Chapter 8

Unsupervised machine learning

Professor Jung Jin Lee
Soongsil University, Korea
New Uzbekistan University, Uzbekistan

Chapter 8 Unsupervised machine learning

8.1 Basic concepts of unsupervised learning and clustering

8.2 Hierarchical clustering model

8.3 K-means clustering model

8.1 Basic concept of unsupervised machine learning and clustering

- Unsupervised learning classifies data whose group affiliation is unknown into homogeneous groups.
  => Clustering analysis

- Clustering analysis;
  - understanding the data structure,
  - identifying cluster characteristics
  - determining the relationships between clusters,
  - allowing for the performance of other analyses.

8.1 Basic concept of unsupervised machine learning and clustering

- Clustering analysis is based on the similarity and relationship between data
- The data in one cluster are similar, and the data in other clusters are different.
- The definition of clusters is not easy.
- Not clear how many clusters to divide into.

8.1 Basic concept of unsupervised machine learning and clustering

- Clustering models are divided into hierarchical and partitional clustering models.
- The hierarchical clustering model allows subclusters within a cluster
  - All data is put into one cluster, divided into subclusters, and then divided into subclusters again.
- The partitional clustering model divides the entire data without overlapping each other,
  - K-means clustering model.

8.1 Basic concept of unsupervised machine learning and clustering

- Clustering models are also divided into exclusive clustering, where one data belongs to one cluster, and inclusive clustering, where one data belongs to multiple clusters.
- The K-means clustering model is an exclusive clustering analysis, and the fuzzy clustering model and the mixed distribution clustering are inclusive clustering.
- Clustering models are also classified into prototype-based models, density-based models, and graph-based models.

8.1 Basic concept of unsupervised machine learning and clustering

- Factors for evaluating clustering models;
  - Clustering tendency for a specific data set
  - Number of accurate clusters
  - Comparison of characteristics of formed clusters

- Evaluation measures for clustering;

$$\text{Cohesion}(G_i) = \sum_{\boldsymbol{x},\, \boldsymbol{y} \in G_i} d(\boldsymbol{x}, \boldsymbol{y})$$

$$\text{Separation}(G_i, G_j) = \sum_{\boldsymbol{x} \in G_i} \sum_{\boldsymbol{x} \in G_j} d(\boldsymbol{x}, \boldsymbol{y})$$

$$\text{Cohesion of total model} = \sum_{i=1}^{K} w_i \times (\text{Cohesion of cluster } G_i)$$

8.2 Hierarchical clustering model

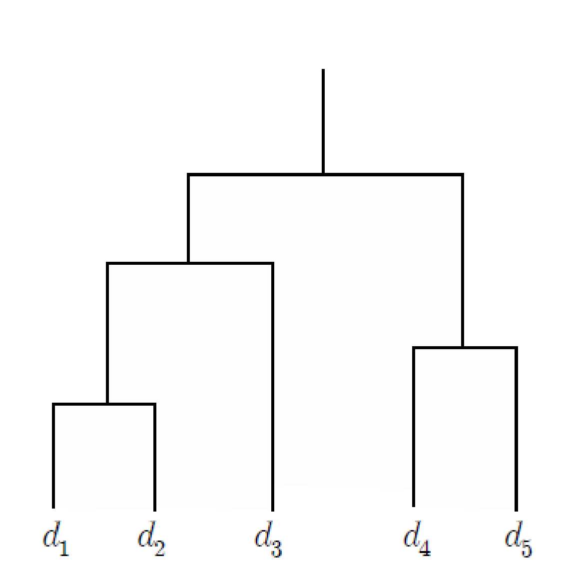
- Agglomerative hierarchical clustering;
  - Starts from one data, and groups the closest clusters in order.
  - Several variations depending on how the distance between clusters is defined.
- Divisive hierarchical clustering;
  - Considers all data as one cluster and divides them in order so that the final cluster becomes one data.
  - Several variations depending on which cluster is first divided and how it is divided. <p>

