MACHINE LEARNING

EXERCISES

Elements of unsupervised machine learning

All the course material is available on the web site

Course web site: https://unica-ml.github.io/#

Exercise 1

Given the following data:

X1=(1 1)’
X2=(1 0)’
X3=(0 1)’
X4=(5 0)’
X5=(4 1)’
X6=(3 2)’

And assuming c=2 (c is the final number of clusters that we want)

1) Apply the k-means algorithm, using as initial centroids C1= (0,0)’; C2= (1,1)’
2) Apply the Single Linkage Clustering algorithm choosing an appropriate measure of “similarity”.

---

1) Apply the k-means algorithm, using as initial centroids C1= (0,0)’; C2= (1,1)’

- **begin initialize** *n*, *c*, $\mathbf{m}_1$, $\mathbf{m}_2$, ... , $\mathbf{m}_c$
- **do** classify *n* patterns according to nearest $\mathbf{m}_i$
- recompute $\mathbf{m}_i$
- **until** no change in $\mathbf{m}_i$
- **return** $\mathbf{m}_1$, $\mathbf{m}_2$, ... , $\mathbf{m}_c$
- **end**

Step 1

C1= [0,0]'; C2= [1,1]';

Distances between patterns and centroids

	C1	C2
x1	1.41	**0.00**
x2	**1.00**	1.00
x3	**1.00**	1.00
x4	5.00	**4.12**
x5	4.12	**3.00**
x6	3.61	**2.24**

Cluster:
C1={x2,x3}; **C**2={x1,x4,x5,x6}
New centroids:
C1= [0.5, 0.5]'; C2= [3.25, 1]';

Step 2
C1= [0.5, 0.5]'; C2= [3.25, 1]';

Distances between patterns and centroids

	C1	C2
x1	**0.71**	2.25
x2	**0.71**	2.46
x3	**0.71**	3.25
x4	4.53	**2.02**
x5	3.54	**0.75**
x6	2.92	**1.03**

Cluster:
C1={x1,x2,x3}; **C**2={x4,x5,x6}
New centroids:
C1= [0.667 0.667]'; C2= [4 1]';

Step 3
C1= [0.667 0.667]'; C2= [4 1]';

Distanze

	C1	C2
x1	**0.47**	3
x2	**0.75**	3.16
x3	**0.75**	4
x4	4.38	**1.41**
x5	3.35	**0**
x6	2.69	**1.41**

Cluster:
C1={x1,x2,x3}; **C**2={x4,x5,x6}
New centroids:
C1= [0.667 0.667]'; C2= [4 1]';

The algorithm has reached the convergence; the final clustering is
{x1, x2, x3} ; {x4, x5, x6}

2) Apply the Single Linkage Clustering algorithm choosing an appropriate measure of similarity.

Single Linkage Clustering Algorithm

1. Choose an appropriate measure of similarity S(a,b) between two "patterns".
   We can choose $S(x1,x2)=1/d_{12}$, where d_{12} is the Euclidean distance between x_1 and x_2
   For simplicity, we define the similarity between two "clusters" as the similarity between their centroids, but other choices are possible (see Chapter 7 of the course).

2. Initialize the algorithm by assuming that each "pattern" is a separate cluster.

3. Identify the two most similar clusters and merge into a new cluster.

4. Repeat step 2 until you obtain *c=2* clusters

STEP 1

Mutual distances

	C1{X1}	C2{X2}	C3{X3}	C4{X4}	C5{X5}	C6{X6}
C1{X1}	0.000	1.000	1.000	4.123	3.000	2.236
C2{X2}		0.000	1.414	4.000	3.162	2.828
C3{X3}			0.000	5.099	4.000	3.162
C4{X4}				0.000	1.414	2.828
C5{X5}					0.000	1.414
C6{X6}						0.000

