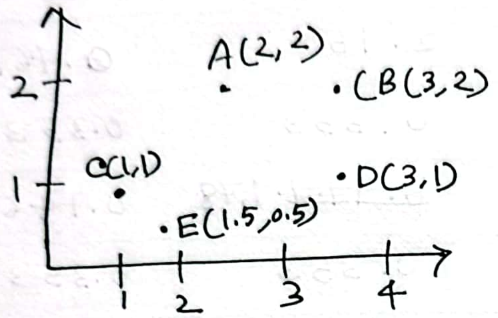
K-means clustering

Use k-Means Algorithm to create two cluster



Assume A(2,2) and C(1,1) are centers of two clusters.

Iteration 1

Given Points	c1(2,2)	c2(1,1)	minDis	Select
A(2,2)	0	1.41	0	c_1
B(3,2)	1	2.24	1	c_1
C(1,1)	1.41	0	0	c_2
D(3,1)	1.41	2	1.41	c_1
E(1.5,0.5)	1.58	0.71	0.7	c_2

New cluster center,

$$c1 = \left(\frac{2+3+3}{3}, \frac{2+2+1}{3}\right) = (2.67, 1.67)$$

$$c_2 = \left(\frac{1+1.5}{2}, \frac{1+0.5}{2}\right) = (1.25, 0.75)$$

Iteration 2

Given Points	C1 (2.67, 1.67)	C2 (1.25, 0.75)	min Dis	Select
A (2,2)	0.746	1.457	0.746	c_1
B (3,2)	0.467	2.15	0.467	C1
C (1,1)	1.799	0.353	0.353	C_2
D (3,1)	0.746	1.78	0.746	C1
E (1.5, 0.5)	1.654	0.353	0.353	C2

New center,

$$C1 = \left(\frac{2+3+3}{3}, \frac{2+2+1}{3} \right) = (2.67, 1.67)$$

$$C2 = \left(\frac{1+1.5}{2}, \frac{1+0.5}{2} \right) = (1.25, 0.75)$$

This is completion of iteration 02.

The centers do not change anymore.

A (2,2), B (3,2), D (3,1) Points belong to C1

C (1,1), E (1.5, 0.5) Points belong to C2.

Practice Problems

Initial clusters are: A1(2,10), A4(5,8), A7(1,2)

I1

Given	C1(2,10)	C_2(5,8)	C_3(1,2)	min Dis	Select
A1(2,10)	0	3.61	8.06	0	c_1
A2(2,5)	5	4.24	3.16	3.16	c3
A3(8,4)	8.48	5	7.28	5	c2
A4(5,8)	3.61	0	7.2	0	c2
A5(7,5)	7.07	3.61	6.71	3.61	c2
A6(6,4)	7.21	4.12	5.39	4.12	c2
A7(1,2)	8.06	7.2	0	0	c3
A8(4,9)	2.24	1.42	7.62	1.42	c2

New center, $c_1 = (2, 10)$
$c_2 = (6, 6)$
$c_3 = (1.5, 3.5)$

I2

Given	c_1(2,10)	c_2(6,6)	c_3(1.5,3.5)	min Dis	Select
A1(2,10)	0	5.65	6.52	0	c_1
A2(2,5)	5	4.13	1.58	1.58	c3
A3(8,4)	8.48	2.83	6.52	2.83	c2
A4(5,8)	3.61	2.24	5.7	2.24	c2
A5(7,5)	7.07	1.41	5.7	1.41	c2
A6(6,4)	7.21	2	4.53	2	c2
A7(1,2)	8.06	6.4	1.58	1.58	c3
A8(4,9)	2.24	3.61	6.04	2.24	c1

$$c_1 = \left(\frac{2+4}{2}, \frac{10+9}{2}\right) = (3, 9.5)$$

$$c_2 = \left(\frac{8+5+7+6}{4}, \frac{4+8+5+4}{4}\right) = (6.5, 5.25)$$

$$c_3 = \left(\frac{2+1}{2}, \frac{5+2}{2}\right) = (1.5, 3.5)$$

I3

Given	c1 (3,9.5)	c2 (6.5,5.25)	c3 (1.5,3.5)	minDis	Select
A1(2,10)	1.11	6.54	6.52	1.11	c1
A2(2,5)	4.61	4.51	1.58	1.58	c3
A3(8,4)	7.43	1.95	6.52	1.95	c2
A4(5,8)	2.5	3.13	5.7	2.5	c1
A5(7,5)	6.02	0.56	5.7	0.56	c2
A6(6,4)	6.26	1.35	4.53	1.35	c2
A7(1,2)	7.76	6.39	1.58	1.58	c3
A8(4,9)	1.12	4.51	6.04	1.12	c1

1. Agglomerative Algorithm : Single Link

Sample No	X	Y
P1	0.4	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.23	0				
P_3	0.22	0.14	0			
P_4	0.37	0.19	0.16	0		
P_5	0.34	0.14	0.28	0.28	0	
P_6	0.24	0.24	0.10	0.22	0.39	0

