Big Data Management and Analytics
Assignment 8

Assignment 8-1

Finding similar items

Suppose that the universal set is given by $\{1,\ldots,10\}$. Construct minhash signatures for the following sets:

(a) $S_1 = \{3,6,9\}$

(b) $S_2 = \{2,4,6,8\}$

(c) $S_3 = \{2,3,4\}$

1. Construct the signatures for the sets using the following list of permutations:
   - (1,2,3,4,5,6,7,8,9,10)
   - (10,8,6,4,2,9,7,5,3,1)
   - (4,7,2,9,1,5,3,10,6,8)

Assignment 8-1

i. Create first a characteristic matrix for each permutation

(a) Set one column with the shingles

(b) Set columns for each document

(c) Fill the document columns by setting a 1 where there is an occurence for each shingle, and 0 else

Element	S_1	S_2	S_3
1	0	0	0
2	0	1	1
3	1	0	1
4	0	1	1
5	0	0	0
6	1	1	0
7	0	0	0
8	0	1	0
9	1	0	0
10	0	0	0

Assignment 8-1

i. Create first a characteristic matrix for each permutation

Element	S_1	S_2	S_3
1	0	0	0
2	0	1	1
3	1	0	1
4	0	1	1
5	0	0	0
6	1	1	0
7	0	0	0
8	0	1	0
9	1	0	0
10	0	0	0

1st permutation

Element	S_1	S_2	S_3
10	0	0	0
8	0	1	0
6	1	1	0
4	0	1	1
2	0	1	1
9	1	0	0
7	0	0	0
5	0	0	0
3	1	0	1
1	0	0	0

2nd permutation

Element	S_1	S_2	S_3
4	0	1	1
7	0	0	0
2	0	1	1
9	1	0	0
1	0	0	0
5	0	1	0
3	1	0	1
10	0	1	0
6	1	1	0
8	0	1	0

3rd permutation

Assignment 8-1

ii. Compute the minhash for each permutation

Element	S_1	S_2	S_3
1	0	0	0
2	0	1	1
3	1	0	1
4	0	1	1
5	0	0	0
6	1	1	0
7	0	0	0
8	0	1	0
9	1	0	0
10	0	0	0

1st permutation

Select the first occurence of 1 per set and get the elements at which they can be found

Which leads to the following minhash:

$h(S_1) = 3$

$h(S_2) = 2$

$h(S_3) = 2$

Assignment 8-1

ii. Compute the minhash for each permutation

Element	S_1	S_2	S_3
10	0	0	0
8	0	1	0
6	1	1	0
4	0	1	1
2	0	1	1
9	1	0	0
7	0	0	0
5	0	0	0
3	1	0	1
1	0	0	0

2nd permutation

The same procedure for the 2nd permutation…

$h(S_1) = 6$
$h(S_2) = 8$
$h(S_3) = 4$

Assignment 8-1

ii. Compute the minhash for each permutation

Element	S_1	S_2	S_3
4	0	1	1
7	0	0	0
2	0	1	1
9	1	0	0
1	0	0	0
5	0	1	0
3	1	0	1
10	0	1	0
6	1	1	0
8	0	1	0

3rd permutation

…and for the 3rd permutation

Which leads to the following minhash:

$h(S_1) = 9$
$h(S_2) = 4$
$h(S_3) = 4$

This yields the following signatures:

$SIG(S_1) = \{3,6,9\}$
$SIG(S_2) = \{2,8,4\}$
$SIG(S_3) = \{2,4,4\}$

Assignment 8-1

2. Instead of using the previously given permutations use hash functions:

$$h_1(x) = x \; mod \; 10$$
$$h_2(x) = (2x + 1) \; mod \; 10$$
$$h_3(x) = (3x + 2) \; mod \; 10$$

Assignment 8-1

2. Instead of using the previously given permutations use hash functions:

i. Set up a table:

Element	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0			
2	0	1	1			
3	1	0	1			
4	0	1	1			
5	0	0	0			
6	1	1	0			
7	0	0	0			
8	0	1	0			
9	1	0	0			
10	0	0	0			

Assignment 8-1

ii. Compute the hash values (except for zero-rows):

Element	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0	-	-	-
2	0	1	1	2	5	8
3	1	0	1	3	7	1
4	0	1	1	4	9	4
5	0	0	0	-	-	-
6	1	1	0	6	3	0
7	0	0	0	-	-	-
8	0	1	0	8	7	6
9	1	0	0	9	9	9
10	0	0	0	-	-	-

Assignment 8-1

iii. Create a table for all hash functions and sets and initialize them with infinite distance:

e	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0	-	-	-
2	0	1	1	2	5	8
3	1	0	1	3	7	1
4	0	1	1	4	9	4
5	0	0	0	-	-	-
6	1	1	0	6	3	0
7	0	0	0	-	-	-
8	0	1	0	8	7	6
9	1	0	0	9	9	9
10	0	0	0	-	-	-

	S_1	S_2	S_3
h_1	∞	∞	∞
h_2	∞	∞	∞
h_3	∞	∞	∞

Assignment 8-1

iii. Create a table for all hash functions and sets and initialize them with infinite distance:

e	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0	-	-	-
2	0	1	1	2	5	8
3	1	0	1	3	7	1
4	0	1	1	4	9	4
5	0	0	0	-	-	-
6	1	1	0	6	3	0
7	0	0	0	-	-	-
8	0	1	0	8	7	6
9	1	0	0	9	9	9
10	0	0	0	-	-	-

Update only S_2, S_3!

