Introduction to Data Mining

Homework 5: Clustering

2012-11598

민두기

1.

1

4

9

16

25

36

49

64

81

1, 4

9

16

25

36

49

64

81

1, 4, 9

16

25

36

49

64

81

1, 4, 9

16, 25

36

49

64

81

1, 4, 9

16, 25

36, 49

64

81

1, 4, 9, 16, 25

36, 49

64

81

1, 4, 9, 16, 25

36, 49

64, 81

Cluster1: 1, 4, 9, 16, 25

Cluster2: 36, 49

Cluster3: 64, 81

2-(1).

(2, 2),

(3, 4),

(5, 2),

(4, 8),

(4, 10),

(6, 8),

(7, 10),

(10, 5),

(9, 3),

(11, 4),

(12, 3),

(12, 6),

(2, 2),

(3, 4),

(5, 2),

(4, 8),

(4, 10),

(6, 8),

(7, 10),

(10, 5), (11, 4),

(9, 3),

(12, 3),

(12, 6),

(2, 2),

(3, 4),

(5, 2),

(4, 8),

(4, 10),

(6, 8),

(7, 10),

(10, 5), (11, 4), (12, 3),

(9, 3),

(12, 6),

(2, 2),

(3, 4),

(5, 2),

(4, 8), (4, 10),

(6, 8),

(7, 10),

(10, 5), (11, 4), (12, 3),

(9, 3),

(12, 6),

(2, 2),

(3, 4),

(5, 2),

(4, 8), (4, 10), (6, 8),

(7, 10),

(10, 5), (11, 4), (12, 3),

(9, 3),

(12, 6),

(2, 2), (3, 4),

(5, 2),

(4, 8), (4, 10), (6, 8),

(7, 10),

(10, 5), (11, 4), (12, 3),

(9, 3),

(12, 6),

(2, 2), (3, 4),

(5, 2),

(4, 8), (4, 10), (6, 8), (7, 10),

(10, 5), (11, 4), (12, 3),

(9, 3),

(12, 6),

(2, 2), (3, 4),

(5, 2),

(4, 8), (4, 10), (6, 8), (7, 10),

(10, 5), (11, 4), (12, 3), (9, 3),

(12, 6),

(2, 2), (3, 4),

(5, 2),

(4, 8), (4, 10), (6, 8), (7, 10),

(10, 5), (11, 4), (12, 3), (9, 3), (12, 6),

(2, 2), (3, 4), (5, 2),

(4, 8), (4, 10), (6, 8), (7, 10),

(10, 5), (11, 4), (12, 3), (9, 3), (12, 6),

Cluster1: (2, 2), (3, 4), (5, 2)

Cluster2: (4, 8), (4, 10), (6, 8), (7, 10)

Cluster3: (10, 5), (11, 4), (12, 3), (9, 3), (12, 6)

2-(2).

(2, 2),

(3, 4),

(5, 2),

(4, 8),

(4, 10),

(6, 8),

(7, 10),

(10, 5),

(9, 3),

(11, 4),

(12, 3),

(12, 6),

(2, 2),

(3, 4),

(5, 2),

(4, 8),

(4, 10),

(6, 8),

(7, 10),

(10, 5), (11, 4),

(9, 3),

(12, 3),

(12, 6),

(2, 2),

(3, 4),

(5, 2),

(4, 8), (4, 10),

(6, 8),

(7, 10),

(10, 5), (11, 4),

(9, 3),

(12, 3),

(12, 6),

(2, 2),

(3, 4),

(5, 2),

(4, 8), (4, 10),

(6, 8),

(7, 10),

(10, 5), (11, 4), (12, 3),

(9, 3),

(12, 6),

(2, 2), (3, 4),

(5, 2),

(4, 8), (4, 10),

(6, 8),

(7, 10),

(10, 5), (11, 4), (12, 3),

(9, 3),

(12, 6),

(2, 2), (3, 4),

(5, 2),

(4, 8), (4, 10),

(6, 8), (7, 10),

(10, 5), (11, 4), (12, 3),

(9, 3),

(12, 6),

(2, 2), (3, 4),

(5, 2),

(4, 8), (4, 10),

(6, 8), (7, 10),

(10, 5), (11, 4), (12, 3), (9, 3),

(12, 6),

(2, 2), (3, 4),

(5, 2),

(4, 8), (4, 10), (6, 8), (7, 10),

(10, 5), (11, 4), (12, 3), (9, 3),

(12, 6),

(2, 2), (3, 4), (5, 2),

(4, 8), (4, 10), (6, 8), (7, 10),

(10, 5), (11, 4), (12, 3), (9, 3),

(12, 6),

(2, 2), (3, 4), (5, 2),

(4, 8), (4, 10), (6, 8), (7, 10),

(10, 5), (11, 4), (12, 3), (9, 3), (12, 6),

Cluster1: (2, 2), (3, 4), (5, 2)

Cluster2: (4, 8), (4, 10), (6, 8), (7, 10)

Cluster3: (10, 5), (11, 4), (12, 3), (9, 3), (12, 6)

3-(1).

Cluster1:

N: 3

SUM: (10, 8)

SUMSQ: (38, 24)

Cluster2:

N: 4

SUM: (21, 36)

SUMSQ: (117, 328)

Cluster3:

N: 5

SUM: (54, 21)

SUMSQ: (590, 95)

3-(2).

Cluster1:

Variance: (1.556, 0.889)

Standard deviation: (1.247, 0.943)

Cluster2:

Variance: (1.688, 1)

Standard deviation: (1.299, 1)

Cluster3:

Variance: (1.36, 1.36)

Standard deviation: (1.166, 1.166)

4.

Mahalanobis distance: 1.375

5.

If centroid, circle’s representative point and ring’s representative point which is on the inner ring’s inner circle are on the line , merge condition is 0.8(i – c) <= d.

More generally, define angle alpha and beta. Alpha is the smallest angle

(0 <= alpha <= 180) between centroid to circle representative point line and centroid to ring’s inner circle representative point line. Beta is the smallest angle (0<= beta <= 180) between centroid to circle representative point line and centroid to ring’s outer circle representative point line.

Merge condition is

(0.8i)^2 + (0.8c)^2 – 2\*(0.8i)(0.8c)\*cos(alpha) <= d^2

or (0.8o)^2 + (0.8c)^2 – 2\*(0.8o)(0.8c)\*cos(beta) <= d^2