Here minimum value is 0.1. Update Distance Matrix

	P_1	P_2	P_3, P_6	P_4	P_5
P_1	0				
P_2	0.23	0			
P_3, P_6	0.22	0.14	0		
P_4	0.37	0.19	0.16	0	
P_5	0.34	0.14	0.28	0.28	0

Here minimum value 0.14. Hence we combine P_2, P_5

	P_1	P_2, P_5	P_3, P_6	P_4
P_1	0			
P_2, P_5	0.23	0		
P_3, P_6	0.22	0.14	0	
P_4	0.37	0.19	0.16	0

Here minimum value 0.14. Hence we combine

(P_2, P_5), (P_3, P_6)

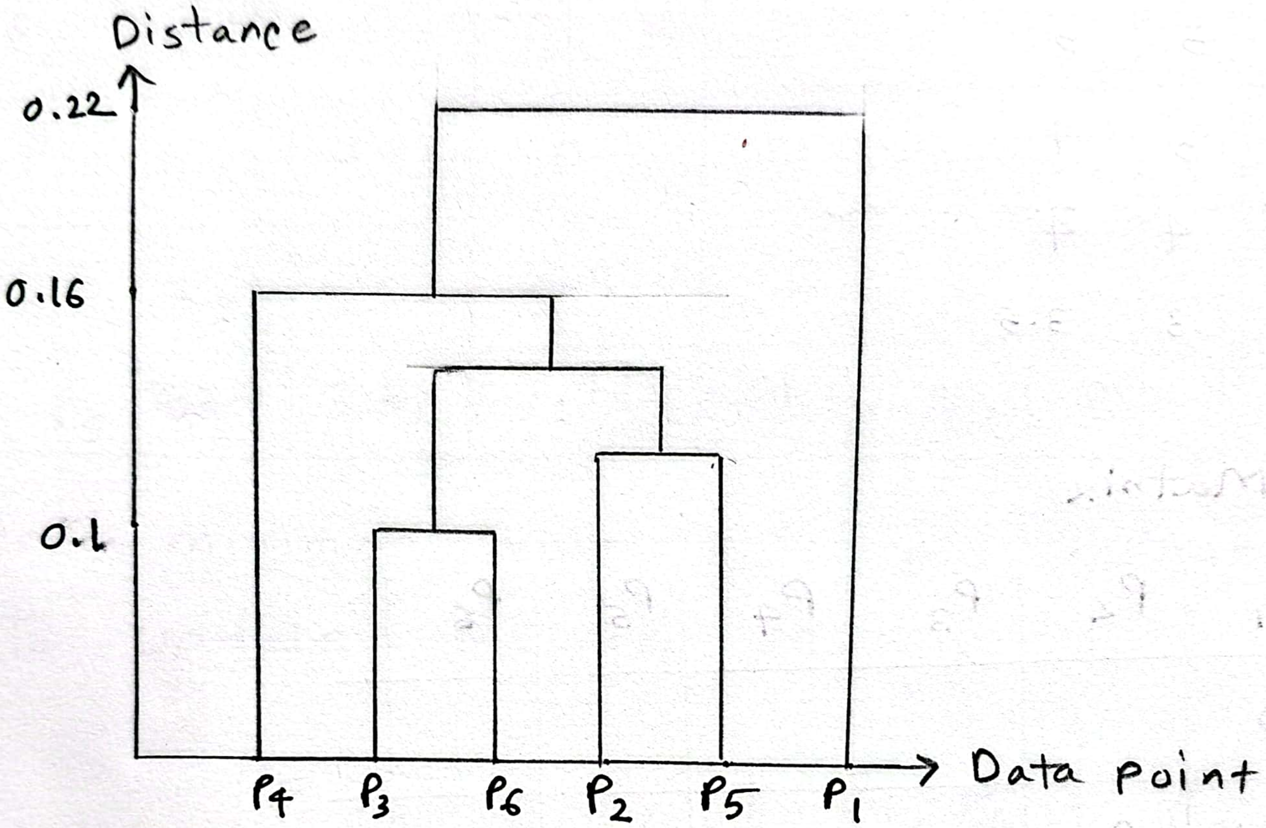
	P_1	$(P_2, P_5), (P_3, P_6)$	P_4
P_1	0		
$(P_2, P_5), (P_3, P_6)$	0.22	0	
P_4	0.37	0.16	0