S_2:
$\min(\infty, 2) = 2$
$\min(\infty, 5) = 5$
$\min(\infty, 8) = 8$

S_3:
$\min(\infty, 2) = 2$
$\min(\infty, 5) = 5$
$\min(\infty, 8) = 8$

Update of 2nd row:

	S_1	S_2	S_3
h_1	∞	2	2
h_2	∞	5	5
h_3	∞	8	8

Assignment 8-1

iii. Create a table for all hash functions and sets and initialize them with infinite distance:

e	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0	-	-	-
2	0	1	1	2	5	8
3	1	0	1	3	7	1
4	0	1	1	4	9	4
5	0	0	0	-	-	-
6	1	1	0	6	3	0
7	0	0	0	-	-	-
8	0	1	0	8	7	6
9	1	0	0	9	9	9

Update only S_1, S_3!

S_1:
$\min(\infty, 3) = 3$
$\min(\infty, 7) = 7$
$\min(\infty, 1) = 1$

S_3:
$\min(2,3) = 2$
$\min(5,7) = 5$
$\min(8,1) = 1$

Update of 3rd row:

	S_1	S_2	S_3
h_1	3	2	2
h_2	7	5	5
h_3	1	8	1

Assignment 8-1

iii. Create a table for all hash functions and sets and initialize them with infinite distance:

e	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0	-	-	-
2	0	1	1	2	5	8
3	1	0	1	3	7	1
4	0	1	1	4	9	4
5	0	0	0	-	-	-
6	1	1	0	6	3	0
7	0	0	0	-	-	-
8	0	1	0	8	7	6
9	1	0	0	9	9	9
10	0	0	0	-	-	-

Update only S_2, S_3!

S_2:
$\min(2,4) = 2$
$\min(5,9) = 5$
$\min(8,4) = 4$

S_3:
$\min(2,4) = 2$
$\min(5,9) = 5$
$\min(1,4) = 1$

Update of 4th row:

	S_1	S_2	S_3
h_1	3	2	2
h_2	7	5	5
h_3	1	4	1

Assignment 8-1

iii. Create a table for all hash functions and sets and initialize them with infinite distance:

e	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0	-	-	-
2	0	1	1	2	5	8
3	1	0	1	3	7	1
4	0	1	1	4	9	4
5	0	0	0	-	-	-
6	1	1	0	6	3	0
7	0	0	0	-	-	-
8	0	1	0	8	7	6
9	1	0	0	9	9	9
10	0	0	0	-	-	-

Update only S_1, S_2!

S_1:
$\min(3,6) = 3$
$\min(7,3) = 3$
$\min(1,0) = 0$

S_2:
$\min(2,6) = 2$
$\min(5,3) = 3$
$\min(4,0) = 0$

Update of 6th row:

	S_1	S_2	S_3
h_1	3	2	2
h_2	3	3	5
h_3	0	0	1

Assignment 8-1

iii. Create a table for all hash functions and sets and initialize them with infinite distance:

e	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0	-	-	-
2	0	1	1	2	5	8
3	1	0	1	3	7	1
4	0	1	1	4	9	4
5	0	0	0	-	-	-
6	1	1	0	6	3	0
7	0	0	0	-	-	-
8	0	1	0	8	7	6
9	1	0	0	9	9	9
10	0	0	0	-	-	-

Update only S_2!

S_2:
$\min(2,8) = 2$
$\min(3,7) = 3$
$\min(0,6) = 0$

Update of 8th row:

	S_1	S_2	S_3
h_1	3	2	2
h_2	3	3	5
h_3	0	0	1

Assignment 8-1

iii. Create a table for all hash functions and sets and initialize them with infinite distance:

e	S_1	S_2	S_3	$h_1(x)$	$h_2(x)$	$h_3(x)$
1	0	0	0	-	-	-
2	0	1	1	2	5	8
3	1	0	1	3	7	1
4	0	1	1	4	9	4
5	0	0	0	-	-	-
6	1	1	0	6	3	0
7	0	0	0	-	-	-
8	0	1	0	8	7	6
9	1	0	0	9	9	9
10	0	0	0	-	-	-

Update only S_1!

S_1:
$\min(3,9) = 3$
$\min(3,9) = 3$
$\min(0,9) = 0$

Update of 8th row:

	S_1	S_2	S_3
h_1	3	2	2
h_2	3	3	5
h_3	0	0	1

Assignment 8-1

iii. Create a table for all hash functions and sets and initialize them with infinite distance:

	S_1	S_2	S_3
h_1	3	2	2
h_2	3	3	5
h_3	0	0	1

This yields the following signatures:

$$SIG(S_1) = (3,3,0)$$
$$SIG(S_2) = (2,3,0)$$
$$SIG(S_3) = (2,5,1)$$

Assignment 8-1

3. How does the estimated Jaccard similarity from (1.) and (2.) compare with the true Jaccard similarity of the original data? How to reduce deviations in the approximated Jaccard similarities?

