Data Mining (CSE542)

Homework 07

ID: ___ Name: __조원석__ Date: __23.05.15__

Task-1

Given the dataset in Figure 14.5, show the dendrogram resulting from the single-link hierarchical agglomerative clustering approach using the L_1-*norm* as the distance between points

$$\|\mathbf{x}-\mathbf{y}\|_1 = \sum_{d=1}^{2} |x_{id} - y_{id}|$$

K=4

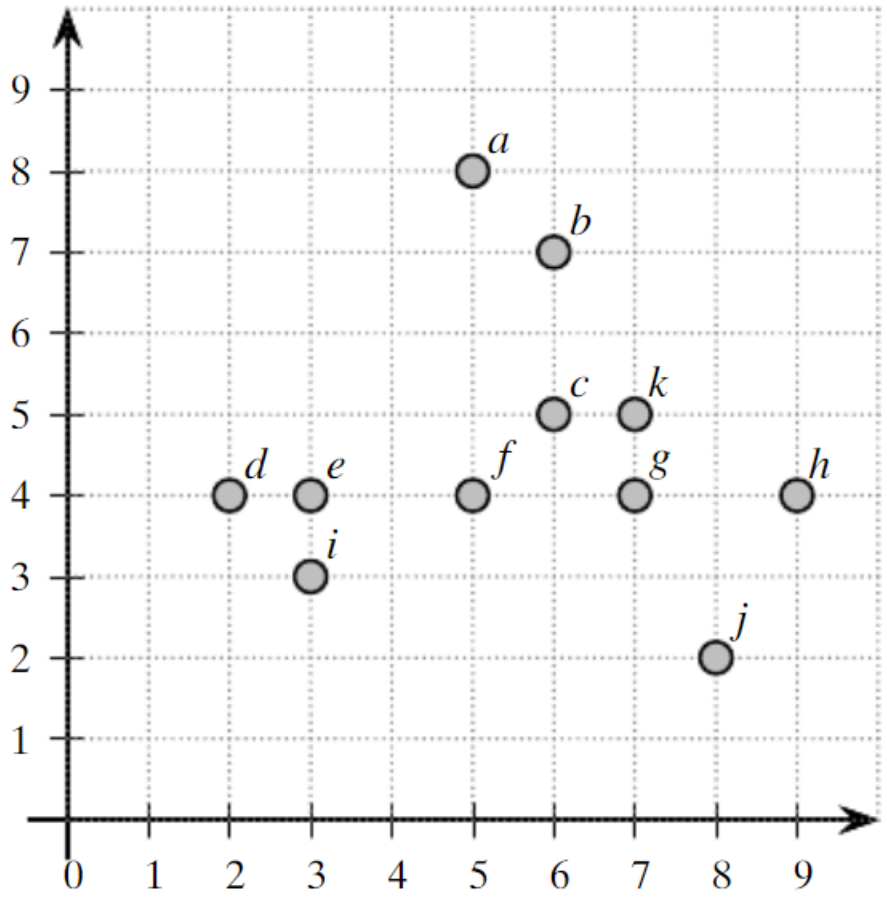


Figure 14.5. Dataset

a(5,8), b(6,7), c(6,5), d(2,4), e(3,4), f(5,4), g(7,4), h(9,4), i(3,3), j(8,2), k(7,5)

	a	b	C	D	E	F	g	h	i	j	k
a	0	2	4	7	6	4	6	8	7	9	5
b	2	0	2	7	6	4	4	6	7	7	3
c	4	2	0	5	4	2	2	4	5	5	1
d	7	7	5	0	1	3	5	7	2	8	6
e	6	6	4	1	0	2	4	6	1	7	5
f	5	4	2	3	2	0	2	4	3	5	3
g	6	4	2	5	4	2	0	2	5	3	1
h	8	6	4	7	6	4	2	0	7	3	3
i	7	7	5	2	1	3	5	7	0	6	6
j	9	7	5	8	7	5	3	3	6	0	4
k	5	3	1	6	5	3	1	3	6	4	0