Here minimum value 0.16. Hence we combine

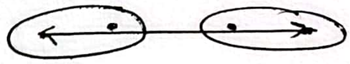
$[(P_2, P_5), (P_3, P_6)], P_4$

	P_1	$((P_2, P_5), (P_3, P_6)), P_4$
P_1	0	
$((P_2, P_5), (P_3, P_6)), P_4$	0.22	0

Dendrogram



2. Agglomerative Algorithm: Complete Link



Sample	X	Y
P_1	1	1
P_2	1.5	1.5
P_3	5	5
P_4	3	4
P_5	4	4
P_6	3	3.5

Distance Matrix

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.71	0				
P_3	5.66	4.95	0			
P_4	3.61	2.92	2.24	0		
P_5	4.24	3.54	1.41	1.0	0	
P_6	3.20	2.5	2.5	0.5	1.12	0

Here minimum value is 0.5. Merge (P_4, P_6). To update distance matrix

	P_1	P_2	P_3	P_4, P_6	P_5
P_1	0				
P_2	0.71	0			
P_3	5.66	4.95	0		
P_4, P_6	3.61	2.92	2.5	0	
P_5	4.24	3.54	1.41	1.12	0

Here minimum value 0.71 and merge (P_1, P_2).

To update distance Matrix

	P_1, P_2	P_3	P_4, P_6	P_5
P_1, P_2	0			
P_3	5.66	0		
P_4, P_6	3.61	2.5	0	
P_5	4.24	1.41	1.12	0

Here minimum value 1.12 and merge P_4, P_6 and P_5.

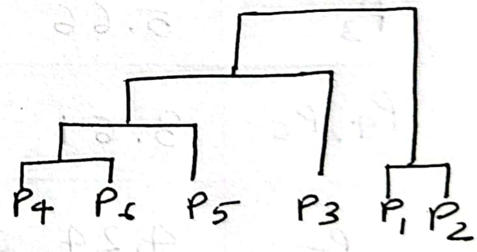
To update distance Matrix

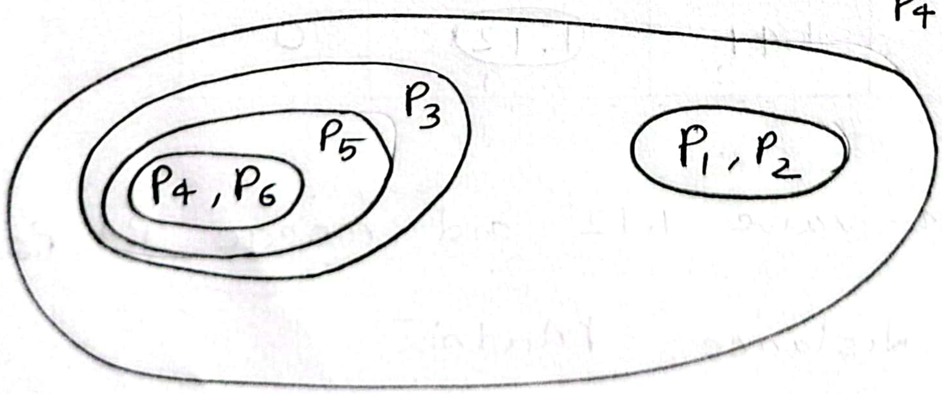
	$P_1 P_2$	P_3	$(P_4, P_6) P_5$
$P_1 P_2$	0		
P_3	5.66	0	
$(P_4, P_6) P_5$	4.24	2.5	0

We get the minimum value as 2.5 and hence combine $(P_4, P_6) P_5$ and P_3.

	P_1, P_2	P_4, P_6, P_5, P_3
P_1, P_2	0	
P_4, P_6, P_5, P_3	5.66	0

Dendrogram





3. Agglomerative : Average Link 

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0					
P_2	0.71	0				
P_3	5.66	4.95	0			
P_4	3.61	2.92	2.24	0		
P_5	4.24	3.54	1.41	1.0	0	
P_6	3.20	2.5	2.5	0.5	1.12	0

↓

	P_1	P_2	P_3	P_4, P_6	P_5
P_1	0				
P_2	0.71	0			
P_3	5.66	4.95	0		
P_4, P_6	3.41	2.71	2.37	0	
P_5	4.24	3.54	1.41	1.06	0

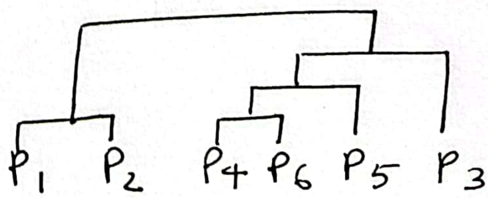
	P_1, P_2	P_3	P_4, P_6	P_5
P_1, P_2	0			
P_3	5.31	0		
P_4, P_6	3.06	2.37	0	
P_5	3.89	1.41	1.06	0

