2011021764
강병욱
2017년 5월 17일

Assignment3
DBSCAN

1. Algorithm summary

DBSCAN has two parameters, eps(dense region radius) and minPts(minimum number of points to form a dense region). It starts with random starting object that has not been visited. Then retrieve all points that is closer than eps and if it contains sufficiently many points, a cluster is started and points is included in the cluster. Else, the point is labeled as noise.

From each point in cluster, its neighbor points in dense region are also added in the cluster. This process continues until the density-connected cluster is completely found. Then, a new unvisited point is visited and the algorithm is repeated. The algorithm can be expressed as following pseudocode(source: Wikipedia).

```
DBSCAN(D, eps, MinPts) {
   C = 0
   for each point P in dataset D {
      if P is visited
         continue next point
      mark P as visited
      NeighborPts = regionQuery(P, eps)
      if sizeof(NeighborPts) < MinPts
         mark P as NOISE
      else {
         C = next cluster
         expandCluster(P, NeighborPts, C, eps, MinPts)
      }
   }
}

expandCluster(P, NeighborPts, C, eps, MinPts) {
   add P to cluster C
   for each point P' in NeighborPts {
      if P' is not visited {
         mark P' as visited
         NeighborPts' = regionQuery(P', eps)
         if sizeof(NeighborPts') >= MinPts
            NeighborPts = NeighborPts joined with NeighborPts'
      }
      if P' is not yet member of any cluster
         add P' to cluster C
   }
}

regionQuery(P, eps)
   return all points within P's eps-neighborhood (including P)
```

2. Function

a) Point class

- region_query(self, data_set, eps)

: Find all objects within eps from self instance and return objects in array.

b) Cluster class

- point_append(self, point)

: Connect point and cluster. The point is added to self cluster and the point's cluster is set as self cluster.

- expand(self, neighbor_point, data_set, eps, minPts)

: Expand cluster from each object in neighbor_point.

c) db_scan module

- db_scan(data_set, eps, minPts)

: Proceed DBScan algorithm metioned above.

3. Execution

a) Requirement

- python 3.6
- numpy
- matplotlib

b) Command line

```
python clustering.py input1.txt 8 15 20
```

(python <execution file> <input file> <cluster_num> <eps> <minPts>)