RECAP: Jaccard similarity:

$$d_{Jaccard}(D_1, D_2) = \frac{|D_1 \cap D_2|}{|D_1 \cup D_2|}$$

Assignment 8-1

RECAP: Jaccard similarity:

$$d_{Jaccard}(D_1, D_2) = \frac{|D_1 \cap D_2|}{|D_1 \cup D_2|}$$

(a) $S_1 = \{3,6,9\}$
(b) $S_2 = \{2,4,6,8\}$
(c) $S_3 = \{2,3,4\}$

i. Actual Jaccard similarity:

	S_1	S_2	S_3
S_1	-	$1/6$	$1/5$
S_2	-	-	$2/5$
S_3	-	-	-

Assignment 8-1

ii. Permutation estimated Jaccard similarity:

(a) $S_1 = \{3,6,9\}$
(b) $S_2 = \{2,8,4\}$
(c) $S_3 = \{2,4,4\}$

	S_1	S_2	S_3
S_1	-	$0/6$	$0/5$
S_2	-	-	$2/3$
S_3	-	-	-

Assignment 8-1

iii. Hash estimated Jaccard similarity:

(a) $S_1 = \{3,3,0\}$
(b) $S_2 = \{2,3,0\}$
(c) $S_3 = \{2,5,1\}$

	S_1	S_2	S_3
S_1	-	$2/3$	$0/5$
S_2	-	-	$1/5$
S_3	-	-	-

Assignment 8-1

iv. Comparison:

Actual J. similarity

	S_1	S_2	S_3
S_1	-	$1/6$	$1/5$
S_2	-	-	$2/5$
S_3	-	-	-

Perm. est. J similarity

	S_1	S_2	S_3
S_1	-	$0/3$	$0/3$
S_2	-	-	$2/3$
S_3	-	-	-

Hash. est. J similarity

	S_1	S_2	S_3
S_1	-	$2/3$	$0/3$
S_2	-	-	$1/3$
S_3	-	-	-

Estimation of the actual J. similarity is rather poor, why?

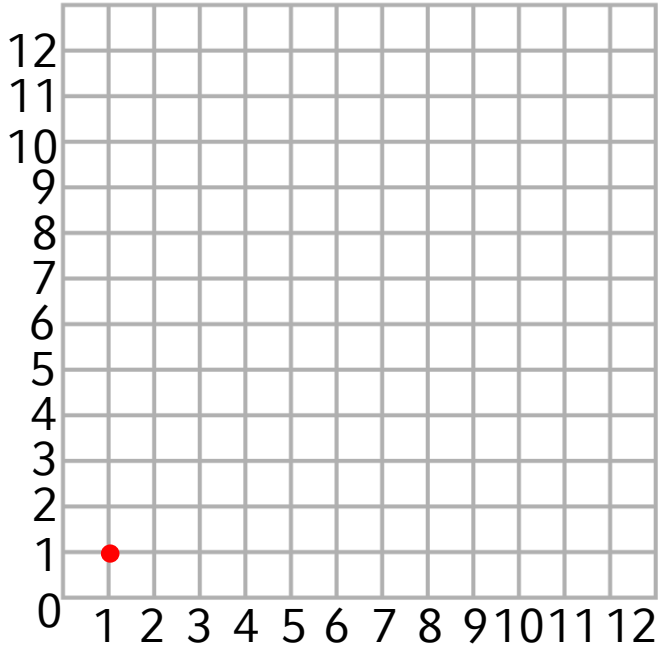
→ Too small minhash vectors. Get more permutations of the universal set or more hash functions to extend the minhash vectors!

Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}2\\1\end{pmatrix}$	$\begin{pmatrix}4\\9\end{pmatrix}$	$\begin{pmatrix}4\\8\end{pmatrix}$	$\begin{pmatrix}10\\4\end{pmatrix}$	$\begin{pmatrix}9\\3\end{pmatrix}$	$\begin{pmatrix}2\\3\end{pmatrix}$	$\begin{pmatrix}11\\3\end{pmatrix}$	$\begin{pmatrix}12\\12\end{pmatrix}$	$\begin{pmatrix}12\\11\end{pmatrix}$	$\begin{pmatrix}11\\12\end{pmatrix}$	$\begin{pmatrix}4\\2\end{pmatrix}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Wait for the first 6 points to arrive