Similarity

	C1{X1}	C2{X2}	C3{X3}	C4{X4}	C5{X5}	C6{X6}
C1 {X1}		1.000	1.000	0.243	0.333	0.447
C2 {X2}			0.707	0.250	0.316	0.354
C3{X3}				0.196	0.250	0.316
C4{X4}					0.707	0.354
C5{X5}						0.707
C6{X6}						

STEP 2

New cluster C1{X1,X2} ; m1= (1 1/2)’

Mutual distance

	C1{X1, x2}	C3{X3}	C4{X4}	C5{X5}	C6{X6}
C1{X1,X2}	0.000	1.118	4.031	3.041	2.500
C3{ X3}		0.000	5.099	4.000	3.162
C4{X4}			0.000	1.414	2.828
C5{X5}				0.000	1.414
C6{X6}					0.000

Similarity

	C1{X1}	C3{X3}	C4{X4}	C5{X5}	C6{X6}
C1{X1,X2}		0.894	0.248	0.329	0.400
C3{ X3}			0.196	0.250	0.316
C4{X4}				0.707	0.354
C5{X5}					0.707
C6{X6}					

STEP 3

New cluster C1{X1,X2,X3} ; m1= (2/3 2/3)'

Mutual distance

	C1{X1,X2,X3}	C4{X4}	C5{X5}	C6{X6}
C1{X1,X2,X3}	0.000	4.384	3.350	2.687
C4{X4}		0.000	1.414	2.828
C5{X5}			0.000	1.414
C6{X6}				0.000

Similarity

	C1{X1,X2,X3}	C4{X4}	C5{X5}	C6{X6}
C1{X1,X2,X3}		0.228	0.299	0.372
C4{X4}			0.707	0.354
C5{X5}				0.707
C6{X6}				

STEP 4

New cluster **C**4{X4,X5 } ;

m1=(2/3,2/3); m4= (9/2 1/2)'; m6=(3 2)'

Mutual distance

	C1{X1,X2,X3}	C4{X4,X5 }	C6{X6}
C1{X1,X2,X3}	0.000	3.837	2.687
C4{X4,X5 }		0.000	2.121
C6{X6}			0.000

Similarity

	C1{X1,X2,X3}	C4{X4,X5 }	C6{X6}
C1{X1,X2,X3}		0.261	0.372
C4{X4,X5 }			0.471
C6{X6}			

Final clusterization: {x1,x2,x3 },{,x4,x5,x6}

the highest similarities that indicate which are the cluster and the pattern to merge are shown in yellow.
Usually, in case of equal value, pair of clusters to be merged is chosen randomly. In this exercise, we have selected the cluster with lower index.

Exercise 2

Cluster 1	Cluster 2
X1=(1 1)'	X4=(5 0)'
X2=(1 0)'	X5=(4 1)'
X3=(0 1)'	X6=(3 2)'

Given the patterns in the above table, check if the two clusters (Cluster 1 and Cluster 2) represent the "natural" classes according to the following "cluster-validity" evaluation functions:

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2 \qquad\qquad J_d = \det(\mathbf{S}_W) = \left| \sum_{i=1}^{c} \mathbf{S}_i \right|$$

compared to the case where the pattern X6= (3 2)' is assigned to the Cluster 1.

Functions for the first case (the two clusters in the above table)

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2$$

$$m_1 = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix}; \;\; m_2 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2 = \sum_{\mathbf{x} \in D_1} \|\mathbf{x} - \mathbf{m}_1\|^2 + \sum_{\mathbf{x} \in D_2} \|\mathbf{x} - \mathbf{m}_2\|^2$$

=4/3+4=16/3≅5.333

$$J_d = \det(\mathbf{S}_W) = \left| \sum_{i=1}^{c} \mathbf{S}_i \right|$$

$$S_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t \Rightarrow$$

$$S_1 = \sum_{\mathbf{x} \in D_1} (\mathbf{x} - \mathbf{m}_1)(\mathbf{x} - \mathbf{m}_1)^t; \; S_2 = \sum_{\mathbf{x} \in D_2} (\mathbf{x} - \mathbf{m}_2)(\mathbf{x} - \mathbf{m}_2)^t$$

$$S_W = \frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} + \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 8 & -7 \\ -7 & 8 \end{pmatrix}$$

$$J_d = \det(\mathbf{S}_W) = 15/9 = 1.6667$$

Functions for the second case (the pattern X6= (3 2)' is assigned to the Cluster 1)

