DATASCIENCE

Assignment3

2011003865 박준형

1. DBSCAN Algorithm

DBSCAN is one of clustering algorithm. DBSCAN need two parameter, eps and minPts. Eps means maximum distance of points. If the number of points that belongs to one object is more than minPts, the object could compose a cluster. In DBSCAN, scan all points in DB and check whether the point could be make core object or not. Check distance between center point and other points, after that, calculate the number of the points that distance is smaller than eps. If number of points is more than minPts, the point could be in a cluster and check neighbor point. if number of points is smaller than minPts and not be overlapped with other cluster, it is noise. If the point is overlapped with other cluster, it going to belong the cluster.



This picture explain DBSCAN. The point colored red, it represent core object, yellow point is neighbor object, blue point is noise.

1. Code description

There are 4 functions in my code.

* DBSCAN function

Check all points in dataset if the point could be make cluster, call expandCluster function, the other case, point is treated like noise.

* regionQuery

Calculate distance between center point and the other points. This function returns list of neighbor points.

* Diff

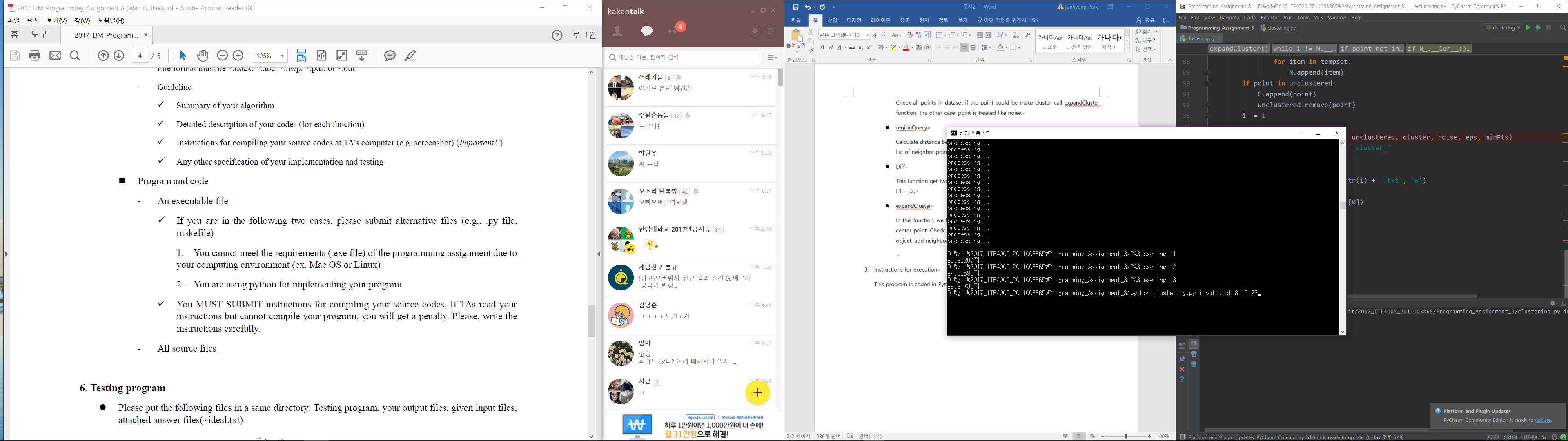
This function get two lists and return list of different elements. Operate like return L1 – L2.

* expandCluster

In this function, we have center point and expand cluster using neighbor point of center point. Check all neighbor points and if the point could be center of core object, add neighbor of the point. Like this way, clusters can be expanded

1. Instructions for execution

This program is coded in Python version 3.5 and could executed using command like this.



It needs python interpreter and 4 parameters, input file name, number of clusters, eps, minPts.

1. Result of algorithm (Test)