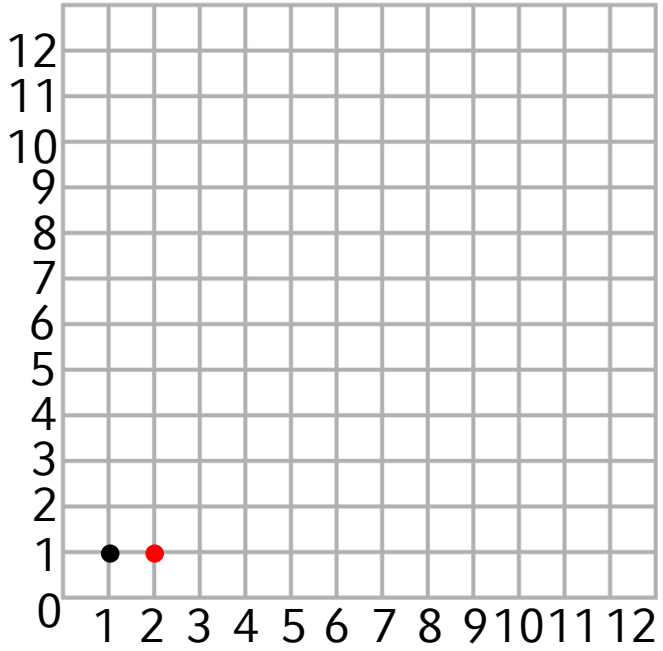


Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}2\\1\end{pmatrix}$	$\begin{pmatrix}4\\9\end{pmatrix}$	$\begin{pmatrix}4\\8\end{pmatrix}$	$\begin{pmatrix}10\\4\end{pmatrix}$	$\begin{pmatrix}9\\3\end{pmatrix}$	$\begin{pmatrix}2\\3\end{pmatrix}$	$\begin{pmatrix}11\\3\end{pmatrix}$	$\begin{pmatrix}12\\12\end{pmatrix}$	$\begin{pmatrix}12\\11\end{pmatrix}$	$\begin{pmatrix}11\\12\end{pmatrix}$	$\begin{pmatrix}4\\2\end{pmatrix}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Wait for the first 6 points to arrive

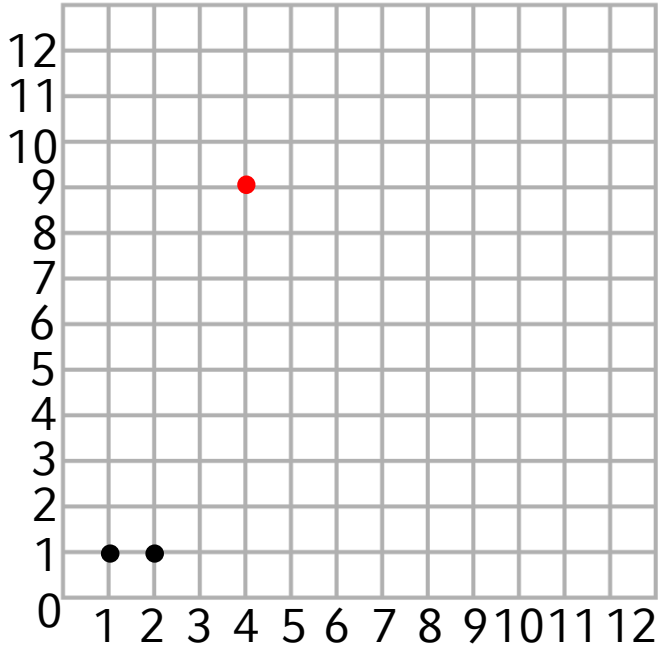


Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Wait for the first 6 points to arrive

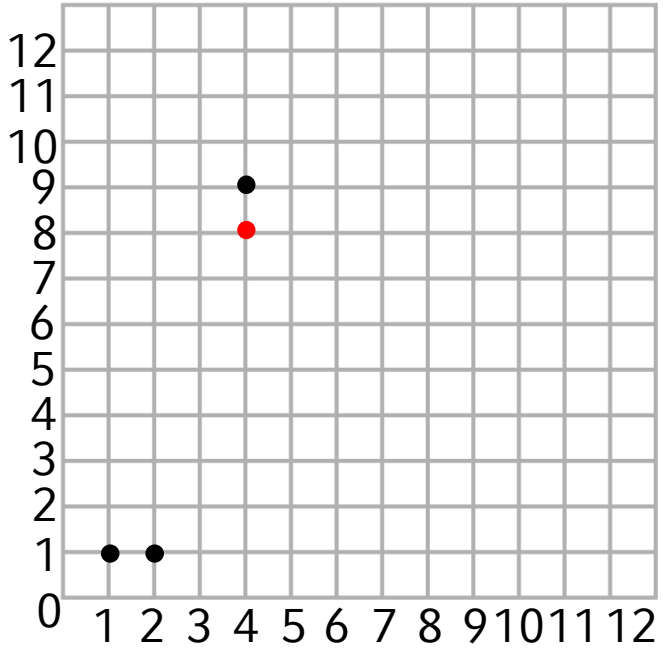


Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\begin{pmatrix}1\\1\end{pmatrix}$	$\begin{pmatrix}2\\1\end{pmatrix}$	$\begin{pmatrix}4\\9\end{pmatrix}$	$\begin{pmatrix}4\\8\end{pmatrix}$	$\begin{pmatrix}10\\4\end{pmatrix}$	$\begin{pmatrix}9\\3\end{pmatrix}$	$\begin{pmatrix}2\\3\end{pmatrix}$	$\begin{pmatrix}11\\3\end{pmatrix}$	$\begin{pmatrix}12\\12\end{pmatrix}$	$\begin{pmatrix}12\\11\end{pmatrix}$	$\begin{pmatrix}11\\12\end{pmatrix}$	$\begin{pmatrix}4\\2\end{pmatrix}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Wait for the first 6 points to arrive