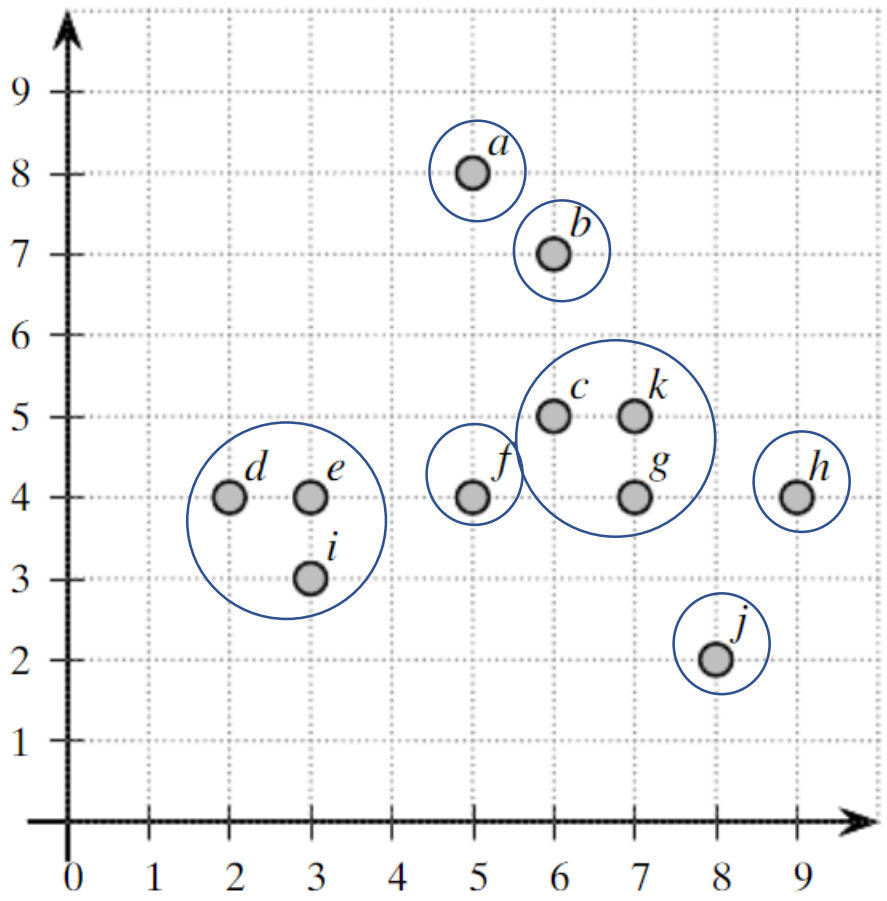
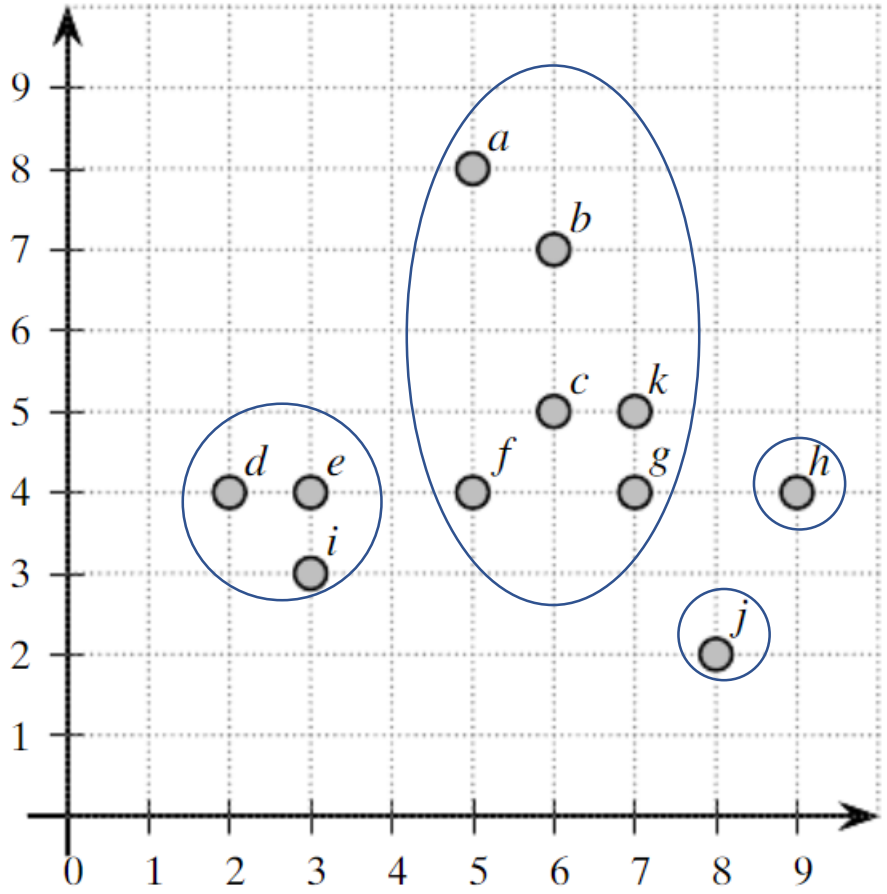