↓

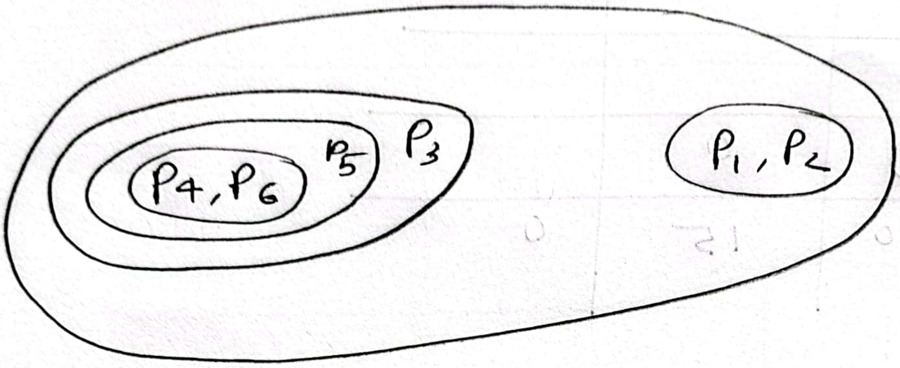
	P_1, P_2	P_3	$(P_4, P_6)\ P_5$
P_1, P_2	0		
P_3	5.31	0	
$(P_4, P_6)\ P_5$	3.48	1.89	0

↓

	P_1, P_2	$((P_4, P_6)\ P_5)\ P_3$
$P_1 P_2$	0	
$((P_4, P_6) P_5) P_3$	4.395	0



Dendrogram



\# Agglomerative Hierarchical clustering 6 one dimensional

18, 22, 25, 42, 27, 43

	18	22	25	27	42	43
18	0					
22	4	0				
25	7	3	0			
27	9	5	2	0		
42	24	20	17	15	0	
43	25	21	18	16	1	0

↓

	18	22	25	27	42, 43
18	0				
22	4	0			
25	7	3	0		
27	9	5	2	0	
42, 43	24	20	17	15	0

	18	22	25,27	42,43
18	0			
22	4	0		
25,27	7	3	0	
42,43	24	20	15	0

↓

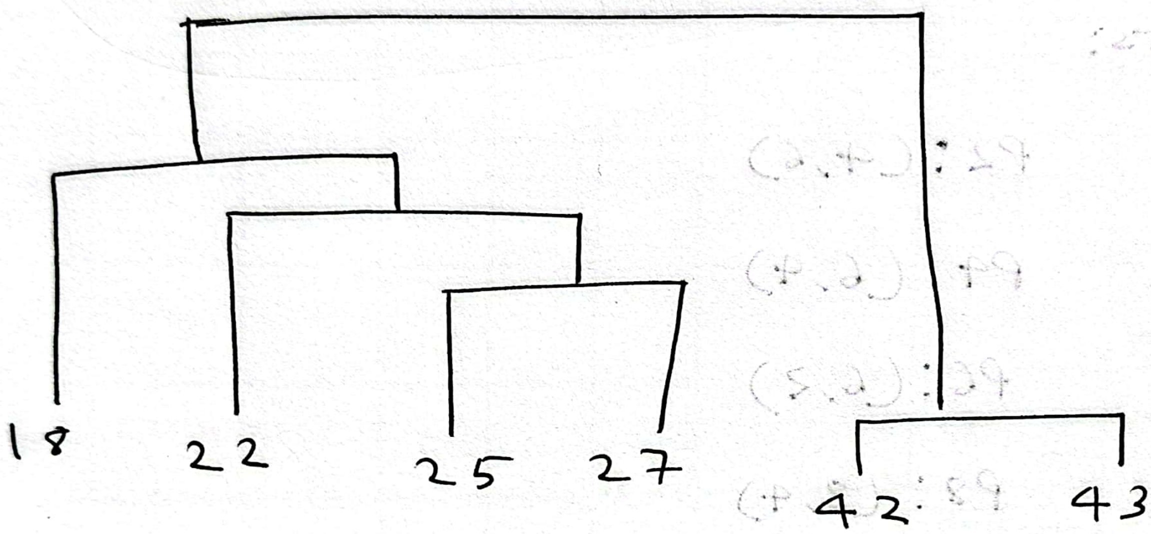
	18	22,(25,27)	42,43
18	0		
22,(25,27)	4	0	
42,43	24	15	0

↓

	18, 22, 25, 27	42, 43
18,22, 25, 27	0	15
42, 43	15	0

↓

	18,22,25,27, 42, 43
18,22,25,27,42,43	0


18
22
25
27
42
43

DBSCAN

Solved Example – 1

minPts = 4 and epsilon (ε) = 1.9

$$Distance(A(x_1, y_1), B(x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Data Points:

P1 : (3, 7)
P2 : (4, 6)

P3 : (5, 5)
P4 : (6, 4)

P5 : (7, 3)
P6 : (6, 2)

P7 : (7, 2)
P8 : (8, 4)

P9 : (3, 3)
P10 : (2, 6)

P11 : (3, 5)
P12 : (2, 4)