Cluster 1	Cluster 2
X1=(1 1)’ X2=(1 0)’ X3=(0 1)’ X6=(3 2)’	X4=(5 0)’ X5=(4 1)’

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2$$

$$m_1 = \begin{pmatrix} 1.25 \\ 1 \end{pmatrix}; \quad m_2 = \begin{pmatrix} 4.5 \\ 0.5 \end{pmatrix}$$

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2 = \sum_{\mathbf{x} \in D_1} \|\mathbf{x} - \mathbf{m}_1\|^2 + \sum_{\mathbf{x} \in D_2} \|\mathbf{x} - \mathbf{m}_2\|^2$$

=6.75 +1=7.75 (in the first case was 5.333)

$$J_d = \det(\mathbf{S}_W) = \left| \sum_{i=1}^{c} \mathbf{S}_i \right|$$

$$S_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t =$$

$$\sum_{\mathbf{x} \in D_1} (\mathbf{x} - \mathbf{m}_1)(\mathbf{x} - \mathbf{m}_1)^t + \sum_{\mathbf{x} \in D_2} (\mathbf{x} - \mathbf{m}_2)(\mathbf{x} - \mathbf{m}_2)^t$$

$$= \begin{pmatrix} 4.75 & 2 \\ 2 & 2 \end{pmatrix} + \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix} = \begin{pmatrix} 5.25 & 1.5 \\ 1.5 & 2.5 \end{pmatrix}$$

$J_d = \det(\mathbf{S}_W) = 10.875$ (in the first case was 1.6667)

To sum up:

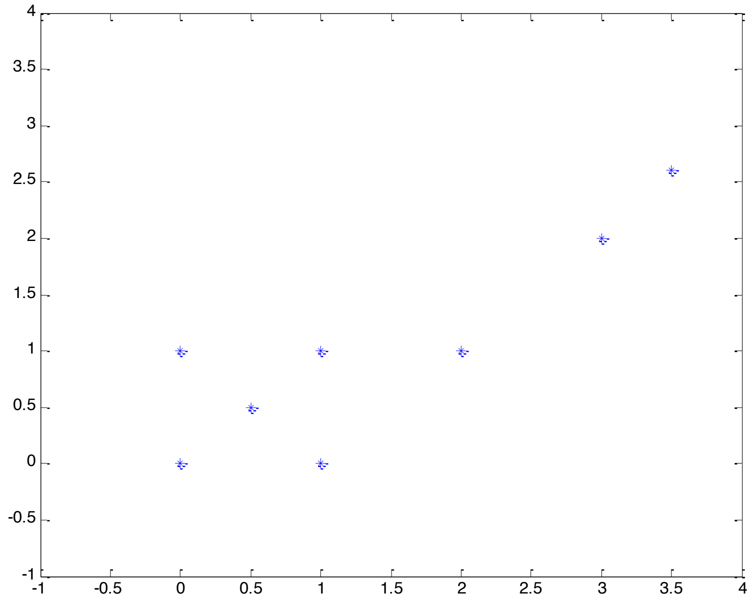
Clustering ‘1’ **Clustering ‘2’**

Clustering ‘1’	Clustering ‘2’
Je=5.33	**Je=7.75**
Jd=1.66	**Jd=10.87**

The “cluster-validity” evaluation functions used shows that the first clustering is the best one.