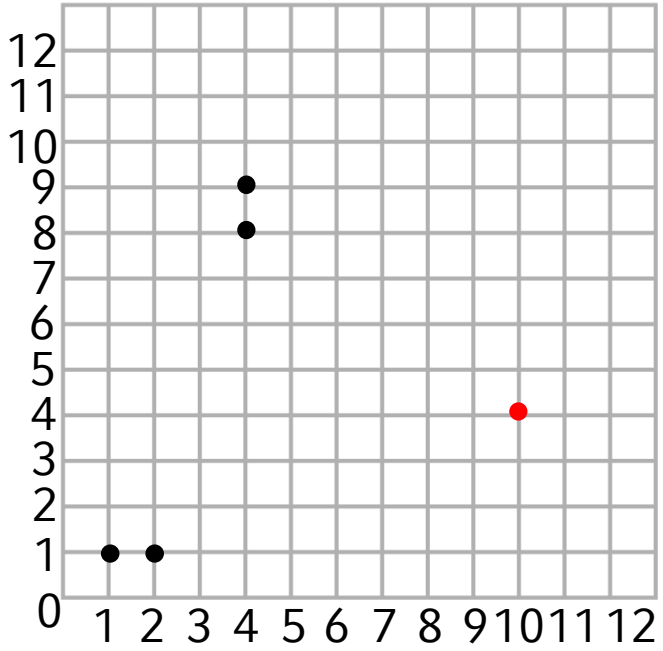


Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Wait for the first 6 points to arrive

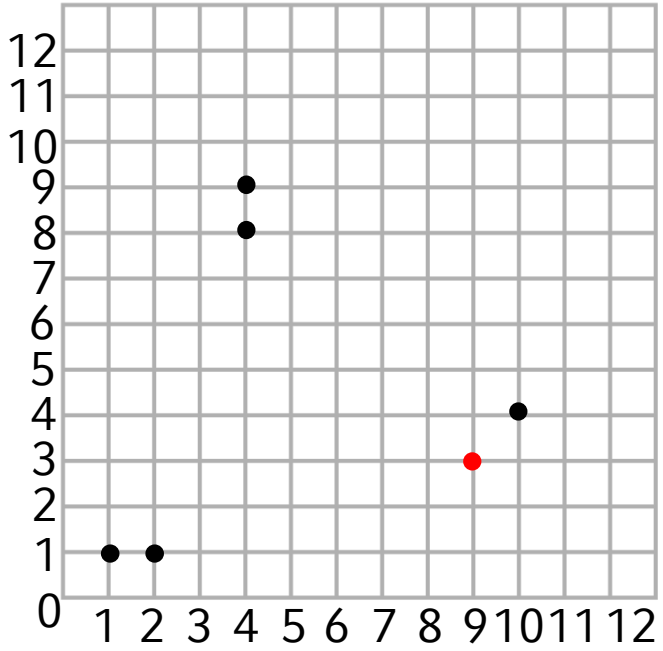


Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Wait for the first 6 points to arrive

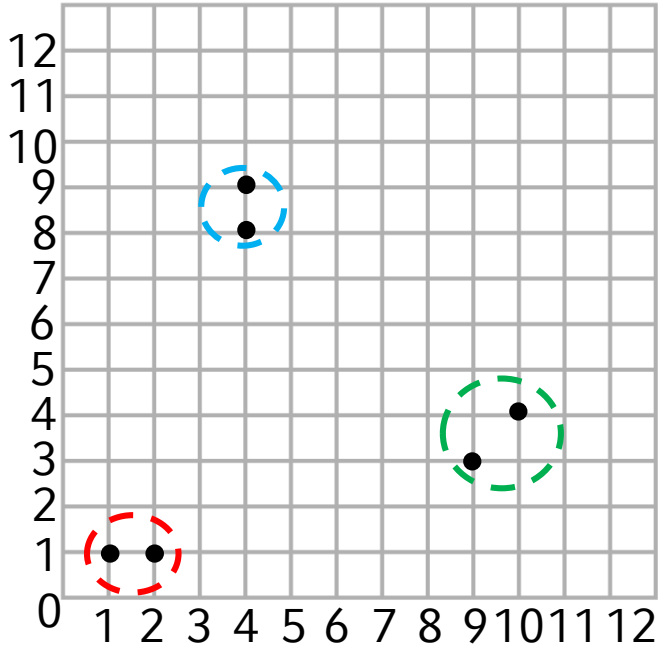


Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Apply standard k-Means to create q clusters.

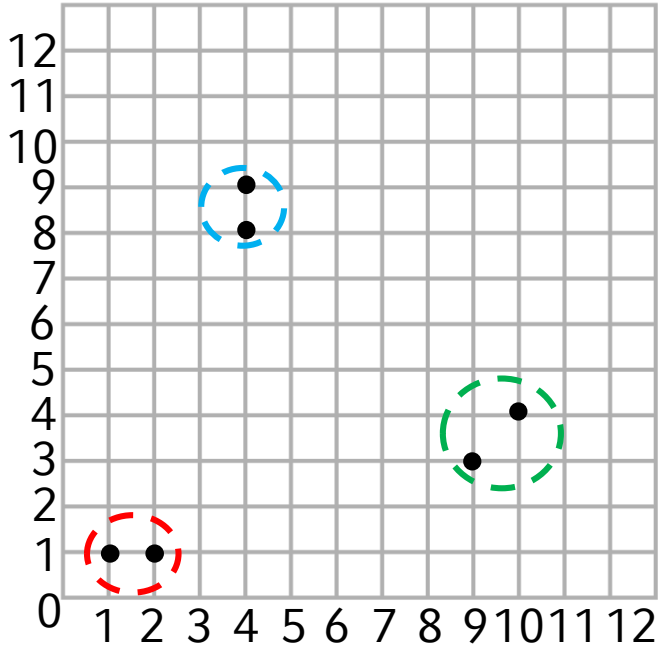


Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

For each discovered cluster, assign a unique ID and create ist micro-cluster summary.