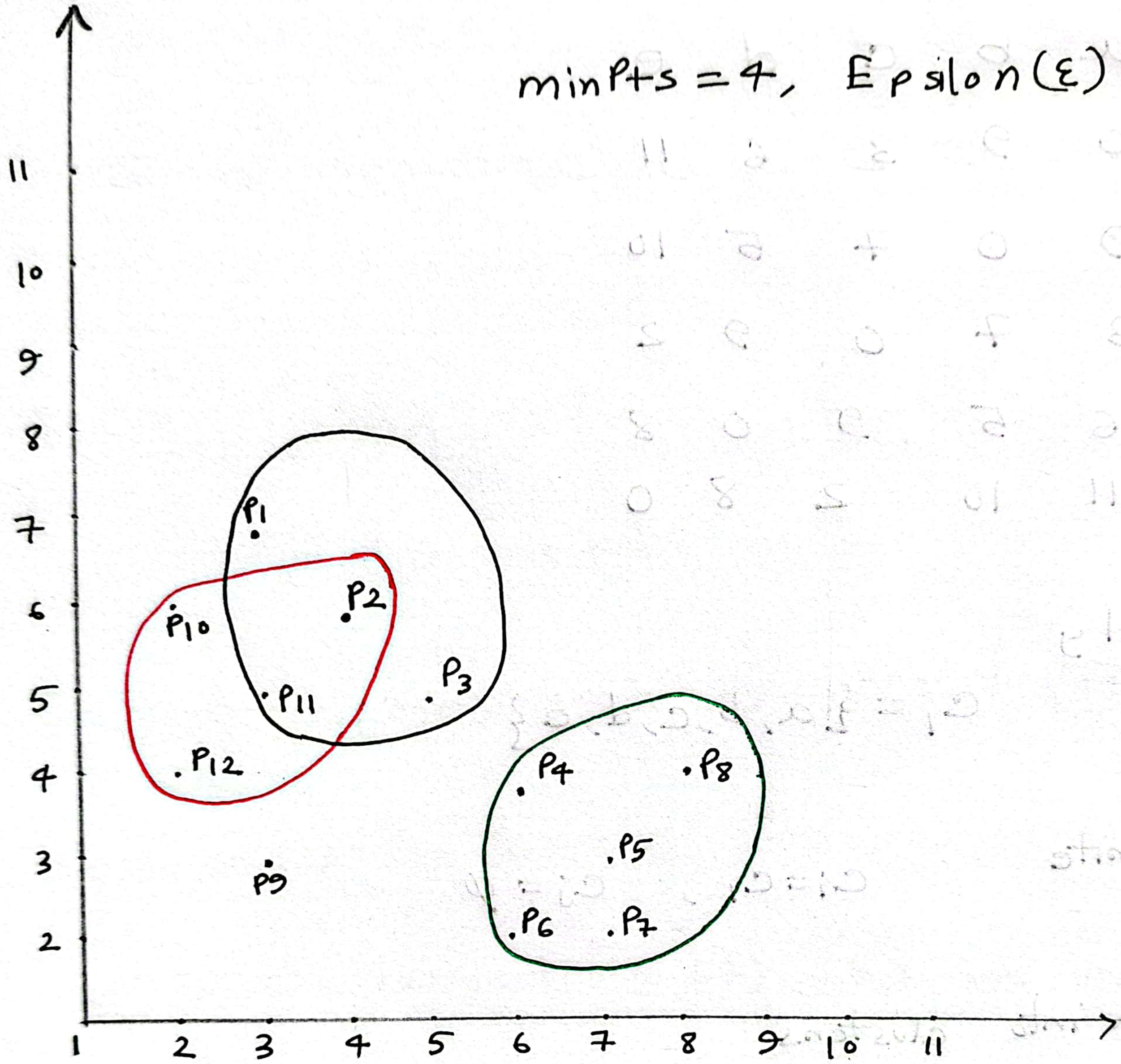
minPts = 4 and epsilon(ε) = 1.5

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	
P_1	0												P_1 : P_2, P_{10}
P_2	1.41	0											P_2 : P_1, P_3, P_{11}
P_3	2.83	1.41	0										P_3 : P_2, P_4
P_4	4.24	2.83	1.41	0									P_4 : P_3, P_5,
P_5	5.66	4.24	2.83	1.41	0								P_5 : P_4, P_6, P_7, P_8
P_6	5.83	4.47	3.16	2.00	1.41	0							P_6 : P_5, P_7,
P_7	6.40	5.00	3.61	2.24	1.00	1.00	0						P_7 : P_5, P_6,
P_8	5.83	4.47	3.16	2.00	1.41	2.83	2.24	0					P_8 : P_5
P_9	4.00	3.16	2.83	3.16	4.00	3.16	4.12	5.10	0				P_9 : P_{12}
P_{10}	1.41	2.00	3.16	4.47	5.83	5.66	6.40	6.32	3.16	0			P_{10} : P_1, P_{11}
P_{11}	2.00	1.41	2.00	3.16	4.47	4.24	5.00	5.10	2.00	1.41	0		P_{11} : P_2, P_{10}, P_{12}
P_{12}	3.16	2.83	3.16	4.00	5.10	4.47	5.39	6.00	1.41	2.00	1.41	0	P_{12} : P_9, P_{11}