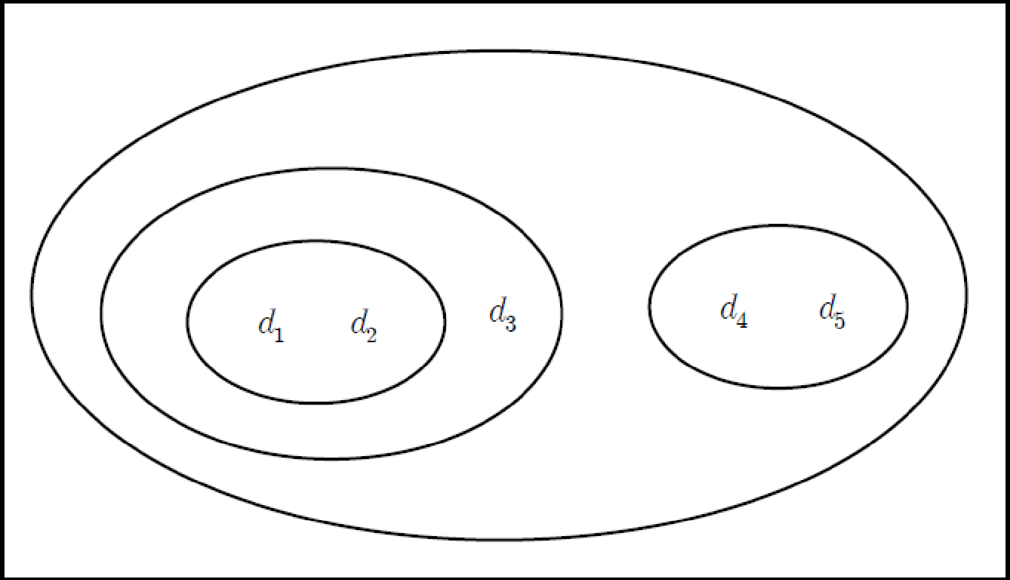
8.2 Hierarchical clustering model

- Resulting graphs of hierarchical clustering;

  * Dendrogram

  *Subset plot





8.2 Hierarchical clustering model

- Agglomerative hierarchical clustering algorithm

Step 1	Consider each data as one cluster and calculate the similarity matrix of all data.
Step 2	repeat
Step 3	Group the two closest clusters into one cluster.
Step 4	Obtain the similarity matrix between all clusters including the newly formed cluster.
Step 5	until (the number of clusters becomes one)

8.2 Hierarchical clustering model

❖ Single linkage or shortest distance

- If the data with the closest distance in the distance matrix D = $\{d_{ij}\}$ U and V, the two data are first grouped to form a cluster (UV).
- The next step calculates the distance between the cluster (UV) and the remaining (n-2) other data or clusters.
- The single linkage distance between the cluster (UV) and cluster W;

$$d_{UV} = \min(d_{UW}, d_{VW})$$

8.2 Hierarchical clustering model

❖ Single linkage or shortest distance

[Example 8.2.1] The five observed data for two variables x_1 and x_2 and the matrix of squared Euclid distances between these data are as follows. Create a hierarchical cluster using the single linkage method.

Table 8.2.1 Five observed data and the matrix of squared Euclid distances

		Distance/th>				
Data	(x_1, x_2)	A	B	C	D	E
A	(1, 5)	0				
B	(2, 4)	2	0			
C	(4, 6)	10	8	0		
D	(4, 3)	13	5	9	0	
E	(5, 3)	20	10	10	1	0

8.2 Hierarchical clustering model

<Answer of Example 8.2.1>

$$d((DE),A) = min(d(D,A), d(E,A)) = min(13,20) = 13$$
$$d((DE),B) = min(d(D,B), d(E,B)) = min(5,10) = 5$$
$$d((DE),C) = min(d(D,C), d(E,C)) = min(9,10) = 9$$

