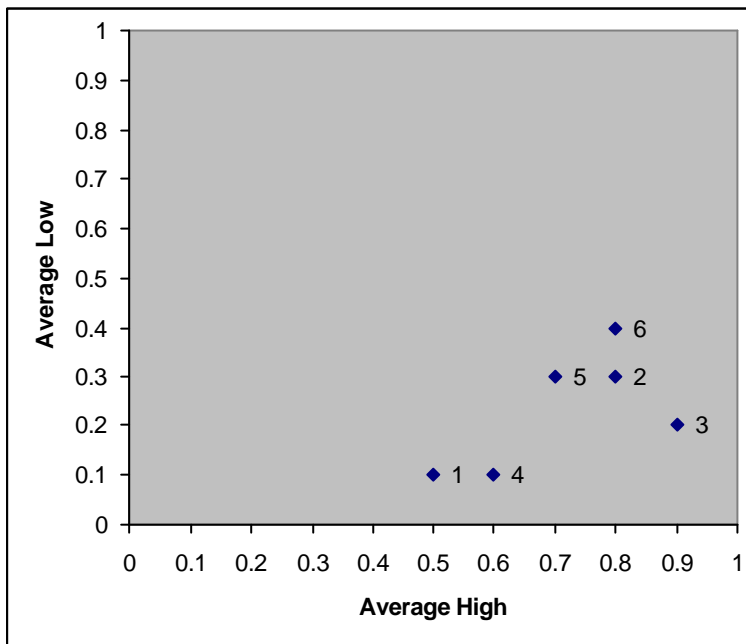# Supplementary Notes #6

COMP 578 Data Mining and Data Warehousing

MScECT, Semester 1, 03-04

Solutions to exercises in Supplementary Notes #3

1) Using Euclidean distance as the distance metric, the following dissimilarity matrix can be generated

Stock No.	Average High	Average Low
1	0.5	0.1
2	0.8	0.3
3	0.9	0.2
4	0.6	0.1
5	0.7	0.3
6	0.8	0.4

	1	2	3	4	5	6
1	0					
2	0.36	0				
3	0.41	0.14	0			
4	0.1	0.28	0.32	0		
5	0.28	0.1	0.22	0.22	0	
6	0.42	0.1	0.22	0.36	0.14	0

	14	2	3	5	6
14	0				
2	0.28	0			
3	0.32	0.14	0		
5	0.22	0.1	0.22	0	
6	0.36	0.1	0.22	0.14	0

	14	25	3	6
14	0			
25	0.22	0		
3	0.32	0.14	0	
6	0.36	0.1	0.22	0

	14	256
14	0	
2356	0.22	0

—

	14	256	3
14	0		
256	0.22	0	
3	0.32	0.14	0

2)

Let the 2 random initial cluster C1 and C2 be sample 1 and sample 2

C1 = {1}, C2 = {2}; Center of C1 = {0.5, 0.1}, Center of C2 = {0.8, 0.3}

Iteration 1:

- Sample 3
  D(3, Center of C1) = 0.41;   D(3, Center of C2) = 0.14
  Sample 3 is assigned to C2

- Sample 4
  D(4, Center of C1) = 0.1;   D(4, Center of C2) = 0.28
  Sample 4 is assigned to C1

- Sample 5
  D(5, Center of C1) = 0.28;   D(5, Center of C2) = 0.1
  Sample 5 is assigned to C2

- Sample 6
  D(6, Center of C1) = 0.42;   D(6, Center of C2) = 0.1
  Sample 6 is assigned to C2

C1 = {1,4}, C2 = {2,3,5,6}

New centers for C1 = (0.55, 0.1) for C2=(0.8, 0.3)

Iteration 2:

- Sample 1
  D(1, Center of C1) = 0.05;   D(1, Center of C2) = 0.36
  Sample 1 is assigned to C1

- Sample 2

D(2, Center of C1) = 0.32;   D(3, Center of C2) = 0

Sample 2 is assigned to C2

- Sample 3

  D(3, Center of C1) = 0.36;   D(3, Center of C2) = 0.14

  Sample 3 is assigned to C2

- Sample 4

  D(4, Center of C1) = 0.05;   D(4, Center of C2) = 0.28

  Sample 4 is assigned to C1

- Sample 5

  D(5, Center of C1) = 0.25;   D(5, Center of C2) = 0.1

  Sample 5 is assigned to C2

- Sample 6

  D(6, Center of C1) = 0.39;   D(6, Center of C2) = 0.1

  Sample 6 is assigned to C2

The centers do not have any changes, the algorithm stops.

3)

From left to right, label them as 1 – 9

Each face has 10 features.

- Addition of 1:

  Assigns 1 to cluster 1 (C1)

- Addition of 2:

  $S(C1, 2) = S(1, 2) = -2$

  Assigns 2 to cluster 2 (C2)

- Addition of 3:

  $S(C1, 3) = S(1, 3) = 4$

  $S(C2, 3) = S(2, 3) = 2$

  Assigns 3 to cluster 1 (C1)

- Addition of 4:

  $S(C1, 4) = S(1, 4) + S(3, 4) = 2 + -2 = 0$

  $S(C2, 4) = S(2, 4) = -4$

  Assigns 4 to cluster 2 (C3)

- Addition of 5:

  $S(C1, 5) = S(1, 5) + S(3, 5) = 2 + 2 = 4$

  $S(C2, 5) = S(2, 5) = 2$

  $S(C3, 5) = S(4, 5) = -2$

  Assigns 5 to cluster 2 (C1)

- Addition of 6:

  $S(C1, 6) = S(1, 6) + S(3, 6) + S(5, 6) = -2 + 2 + 2 = 2$

  $S(C2, 6) = S(2, 6) = 6$

  $S(C3, 6) = S(4, 6) = 0$

  Assigns 6 to cluster 2 (C2)

- Addition of 7:

  $S(C1, 7) = S(1, 7) + S(3, 7) + S(5, 7) = -6 + -4 + 0 = -10$

  $S(C2, 7) = S(2, 7) + S(6, 7) = -2 + -4 = -6$

  $S(C3, 7) = S(4, 7) = -4$

  Assigns 7 to cluster 4 (C4)

- Addition of 8:

  $S(C1, 8) = S(1, 8) + S(3, 8) + S(5, 8) = 2 + 2 + 4 = 8$

  $S(C2, 8) = S(2, 8) + S(6, 8) = -2 + -2 = -4$

  $S(C3, 8) = S(4, 8) = 0$

  $S(C4, 8) = S(7, 8) = 0$

  Assigns 8 to cluster 1 (C1)

- Addition of 9:

  $S(C1, 9) = S(1, 9) + S(3, 9) + S(5, 9) + S(8, 9) = 0 + -2 + -6 + -4 = -12$

  $S(C2, 9) = S(2, 9) + S(6, 9) = -8 + -6 = -14$

  $S(C3, 9) = S(4, 9) = -2$

  $S(C4, 9) = S(7, 9) = -4$

  Assigns 9 to cluster 5 (C5)

Cluster 1:	Cluster 4:
Cluster 2:	Cluster 5:
Cluster 3:	