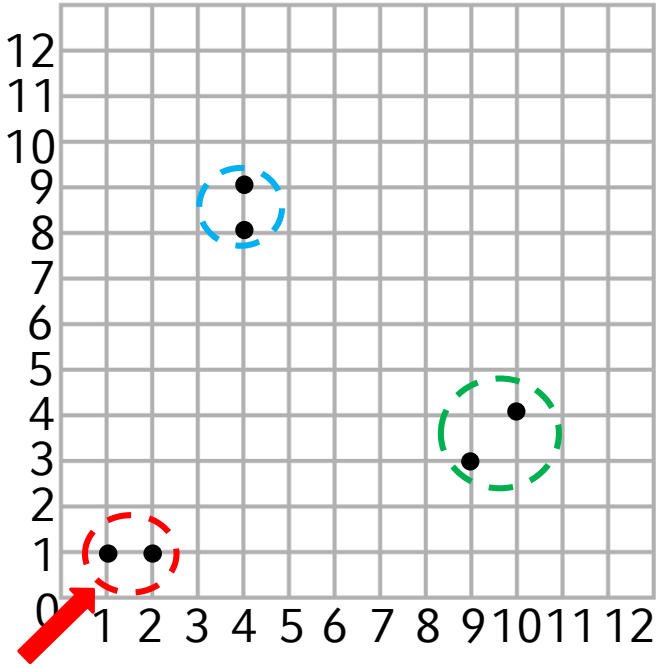


Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

For each discovered cluster, assign a unique ID and create ist micro-cluster summary.



$$CFT_1 = (CF2^X, CF1^X, CF2^t, CF1^t, n)$$

$$\binom{1^2 + 2^2}{1^2 + 1^2} = \binom{5}{2} \qquad \binom{1+2}{1+1} = \binom{3}{2} \qquad 1^2 + 2^2 = 5 \qquad 1 + 2 = 3 \qquad 2$$

Assignment 8-2 - CluStream

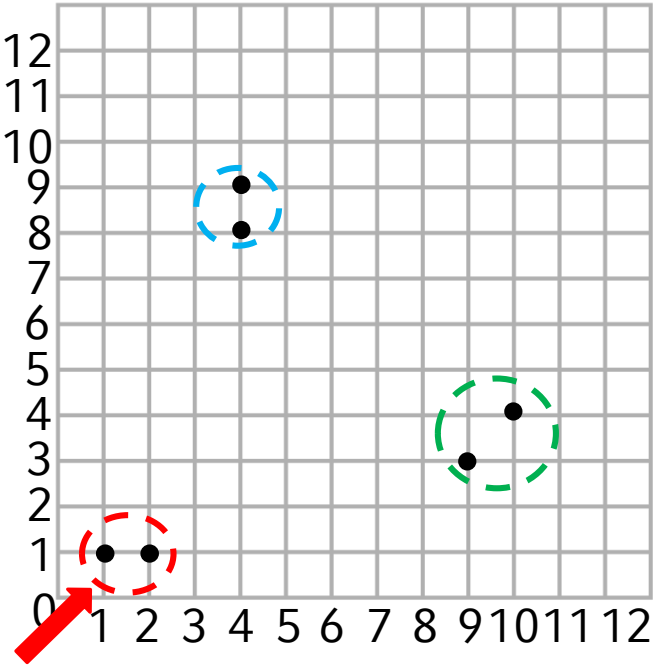
t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

For each discovered cluster, assign a unique ID and create ist micro-cluster summary.

$$CFT_1 = (\binom{5}{2}, \binom{3}{2}, 5,3,2)$$

$$center = {}^{\binom{3}{2}}/_{2} = \binom{1.5}{1}$$



Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

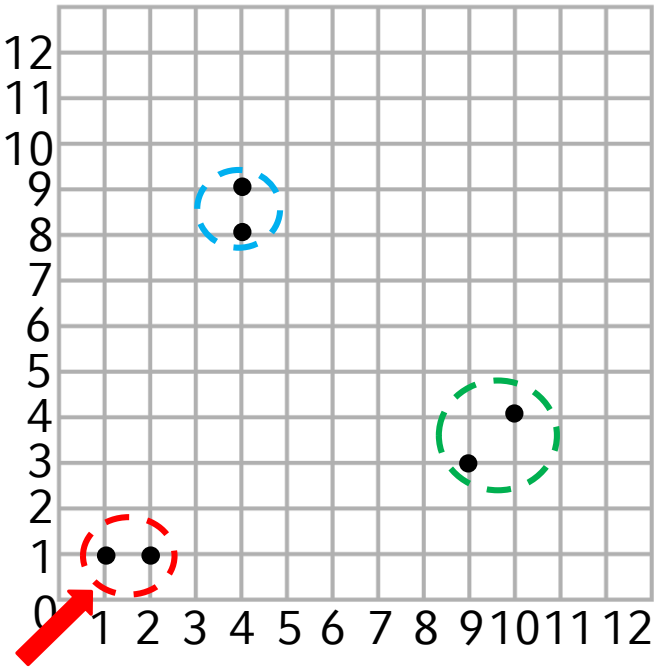
$$CFT_1 = (\binom{5}{2}, \binom{3}{2}, 5,3,2)$$