minPts = 4, Epsilon (ε) = 1.9

	Point	Status	
P_1 : P_2, P_{10}	P_1	Noise	Border
P_2 : P_1, P_3, P_{11}	P_2	Core	
P_3 : P_2, P_4	P_3	Noise	Border
P_4 : P_3, P_5	P_4	Noise	Border
P_5 : P_4, P_6, P_7, P_8	P_5	Core	
P_6 : P_5, P_7	P_6	Noise	Border
P_7 : P_5, P_6	P_7	Noise	Border
P_8 : P_5	P_8	Noise	Border
P_9 : P_{12}	P_9	Noise	
P_{10} : P_1, P_{11}	P_{10}	Noise	Border
P_{11} : P_2, P_{10}, P_{12}	P_{11}	Core	
P_{12} : P_9, P_{11}	P_{12}	Noise	Border

minPts = 4, Epsilon(ε) = 1.9



DIANA (DIvisive ANAlysis)

	a	b	c	d	e
a	0	9	3	6	11
b	9	0	7	5	10
c	3	7	0	9	2
d	6	5	9	0	8
e	11	10	2	8	0

① Initially

$c_1 = \{a, b, c, d, e\}$

② We write $c_i = c_1$, $c_j = \phi$

③ Division into clusters

ⓐ initial iteration

Average dissimilarity of $a = \frac{1}{4}(d(a,b) + d(a,c) + d(a,d) + d(a,e))$

$= \frac{1}{4}[9 + 3 + 6 + 11) = 7.25$

$b = 7.75$

$c = 5.25$

$d = 7$

$e = 7.75$

7.75 Highest.

now, we ~~we~~ move b to c_j.

$c_i = \{a, c, d, e\}$, $c_j = \varphi \cup \{b\} = \{b\}$

ⓑ Remaining iterations

(i) 2nd iteration

$$D_a = \frac{1}{3}[d(a,c) + d(a,d) + d(a,e)] - \frac{1}{1}[d(a,b)]$$

$$= \frac{20}{3} - 9 = -2.33$$

$$D_c = \frac{1}{3}[d(c,a) + d(c,d) + d(c,e)] - \frac{1}{1}[d(c,b)] = -2.33$$

$D_d = 0.67$ $D_e = 0$

D_d is the largest (0.67). So we move, d to c_j.

we no have,

$c_i = \{a, c, e\}$, $c_j = \{b\} \cup \{d\} = \{b, d\}$

(ii) 3rd iteration

$$D_a = \frac{1}{2}[d(a,c) + d(a,e)] - \frac{1}{2}[d(a,b) + d(a,d)]$$

$$= \frac{1}{2}[3 + 11] - \frac{1}{2}[6 + 9] = -0.5$$

$D_c = -13.5$ $D_e = -2.5$

All are negative. So we stop and form the clusters C_i and C_j.

④ To divide, C_i and C_j, we compute their diameters.

$$\text{diameter}(C_i) = \max\{d(a,c), d(a,e), d(c,e)\}$$
$$= \max\{3, 11, 6\} = 11$$

$$\text{diameter}(C_j) = \max\{d(b,d)\} = 5$$

The cluster with the largest diameter is C_i. So we now split C_i.

Association Rules Exercise

① The support level 33 %

confidence level 50%

Rule: $X \to Y$

$$\text{Support} = \frac{frq(X,Y)}{N}$$

$$\text{Confidence} = \frac{frq(X,Y)}{frq(X)}$$

①

1-item sets	Frequency
Milk	→ 9
Bread	→ 10
Butter	→ 10
Egg	→ 3 $\frac{3}{12} = 0.25 < 33\%$
Cookies	→ 5 $\frac{5}{12} = 0.42 > 0.33\%$
Ketchup	→ 3

Frequent 1-item sets	Frequency
Milk	9
Bread	10
Butter	10
Cookies	5

②

2-item sets	Frequency
Milk, Bread	→ 7
Milk, Butter	→ 7
Milk, cookies	→ 3
Bread, Butter	→ 9
Bread, cookies	→ 4
Butter, cookies	→ 3

Frequent 2-item sets	Frequency
Milk, Bread	7
Milk, Butter	7
Bread, Butter	9
Bread, Cookies	4

Milk, Butter, Bread, Cookies

3-item sets	Frequency
Milk, Butter, Bread →	6
Milk, Butter, Cookies →	1
Milk, Bread, Cookies →	2
Butter, Bread, Cookies →	3