Exercise 3

Cluster 1	Cluster 2
X1=(0 0)’ X2=(0 1)’ X3=(0.5 0.5)’ X4=(1 0)’ X5=(1 1)’ X6=(2 1)’	X7=(3 2)’ X8=(3.5 2.6)’



Given the patterns in the table, check if the two clusters (Cluster 1 and Cluster 2) represent the “natural” classes according to the following “cluster-validity” evaluation functions:

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2 \qquad\qquad J_d = \det(\mathbf{S}_W) = \left|\sum_{i=1}^{c} \mathbf{S}_i\right|$$

compared to the case where the pattern X6= (3 2)' is assigned to the cluster 2

Functions for the first case (the two clusters in the above table)

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x}\in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2$$

$$m_1 = \begin{pmatrix} 0.75 \\ 0.5833 \end{pmatrix}; \quad m_2 = \begin{pmatrix} 3.25 \\ 2.3 \end{pmatrix}$$

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x}\in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2 = \sum_{\mathbf{x}\in D_1} \|\mathbf{x} - \mathbf{m}_1\|^2 + \sum_{\mathbf{x}\in D_2} \|\mathbf{x} - \mathbf{m}_2\|^2$$ =4.083 + 0.305 =4.388

$$J_d = \det(\mathbf{S}_W) = \left|\sum_{i=1}^{c} \mathbf{S}_i\right|$$

$$S_i = \sum_{\mathbf{x}\in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t \Rightarrow$$

$$S_1 = \sum_{\mathbf{x}\in D_1} (\mathbf{x} - \mathbf{m}_1)(\mathbf{x} - \mathbf{m}_1)^t; \; S_2 = \sum_{\mathbf{x}\in D_2} (\mathbf{x} - \mathbf{m}_2)(\mathbf{x} - \mathbf{m}_2)^t$$

$$S_W = \begin{pmatrix} 2.875 & 0.625 \\ 0.625 & 1.208 \end{pmatrix} + \begin{pmatrix} 0.125 & 0.15 \\ 0.15 & 0.18 \end{pmatrix} = \begin{pmatrix} 3 & 0.7750 \\ 0.775 & 1.3883 \end{pmatrix}$$

$$J_d = \det(\mathbf{S}_W) = 3.5644$$

Functions for the second case (the pattern X6= (3 2)' is assigned to the cluster 2)

Cluster 1	Cluster 2
X1=(0 0)’ X2=(0 1)’ X3=(0.5 0.5)’ X4=(1 0)’ X5=(1 1)’	X6=(2 1)’ X7=(3 2)’ X8=(3.5 2.6)’

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x}\in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2$$

$$m_1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}; \quad m_2 = \begin{pmatrix} 2.833 \\ 1.866 \end{pmatrix}$$

$$J_e = \sum_{i=1}^{c} \sum_{\mathbf{x}\in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2 = \sum_{\mathbf{x}\in D_1} \|\mathbf{x} - \mathbf{m}_1\|^2 + \sum_{\mathbf{x}\in D_2} \|\mathbf{x} - \mathbf{m}_2\|^2$$

= 2.000 + 2.473 =4.473 (in the first case was 4.388)

$$J_d = \det(\mathbf{S}_W) = \left|\sum_{i=1}^{c} \mathbf{S}_i\right|$$

$$S_i = \sum_{\mathbf{x}\in D_i} (\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^t = \sum_{\mathbf{x}\in D_1} (\mathbf{x}-\mathbf{m}_1)(\mathbf{x}-\mathbf{m}_1)^t + \sum_{\mathbf{x}\in D_2} (\mathbf{x}-\mathbf{m}_2)(\mathbf{x}-\mathbf{m}_2)^t =$$

$$= \begin{pmatrix} 2.167 & 1.233 \\ 1.233 & 2.307 \end{pmatrix}$$

$J_d = \det(\mathbf{S}_W) = 3.4767$ (in the first case was 3.564)

To sum up:

Clustering '1'	Clustering '2'
Je=4.388	Je=4.473
Jd=3.564	Jd=3.477

The "cluster-validity" evaluation functions used show **conflicting results**. The results of the function Je indicate that the first clustering is the best one; Jd indicates that the second clustering is the best one.