$$radius = \sqrt{|CF2^X|/n - \left(|CF1^X|/n\right)^2}$$

$$= \sqrt{\binom{5}{2}/2 - \left(\binom{3}{2}/2\right)^2}$$

$$= \sqrt{\binom{2.5}{1} - \left(\binom{1.5}{1}\right)^2} \quad = \sqrt{\binom{2.5}{1} - \binom{2.25}{1}}$$

$$= \sqrt{(2.5 - 2.25) + (1 - 1)} = 0.5$$



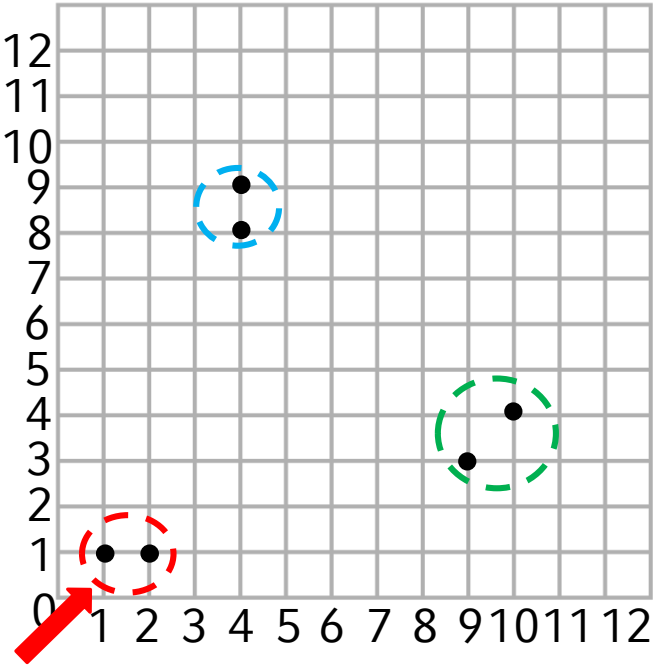
Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

For each discovered cluster, assign a unique ID and create ist micro-cluster summary.

id	$CF2^x$	$CF1^x$	$CF2^t$	$CF1^t$	n	cen	rad
1	$\binom{5}{2}$	$\binom{3}{2}$	5	3	2	$\binom{1.5}{1}$	0.5



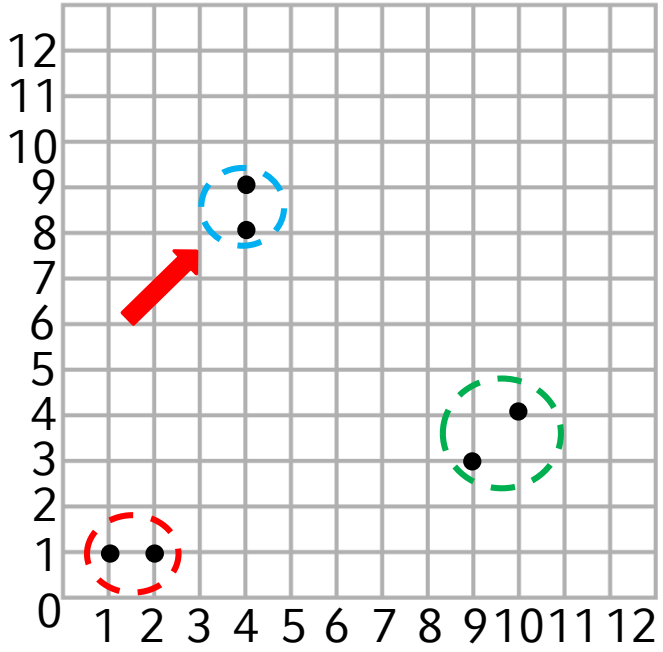
Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

For each discovered cluster, assign a unique ID and create ist micro-cluster summary.

id	$CF2^x$	$CF1^x$	$CF2^t$	$CF1^t$	n	cen	rad
1	$\binom{5}{2}$	$\binom{3}{2}$	5	3	2	$\binom{1.5}{1}$	0.5
2	$\binom{32}{145}$	$\binom{8}{17}$	25	7	2	$\binom{4}{8.5}$	0.5



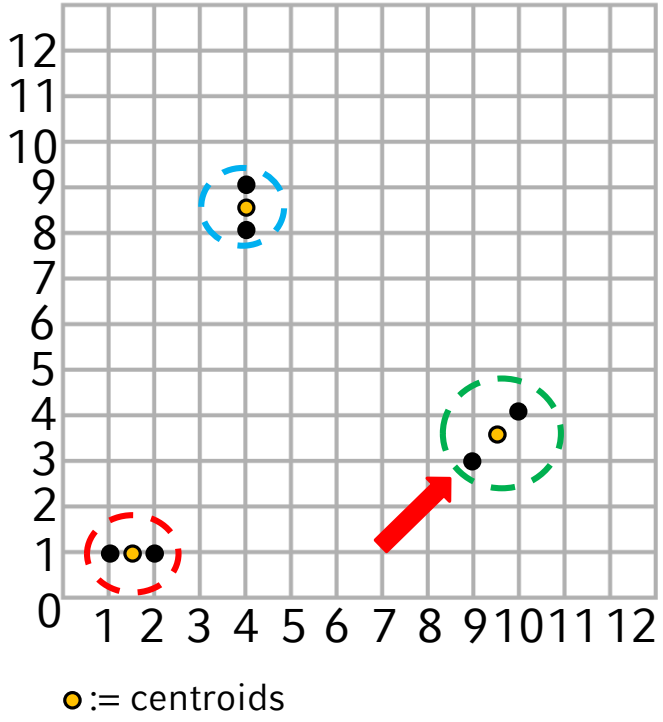
Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