Table 8.2.2 Modified distance matrix with cluster (DE) using the single linkage

	Distance/th>			
Cluster	A	B	C	(DE)
A	0			
B	2	0		
C	10	8	0	
(DE)	13	5	9	0

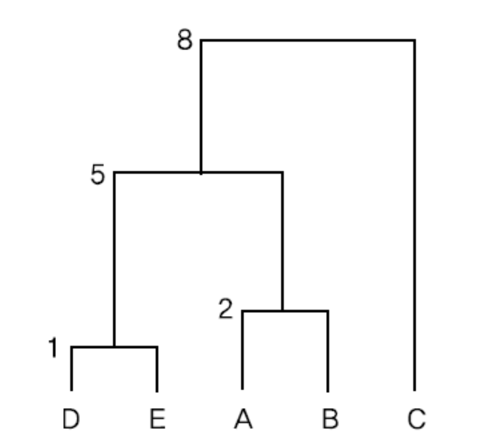
8.2 Hierarchical clustering model

<Answer of Example 8.2.1>

$$d((AB),C) = min(d(A,C),d(B,C)) = min(10,8) = 8$$
$$d((AB),(DE)) = min(d(A,(DE)),d(B,(DE)) = min(13,5) = 5$$

Table 8.2.3 Modified distance matrix with cluster (AB) using the single linkage

	Distance/th>		
Cluster	(AB)	C	(DE)
(AB)	0		
C	8	0	
(DE)	5	9	0



$$d((AB)(DE),C) = min(d((AB),C),d((DE),C)) = min(8,9) = 8$$

8.2 Hierarchical clustering model

❖ Complete linkage or maximum distance

- If the data with the closest distance in the distance matrix D = $\{d_{ij}\}$ U and V, the two data are first grouped to form a cluster (UV).
- The next step calculates the distance between the cluster (UV) and the remaining (n-2) other data or clusters.
- The single linkage distance between the cluster (UV) and cluster W;

$$d_{UV} = \max(d_{UW}, d_{VW})$$

8.2 Hierarchical clustering model

❖ Complete linkage or maximum distance

[Example 8.2.1] The five observed data for two variables $x_1$and x_2 and the matrix of squared Euclid distances between these data are as follows. Create a hierarchical cluster using the complete linkage method.

Table 8.2.1 Five observed data and the matrix of squared Euclid distances

		Distance/th>				
Data	(x_1, x_2)	A	B	C	D	E
A	(1, 5)	0				
B	(2, 4)	2	0			
C	(4, 6)	10	8	0		
D	(4, 3)	13	5	9	0	
E	(5, 3)	20	10	10	1	0

8.2 Hierarchical clustering model

<Answer of Example 8.2.2>

$$d((DE),A) = max(d(D,A), d(E,A)) = max(13,20) = 20$$
$$d((DE),B) = max(d(D,B), d(E,B)) = max(5,10) = 10$$
$$d((DE),C) = max(d(D,C), d(E,C)) = max(9,10) = 10$$

Table 8.2.5 Modified distance matrix with cluster (DE) using the complete linkage

	Distance/th>			
Cluster	A	B	C	(DE)
A	0			
B	2	0		
C	10	8	0	
(DE)	20	10	10	0

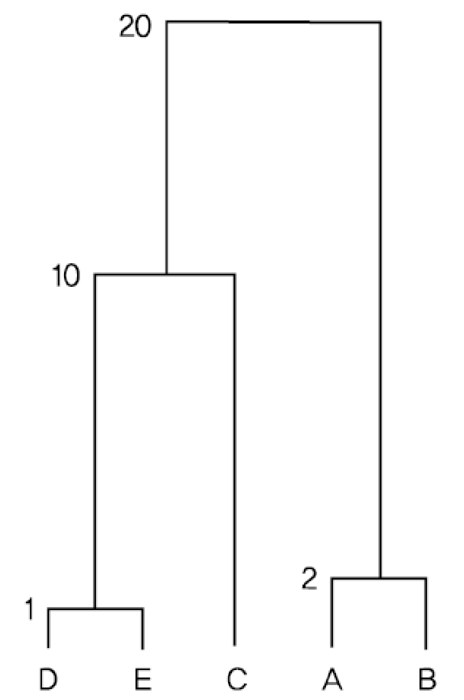
8.2 Hierarchical clustering model

<Answer of Example 8.2.2>

$$d((AB), C) = max(d(A, C), d(B, C)) = max(10, 8) = 10$$
$$d((AB), (DE)) = max(d(A, (DE)), d(B, (DE)) = max(20, 10) = 20$$