Exercise 4

Apply the Basic Iterative Minimum-Square-Error Clustering algorithm to the patterns in the table below to obtain c=3 clusters.

X1	X2	X3	X4	X5	X6	X7	X8	X9
1	1	0	5	4	4	6	7	7.5
1	0	1	8	7	7.5	-1	0.5	1

Note 1: usually, the initial clusters are selected by c random "centroids" and assigning each pattern to the nearest centroid. In this exercise, we start from a fully random clustering as follows:

Cluster 1: 1 2 9
Cluster 2: 3 4 5
Cluster 3: 6 7 8

Note 2: this clustering algorithm requires that a pattern is selected randomly for each iteration of the clustering algorithm.
In this exercise, for simplicity, patterns are selected following the order x1, ..., x9.

Basic Iterative Minimum - Square - Error Clustering

begin initialize $n, c, \mathbf{m}_1, \ldots, \mathbf{m}_c$

  do randomly select a pattern $\mathbf{x}'$

    $i \leftarrow \arg\min_{i'} \|\mathbf{m}_{i'} - \mathbf{x}'\|$ (classify $\mathbf{x}'$)

    if $n_i \neq 1$ **then** *compute*

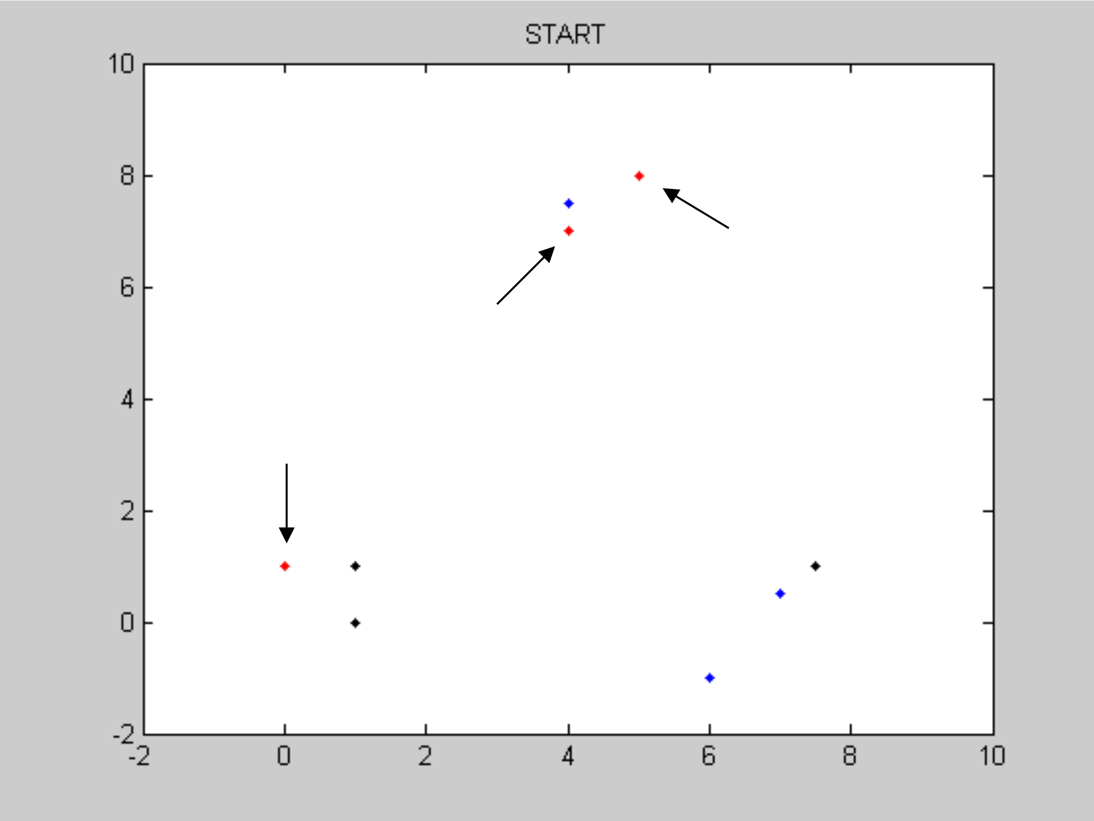
$$\rho_j = \begin{cases} \frac{n_j}{n_j+1}\|\mathbf{x}' - \mathbf{m}_j\|^2 & if \quad j \neq i \\ \frac{n_j}{n_j-1}\|\mathbf{x}' - \mathbf{m}_i\|^2 & if \quad j = i \end{cases}$$

