Data Science

Assignment 3 Clustering

Prof. Sangwook Kim  
TAs. Jangwan Koo, Yeonchang Lee

2013011424  
Yedarm Seong  
Division of Computer Science and Engineering

1. Environment

OS: Windows 10  
Language: Python 3.5.2

2. Summary of algorithm

(1) region query

P <- Target point  
N <- Empty list  
for point in data set  
 if the point is a neighbor of P  
 N.add(point)

return N

(2) expand cluster

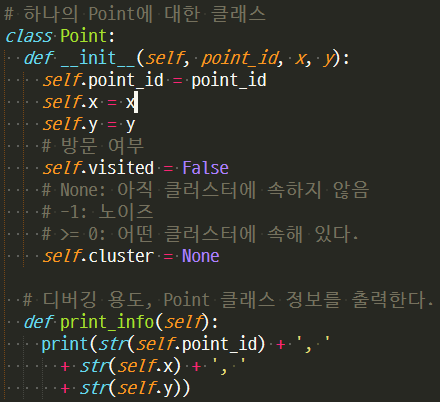
C <- Target cluster  
for n in N  
 C.add(n)  
 if n is not visited  
 n.visited <- True  
 new\_n <- region\_query(n)  
 if new\_n is satisfied min pts  
 N.add(new\_n)  
 if n is not in the cluster  
 C.add(n)

(3) dbscan

for point in dataset  
 if point is not visited  
 point.visited <- True  
 n <- region\_query(point)  
 if n is satisfied in pts  
 expand\_cluster(n)  
 else  
 n is noise

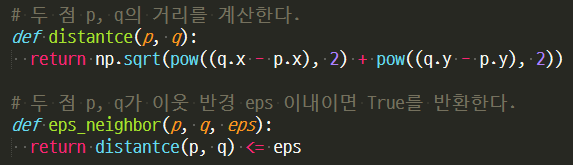
3. Detailed description of codes (for each function)

Point class



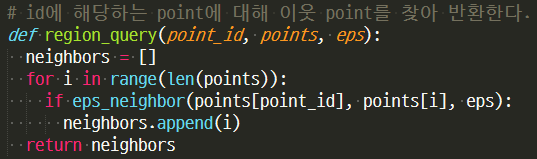
The point class has the ID of each point and the x, y coordinates, visited, and clusters to which it belongs.  
If the value of the cluster member variable is -1, it is noise.

Calculate distance & is Determine neighbor



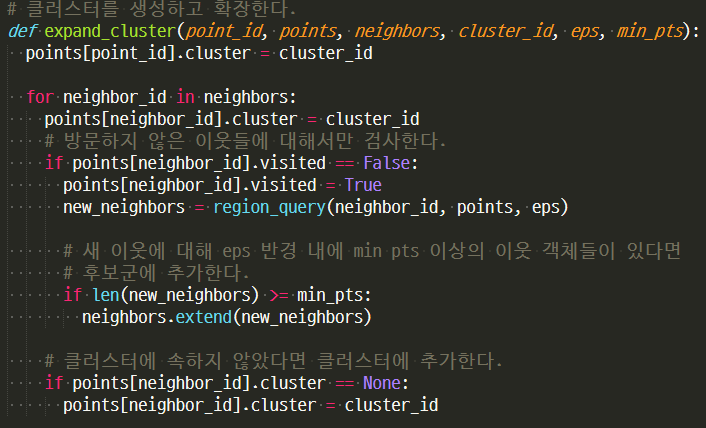
Calculate the Euclidean distance.  
Determine if the two input points are neighbors.

Get neighbors



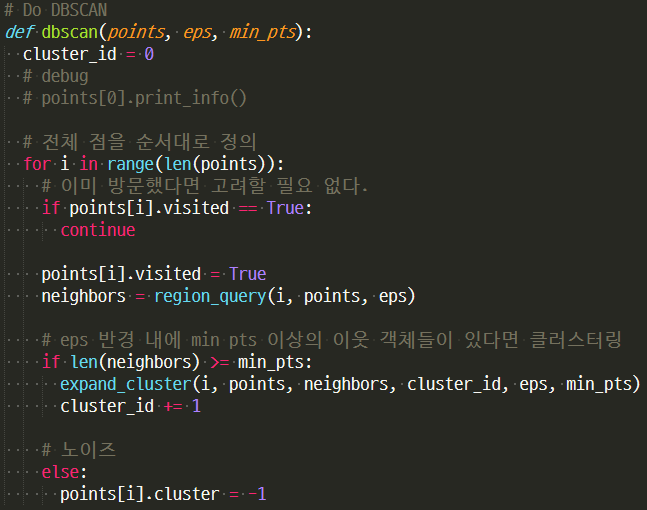
Returns the neighbor of the input point id and returns it.

Expanding clusters



Add the input point id to cluster C, find a new neighbor for the neighbors of point id, and add it to cluster C if the neighbors satisfy the condition.

Doing DBSCAN algorithm



Compute the neighbors for points in the entire dataset, and expand the cluster if the conditions are met for neighbors.  
If the condition is not satisfied, it is regarded as noise.

4. Instructions for compiling source codes at TA's computer (e.g. screenshot) (*Important!!*)

**Python does not need a compile process**

If you already install pyinstaller for python3, in the directory where the clustering.py file is located, type the following command:

Windows:  
pyinstaller -F clustering.py

Ubuntu:  
pyinstaller -F clustering.py

Then the dist folder is created in the directory where the command is executed, and there is ‘clustering.exe’ file in it.(For linux, the ‘clustering’ file)

Usage: clustering.exe [input file] [number of cluster] [epsilon] [min pts]

After using pyinstaller, there is an exe file in the dist folder:  
ex) dist/clutsering.exe [input file] [number of cluster] [epsilon] [min pts]

If you move the exe file to the same location as input file, you can use it like this:  
ex) clutsering.exe [input file] [number of cluster] [epsilon] [min pts]

5. Any other specification of implementation and testing

Another execution method:  
Running as a Python file

ex) clutsering.py [input file] [number of cluster] [epsilon] [min pts]