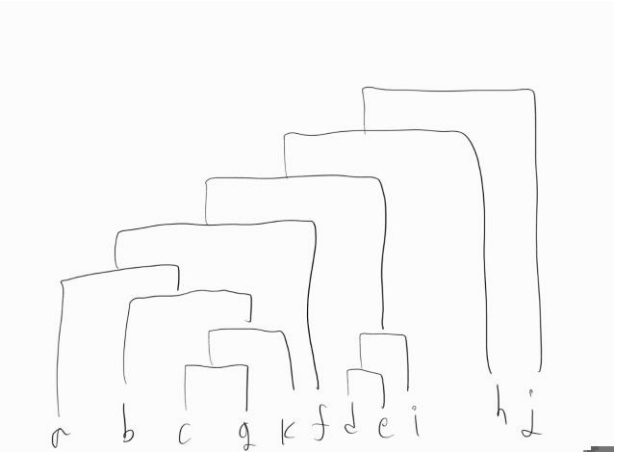
Figure 14.5. Dataset

	a	b	c,g,k	d,e,i	f	h	j
a	0	2	4	6	4	8	9
b	2	0	2	6	4	6	7
c,g,k	4	2	0	2	4	2	2
d,e,i	6	6	2	0	2	6	6
f	4	4	4	2	0	4	5
h	8	6	2	6	4	0	3
j	9	7	2	6	5	3	0



	a,b,c,f,g,k	d,e,i	h	j
a,b,c,f,g,k	0	2	2	3
d,e,i	2	0	6	6
h	2	6	0	3
j	3	6	3	0

The dendrogram is that>



Task-2

Consider Figure 15.12 and answer the following questions, assuming that we use the Euclidean distance between points, and that $\epsilon = 2$ and $minpts = 3$

(a) List all the core points.

(b) Is a directly density reachable from d?

(c) Is o density reachable from i? Show the intermediate points on the chain or the point where the chain breaks.

(d) Is density reachable a symmetric relationship, that is, if x is density reachable from y, does it imply that y is density reachable from x? Why or why not?

(e) Is l density connected to x? Show the intermediate points that make them density connected or violate the property, respectively.

(f) Is density connected a symmetric relationship?

(g) Show the density-based clusters and the noise points.

