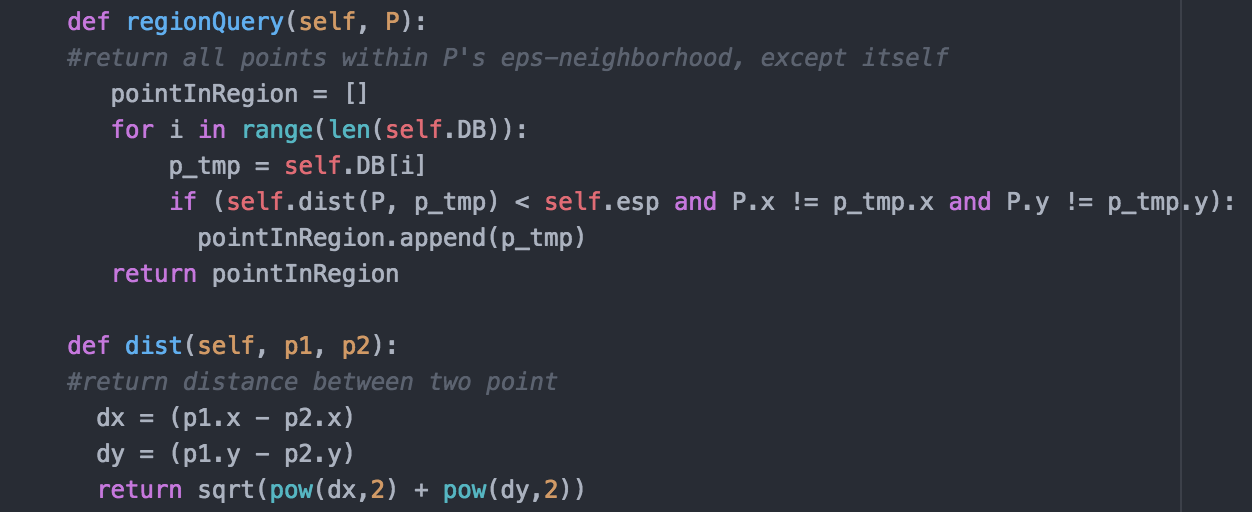


1. checkMembership Function will return ‘true’ if point is belonged to some cluster. It use for check member of cluster if it is not memember of any cluster, then add new cluster.



1. regionQuery Function is for getting all neighborhood of points within P’s(parameter..one of point) epsilon.

It calculates distance function.(2 root..a^2 + b^2 )



1. Main Function calls DBSCAN fuction and save result(object\_id of each cluster) into outputfile.



1. Insrtuctions for compiling my source code(e.g screenshot)

Use this command…

**python dbscan.py input\_data\_File\_Name n epsilon minPoints**

* 🡪 ***parameter…*** - ***n***: number of clusters for the corresponding input data
* - ***Eps***: maximum radius of the neighborhood
* - ***MinPts***: minimum number of points in an Eps-neighborhood of a given point