    if $\rho_k \leq \rho_j \quad \forall j$ **then** transfer $\mathbf{x}'$ in D_k

      recompute J_e, $\mathbf{m}_i$, $\mathbf{m}_k$

  until no change in J_e n attempts

end



Start with a random “clustering” (see the above picture).

The colors represent the initial (wrong) allocation of patterns to clusters. It is seen that, for example, the cluster 'red' has its patterns (indicated by arrows) belonging to two different (natural) clusters.

Calculate the initial centroids:

m_1 =(3.1667   0.6667)’
m_2 =(3.0000   5.3333)’
m_3 =(5.6667   2.3333)’

Let’s recall the basic equations of the Iterative Minimum-Square-Error Clustering algorithm:

$$ro\_destination = \frac{n_j}{n_j+1}\left\|\mathbf{x}' - \mathbf{m}_j\right\|^2; \quad ro\_source = \frac{n_j}{n_j-1}\left\|\mathbf{x}' - \mathbf{m}_j\right\|^2$$

“Source” and “Destination” indicates respectively the clusters that give and receive the pattern x’.

We move a given pattern x’ from the “source” cluster to the “destination” cluster.

pattern 1

cluster “source” 1

$$\rho_1 = \frac{n_1}{n_1-1}\left\|x - m_1\right\|^2 = \frac{3}{2}\left\|\begin{matrix}1-3.16 \\ 1-0.\overline{6}\end{matrix}\right\|^2 \cong \frac{3}{2}\cdot 4.8055 \cong 7.2$$

cluster “destination” 2

$$\rho_2 = \frac{n_2}{n_2+1}\|x-m_2\|^2 = \frac{3}{4}\left\|\begin{matrix}1-3\\1-5.\overline{3}\end{matrix}\right\|^2 = \frac{3}{4}\cdot 22.\overline{7} \cong 17.08$$

(we can do the same calculation for cluster “destination” 3)

We obtain the following values:

cluster “source” 1: 7.2083
cluster “destination” 2: 17.0833
cluster “destination” 3: 17.6667

Therefore:
$\rho_1 < \rho_i \ \forall i$

ro 'source'<= ro 'destination' -> DO NOT move the pattern

The reduction of the error due to the elimination of the pattern from the cluster 1 is less than the increase of the error due to the adding of the pattern to the other clusters.

pattern 2
cluster 'source'1
ro cluster 'source': 7.7083
ro cluster 'destination' : 24.3333, 20.4167
ro 'source'<= ro 'destination' -> DO NOT move the pattern

pattern 3
cluster 'source'2

$$\rho_2 = \frac{n_2}{n_2-1}\|x-m_2\|^2 = \frac{3}{2}\left\|\begin{matrix}0-3\\1-5.\overline{3}\end{matrix}\right\|^2 \cong \frac{3}{2}\cdot 27.\overline{7} \cong 41.\overline{6}$$

$$\rho_1 = \frac{n_1}{n_1+1}\|x-m_1\|^2 = \frac{3}{4}\left\|\begin{matrix}0-3.16\\1-0.67\end{matrix}\right\|^2 = \frac{3}{4}\cdot 10.139 \cong 7.6$$