Frequent 3-item sets	Frequency
Milk, Bread, Butter	6

Frequent 3-item set = 1 ⇒ { Milk, Bread, Butter}

Non-Empty subset are

{ {Milk}, {Bread}, {Butter}, {Milk, Bread}, {Milk, Butter}, {Bread, Butter} }

Min support = 30%   Min confidence = 60%

Association Rule

$S \rightarrow (I-S)$

If $\frac{\text{Support (I)}}{\text{Support (S)}} \geq$ minimum confidence

Rule 1: {Milk} → {Bread, Butter}

$\text{Support} = \frac{6}{12} = 50\%$

$\text{Confidence} = \frac{\text{Support(Milk, Bread, Butter)}}{\text{Support(Milk)}} = \frac{6/12}{9/12} = 0.6$

Valid

Rule2: {Bread} → {Milk, Butter}

Support = $\frac{6}{12}$ = 50%

Confidendence = $\frac{\text{Support(Milk, Bread, Butter)}}{\text{Support (Bread)}} = \frac{6/12}{10/12}$ = 0.6
=60%

Valid

Rule 3: {Butter} → {Milk, Bread}

Support = 50%.

Confidence = $\frac{6/12}{\text{Support(Butter)}} = \frac{6/12}{10/12}$ = 0.6 = 60%

Valid

Rule 4: {Milk, Bread} → {Butter}

Support = 50%.

Confidence = $\frac{6/12}{\text{Support(Milk, Bread)}} = \frac{6/12}{7/12}$ = 85.7%

Valid

Rule 5: {Milk, Butter} → {Bread}

Support = 50%.

Confidence = $\frac{6/12}{7/12}$ = 85.7%

Valid

Rule 6: {Bread, Butter} → {Milk}

Support = 50%

$$\text{confidence} = \frac{6/12}{9/12} = 66.67\%$$

Valid.

Solved Example 3

TID	List of Items IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

min support = 2

Confidence = 50%

1-item sets	Frequency
I1	6
I2	7
I3	6
I4	2
I5	2

Frequent 1-item sets	Frequency
I1	6
I2	7
I3	6
I4	2
I5	2

2-item sets	Frequency
I1 I2 →	4
I1 I3 →	4
I1 I4 →	1
I1 I5 →	2
I2 I3 →	4
I2 I4 →	2
I2 I5 →	2
I3 I4 →	0
I3 I5 →	1
I4 I5 →	0

Frequent 2-item sets	Frequency
I1 I2	4
I1 I3	4
I1 I5	2
I2 I3	4
I2 I4	2
I2 I5	2

3-item sets   Frequency

I1 I2 I3 → 2

I1 I2 I4 → 1

I1 I2 I5 → 2

I1 I3 I4 → 0

I1 I3 I5 → 1

I1 I4 I5 → 0

I2 I3 I4 → 0

I2 I3 I5 → 1

I2 I4 I5 → 0

I3 I4 I5 → 0

I1 I2 I3 I4 I5

Frequent 3-item sets	Frequency
I1 I2 I3	2
I1 I2 I5	2

Frequent 3-item set = 1 ⇒ {1, 2, 3} and {1, 2, 5}

Non-empty subsets are

{{1}, {2}, {3}, {1, 2}, {1, 3}, {2, 3}}

{{1}, {2}, {5}, {1, 2}, {1, 5}, {2, 5}}

Association Rule

$S \rightarrow (1 - S)$

if support(1) / support(S) ≥ min confidence

$I = \{1, 2, 3\}$

Rule 1: $\{1\} \rightarrow \{2, 3\}$

$$\text{Support} = \frac{2}{9} = 22.22\%$$

$$\text{Confidence} = \frac{2/9}{\text{Support}(1)} = \frac{2/9}{6/9} = 33.33\% < 50\%$$

Invalid

Rule 2: $\{2\} \rightarrow \{1, 3\}$

$$\text{Support} = 22.22\%$$

$$\text{Confidence} = \frac{2/9}{\text{Support}\{2\}} = \frac{2/9}{7/9} = 28.57 < 50\%$$

Invalid

Rule 3: $\{3\} \rightarrow \{1, 2\}$

$$\text{Support} = 22.22\%$$

$$\text{Confidence} = \frac{2/9}{\text{Support}\{3\}} = \frac{2/9}{6/9} = 33.33\% < 50\%$$