For each discovered cluster, assign a unique ID and create ist micro-cluster summary.

id	$CF2^X$	$CF1^X$	$CF2^t$	$CF1^t$	n	cen	rad
1	$\binom{5}{2}$	$\binom{3}{2}$	5	3	2	$\binom{1.5}{1}$	0.5
2	$\binom{32}{145}$	$\binom{8}{17}$	25	7	2	$\binom{4}{8.5}$	0.5
3	$\binom{181}{25}$	$\binom{19}{7}$	61	11	2	$\binom{9.5}{3.5}$	0.7



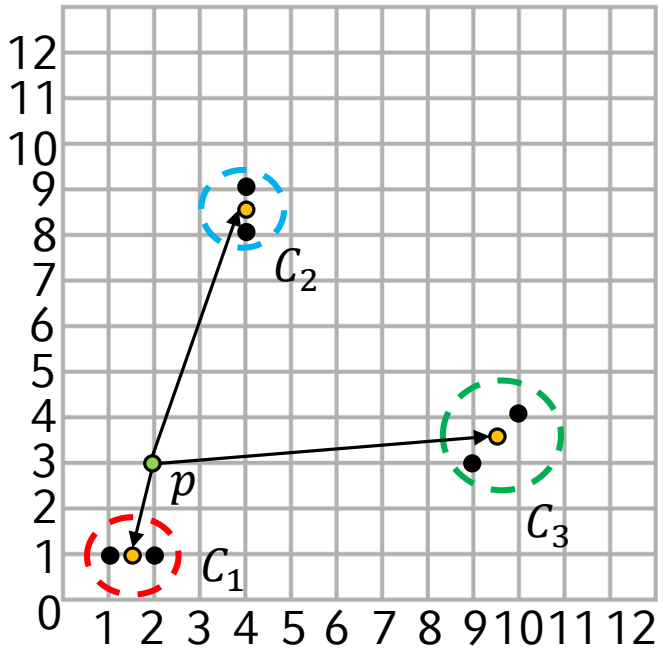
Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Compute distance between point p and each of the q maintained micro-cluster centroids.

$$d_{L2}(C_1, p) = 2.06$$
$$d_{L2}(C_2, p) = 5.85$$
$$d_{L2}(C_3, p) = 7.51$$



Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

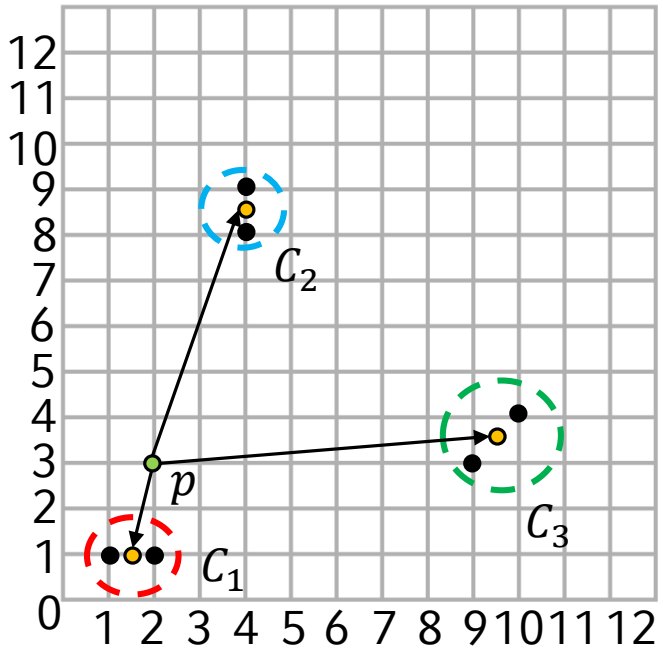
- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Select a cluster clu as the closest micro-cluster to p.

$d_{L2}(C_1, p) = 2.06$

$d_{L2}(C_2, p) = 5.85$

$d_{L2}(C_3, p) = 7.51$



Assignment 8-2 - CluStream

t	1	2	3	4	5	6	7	8	9	10	11	12
p	$\binom{1}{1}$	$\binom{2}{1}$	$\binom{4}{9}$	$\binom{4}{8}$	$\binom{10}{4}$	$\binom{9}{3}$	$\binom{2}{3}$	$\binom{11}{3}$	$\binom{12}{12}$	$\binom{12}{11}$	$\binom{11}{12}$	$\binom{4}{2}$

- initPoints = 6
- q = 3
- Factor of clu radius t = 5

Find the maximum boundary of clu
:= factor of t of clu radius= $5 \cdot 0.5 = 2.5$

$d_{L2}(C_1, p) = 2.06$

is p within the maximum cluster boundary of C_1?

$(x - x_{c_1 centroid})^2 + (y - y_{c_1 centroid})^2 \leq (t \cdot rad_{C1})^2$?

$(2 - 1.5)^2 + (3 - 1)^2 \leq (5 \cdot 0.5)^2$?