(same calculation for cluster 3)

ro cluster 'source'(2) 41.6667
ro cluster 'destination' (1, 3) **7.60417** 25.4167
ro 'source'> ro 'destination' -> assign the pattern to the cluster 1

New clusters
C{1} = 1 2 3 9
C{2} = 4 5
C{3} = 6 7 8

Means:
m_1 =(2.3750 0.7500)'
m_2 =(4.5 7.5)'
m_3 =(5.6667 2.3333)' (unchanged)

pattern 4
cluster 'source' 2
ro cluster 'source' 1.0
ro cluster 'destination' 47.5625 24.4167
ro 'source'<= ro 'destination' -> DO NOT move the pattern

pattern 5
cluster 'source' 2
ro cluster 'source' 1.0
ro cluster 'destination' 33.3625 18.4167
ro 'source'<= ro 'destination' -> DO NOT move the pattern

pattern 6
cluster 'source' 3
ro cluster 'source'44.2083
ro cluster 'destination' (1,2) 38.5625 **0.166667**

ro 'source'> ro 'destination' -> assign the pattern to the cluster 2

New clusters
C{1} = 1 2 3 9
C{2} = 4 5 6
C{3} = 7 8

New means
m_1 =(2.3750 0.7500)' (unchanged)
m_2 =(4. 3333 7.5)'
m_3 =(6.5 -0.25)'

pattern 7
cluster 'source'3
ro cluster 'source'1.625
ro cluster 'destination' 12.9625 56.2708
ro 'source'<= ro 'destination' -> DO NOT move the pattern

pattern 8
cluster 'source'3
ro cluster 'source'1.625
ro cluster 'destination' 17.1625 42.0833
ro 'source'<= ro 'destination' -> DO NOT move the pattern

pattern 9
cluster 'source'1
ro cluster 'source'35.1042
ro cluster 'destination' (2,3) 39.2083 **1.70833**

ro 'source'> ro 'destination' -> assign the pattern to the cluster 3

New clusters
C{1} = 1 2 3
C{2} = 4 5 6
C{3} = 7 8 9

new means
m_1 =(0.6667 0.6667)'
m_2 =(4. 3333 7.5)' (unchanged)
m_3 =(6.8333 0.1667)'

pattern 1
cluster 'source'1
ro cluster 'source'0.33333
ro cluster 'destination' 40.0208 26.0417
ro 'source'<= ro 'destination' -> DO NOT move the pattern
pattern 2
cluster 'source'1
ro cluster 'source'0.83333
ro cluster 'destination' 50.5208 25.5417
ro 'source'<= ro 'destination' -> DO NOT move the pattern
pattern 3
cluster 'source'1
ro cluster 'source'0.83333
ro cluster 'destination' 45.7708 35.5417
ro 'source'<= ro 'destination' -> DO NOT move the pattern
pattern 4
cluster 'source'2
ro cluster 'source'1.0417
ro cluster 'destination' 54.4167 48.5417
ro 'source'<= ro 'destination' -> DO NOT move the pattern
pattern 5
cluster 'source'2
ro cluster 'source'0.54167
ro cluster 'destination' 38.4167 41.0417
ro 'source'<= ro 'destination' -> DO NOT move the pattern
pattern 6
cluster 'source'2
ro cluster 'source'0.16667
ro cluster 'destination' 43.3542 46.3542
ro 'source'<= ro 'destination' -> DO NOT move the pattern
pattern 7
cluster 'source'3
ro cluster 'source'3.0833
ro cluster 'destination' 23.4167 56.2708
ro 'source'<= ro 'destination' -> DO NOT move the pattern
pattern 8
cluster 'source'3
ro cluster 'source'0.20833

ro cluster 'destination' 30.1042   42.0833
 ro 'source'<= ro 'destination' -> DO NOT move the pattern

pattern 9

cluster 'source'3
ro cluster 'source'1.7083
ro cluster 'destination' 35.1042   39.2083
 ro 'source'<= ro 'destination' -> DO NOT move the pattern