**Assignment3. DBSCAN**

2017029561 Nam Jihoon

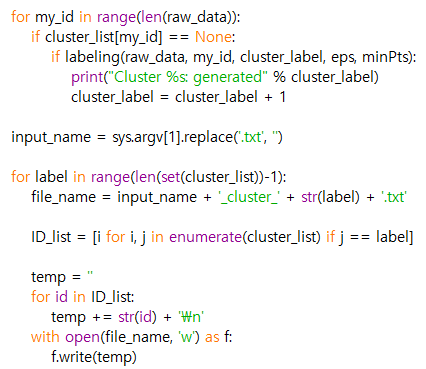
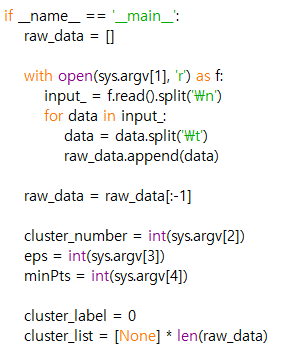
1. **Summary**

DBSCAN stands for 'Density-based spatial clustering of applications with noise.' The method of classifying data is a kind of clustering, and clustering according to density is called DBSCAN.

Because data is grouped and classified, there is no need for labeling for each data. Therefore, this method is also called 'unsupervised learning'.

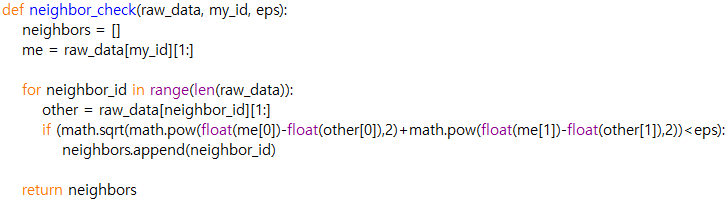
After presenting all the data in the coordinates, determine how many neighbors are at a certain distance for each point. If a neighbor exists above a critical point, that point is called a 'core point' and plays an important role as a criterion for determining clusters. Even if it is not a 'core point', if there is a core point in the neighbor's data, it belongs to the same cluster. The clusters are set based on these dense areas, and the points that cannot be 'core points' are called 'border points'.

1. **Functions description**
2. **Main function**



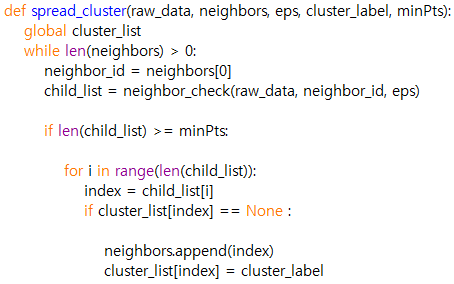
It is the main function that is automatically called as soon as the program starts. Parameters can be received through a console stored in sys.argv. In this task, it is used to address input and output files. The periodic function is to receive the address of the input file, process the data, and then read the input file to save it in the list. Subsequently, to configure and utilize dbscan, another function is called and the result is saved to the specified output file address.

1. **Neighbor\_check()**



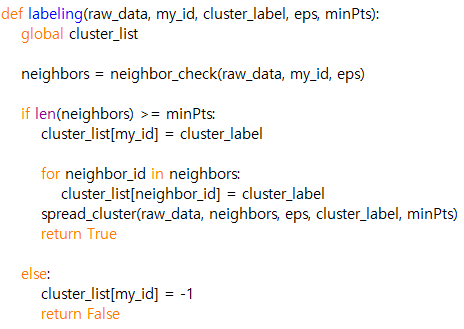
This function is a function that finds and places all neighbors in the data of that ID in the array. Returns the list of all data closer than the critical distance of eps in an array.

1. **Spread\_cluster()**



It is a 'spread\_cluster()' function that allows a neighbor's other neighbors to spread to the same cluster when the neighborhood is cluster by different core points. When a cluster of data is determined, it works by finding a neighbor of each data and setting a new cluster.

1. **Labeling()**



Because it is 'unsupervised learning', labelling is carried out during data analysis. It works by giving a label when a neighbor detects a lot of core points.

1. **Compiling Instruction**

The Program written by Python Idle and used Python 3.7.2.

This program was tested on the Windows 10 command prompt

[1] download dbscan.py and input text files in same directory

[2] get in to directory that dbscan.py is downloaded and enter the command “clustering.py input1.txt 8 15 22”

[3] repeat procedure[2] for input2.txt and input3.txt

[4] To check the test result, enter the command “PA3.exe input1”, “PA3.exe input2”, “PA3.exe input3”

1. **Result**