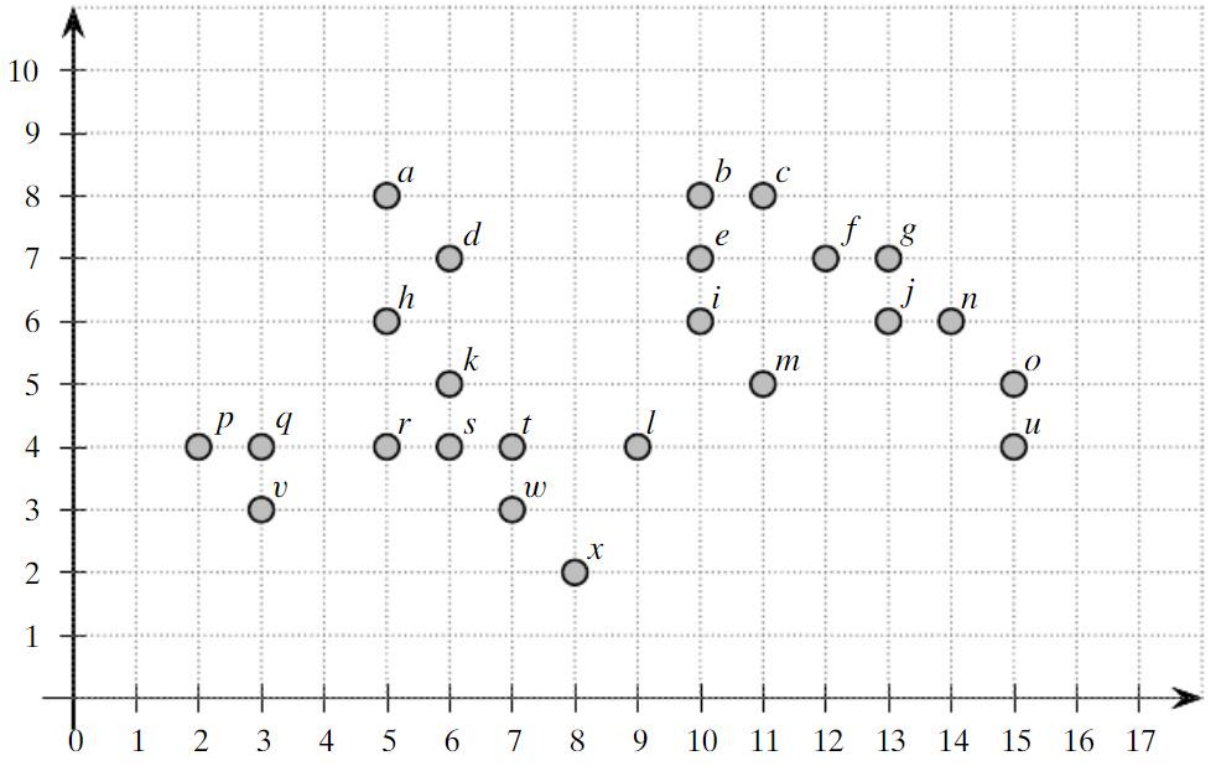


Figure 15.12. Dataset

(a) List all the points
: a, b, c, d, e, f, g, h, i, j, k, n, o, p, q, r, s, t, v, w

(b) Is 'a' directly density reachable from 'd'?
: Yes => d is a core object and a belongs to N_2(d)

(c) Is 'o' density reachable from 'i'? Show the intermediate points on the chain or the point where the chain breaks.
: Yes => { i, e, b, c, f, g, j, n, o } or { i, e, f, j, n, o }

(d) Is density reachable a symmetric relationship, that is, if 'x' is density reachable from y, does it imply that 'y' is density reachable from 'x'? Why or why not?
: Density reachable is not a symmetric relationship, since a non-coreobject may be reachable from a core object, but the reverse is not necessarilytrue. For example uis density reachable from nbut nis not density reachablefrom u.

(e) Is 'l' density connected to 'x'? Show the intermediate points that make them density connected or violate the property, respectively.
: Yes => Yes, for example, via t, since l is density-reachable from t and x is also density-reachable from t.

(f) Is density connected a symmetric relationship?
: Yes => by definition. In other words for any two points, there exists acore point that reaches both of them.

(g) Show the density-based clusters and the noise points.
: C_1 => { a, d, h, k, p, q, r, s, t, l, v, w, x}
C_2 => { b, c, e, f, g, l, j, n, m, o, u }