Invalid

Rule 4: $\{1, 2\} \rightarrow \{3\}$

$$\text{Support} = 22.22\%$$

$$\text{Confidence} = \frac{2/9}{4/9} = 50\% \geq 50\%$$

Valid

Rule 5: $\{1,3\} \rightarrow \{2\}$

Support = 22.22%

$$\text{Confidence} = \frac{2/9}{4/9} = 50\%$$

Valid

Rule 6: $\{2,3\} \rightarrow \{1\}$

Support = 22.22%

$$\text{Confidence} = \frac{2/9}{4/9} = 50\%$$

Valid

Now,

$I = \{1, 2, 5\}$

Rule 1: $\{1\} \rightarrow \{2,5\}$

$$\text{Support} = \frac{2}{9} = 22.22\%$$

$$\text{Confidence} = \frac{2/9}{6/9} = 33.33\% < 50\%$$

Invalid

Rule 2: $\{2\} \rightarrow \{1,5\}$

Support = 22.22%

$$\text{Confidence} = \frac{2/9}{7/9} = 28.57\%$$

Rule 3: $\{5\} \rightarrow \{1, 2\}$

Support = 22.22%

Confidence = $\frac{2/9}{2/9}$ = 100%

Valid

Rule 4: $\{1, 2\} \rightarrow \{5\}$

Support = 22.22%

Confidence = $\frac{2/9}{4/9}$ = 50%

Valid

Rule 5: $\{1, 5\} \rightarrow \{2\}$

Support = 22.22%

Confidence = $\frac{2/9}{2/9}$ = 100%.

Valid

Rule 6: $\{2, 5\} \rightarrow \{1\}$

Support = 22.22%

Confidence = $\frac{2/9}{2/9}$ = 100%.

Valid

Aprioni Algorithm Example 1

Transaction ID	Items
T_1	Hot Dogs, Buns, ketchup
T_2	Hot Dogs, Buns
T3	Hot Dogs, Coke, chips
T4	chips, Coke
T5	chips, ketchup
T6	Hot Dogs, Coke, chips

minimum ~~support~~ threshold (c) = 33.33 %

minimum confident threshold c = 60%

Minimum support count = $\frac{33.33}{100} \times 6 = 2$

Item Set	Sup count
Hot Dogs	4
Buns	2
ketchup	2
Coke	3
chips	4

→

Itemset	Sup Count
Hot Dogs	4
Buns	2
ketchup	2
Coke	3
chips	4

CamScanner

2 item set Sup-count

Hot Dogs, Buns → 2

Hot Dogs, ketchup → 1

Hot Dogs, coke → 2

Hot Dogs, chips → 2

Buns, ketchup → 1

Buns, coke → 0

Buns, chips → 0

ketchup, coke → 0

ketchup, chips → 1

Coke, chips → 3

→

Frequent 2 item sets	Sup Count
Hot Dogs, Buns	2
Hot Dogs, Coke	2
Hot Dogs, Chips	2
Coke, chips	3

3 item sets	Sup-count
Hot Dogs, Buns, Coke	0
Hot Dogs, Buns, chips	0
Hot Dogs, Coke, Chips	2

→

Frequent-3	Sup-count
Hot Dogs, Coke, chips	2

Frequent itemset(I) = { Hot Dogs, Coke, Chips}

Non-empty subsets are –

{ {Hot Dogs} , {Coke}, {chips}, {Hot Dogs, Coke} ,
{Hot Dogs, Chips} , {Coke, chips} }

Rule 1: {Hot Dogs} → {Coke, chips}

$$\text{Support} = \frac{2}{6} = 33.33\%$$

$$\text{Confidence} = \frac{2/6}{4/6} = 50\%$$ Rejected

Rule 2: {Coke} → {Hot Dogs, chips}

$$\text{Support} = \frac{2}{6} = 33.33\%$$

$$\text{Confidence} = \frac{2/6}{3/6} = 66.67\%$$ Selected

Rule 3: {Chips} → {Hot Dogs, Coke}

$S \rightarrow 33.33\%$

$$C \rightarrow \frac{2/6}{4/6} = 50\%$$ Rejected

Rule 4: {Hot Dogs, Coke} → {Chips}

$S \rightarrow 33.33\%$

$$C \rightarrow \frac{2/6}{2/6} = 100\%$$ Selected

Rule 5: {Hot Dogs, Chips} → {Coke}

$S \rightarrow 33.33\%$

$$C \rightarrow \frac{2/6}{2/6} = 100\%$$ Selected

Rule 6: {Coke, Chips} → {Hot Dogs}

$S \rightarrow 33.33\%$

$C \rightarrow \frac{2/6}{3/6} = 66.67\% > 60\%$ Selected

Apriori Algorithm Example-2

TID	items
T1	Hot Dogs, Buns, ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, chips
T4	Chips, Coke
T5	chips, Ketchup
T6	Hot Dogs, Coke, chips

Minimum Support (S=3)

Frequent 1 item set	Sup-count
Hot Dogs	4
Coke	3
Chips	4

Frequent 2 item set	Sup-Count
Coke, chips	3

Frequent item (1) → {coke, chips}

Rule 1: {Coke} → {chips}

$S \rightarrow \frac{3}{6} = 50\%$

$C \rightarrow \frac{3/6}{3/6} = 100\%$ Selected

Rule 2: {chips} → {Coke}

$S \rightarrow 50\%$

$C \rightarrow \frac{3/6}{4/6} = 75\%$ Selected.