Table 8.2.6 Modified distance matrix with cluster (AB) using the complete linkage

	Distance		
Cluster	(AB)	C	(DE)
(AB)	0		
C	10	0	
(DE)	20	10	0



$$d((AB), C(DE)) = max(d((AB), C), d((AB), (DE))) = max(10, 20) = 20$$

8.2 Hierarchical clustering model

❖ Average linkage

- If the data with the closest distance in the distance matrix D = $\{d_{ij}\}$ U and V, the two data are first grouped to form a cluster (UV).
- The next step calculates the average distance between cluster (UV) and cluster W.

$$d_{(UV)W} = \frac{\sum_{\boldsymbol{x}_i \in (UV)} \sum_{\boldsymbol{x}_j \in W} d(\boldsymbol{x}_i, \boldsymbol{x}_j)}{n_{(UV)} \times n_W}$$

8.2 Hierarchical clustering model

❖ Average linkage

[Example 8.2.3] The five observed data for two variables x_1 and x_2 and the matrix of squared Euclid distances between these data are as follows. Create a hierarchical cluster using the average linkage method.

Table 8.2.1 Five observed data and the matrix of squared Euclid distances

		Distance/th>				
Data	(x_1, x_2)	A	B	C	D	E
A	(1, 5)	0				
B	(2, 4)	2	0			
C	(4, 6)	10	8	0		
D	(4, 3)	13	5	9	0	
E	(5, 3)	20	10	10	1	0

8.2 Hierarchical clustering model

<Answer of Example 8.2.3>

$$d((DE),A) = \frac{d(D,A)+d(E,A)}{2\times 1} = \frac{13+20}{2} = 16.5$$
$$d((DE),B) = \frac{d(D,B)+d(E,B)}{2\times 1} = \frac{5+10}{2} = 7.5$$
$$d((DE),C) = \frac{d(D,C)+d(E,C)}{2\times 1} = \frac{9+10}{2} = 9.5$$

Table 8.2.8 Modified distance matrix with cluster (DE) using the single linkage

	Distance/th>			
Cluster	A	B	C	(DE)
A	0			
B	2	0		
C	10	8	0	
(DE)	16.5	7.5	9.5	0

8.2 Hierarchical clustering model

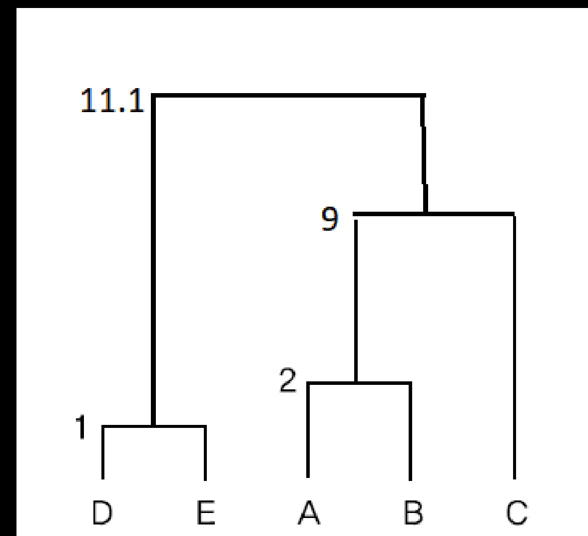
<Answer of Example 8.2.3>

$$d((AB),C) = \frac{d(A,C)+d(B,C)}{2\times1} = \frac{10+8}{2} = 9$$

$$d((AB),(DE)) = \frac{d(A,D)+d(A,E)+d(B,D)+d(B,E)}{2\times2} = \frac{13+20+5+10}{4} = 12$$

Table 8.2.9 Modified distance matrix with cluster (AB) using the average linkage

	Distance/th>		
Cluster	(AB)	C	(DE)
(AB)	0		
C	9	0	
(DE)	12	9.5	0



$$d((AB)C,(DE)) = \frac{d(A,D)+d(A,E)+d(B,D)+d(B,E)+d(C,D)+d(C,E)}{3\times2} = \frac{12+20+5+10+9+10}{6} = 11.1$$

8.2 Hierarchical clustering model

❖ Centroid linkage

- If the data with the closest distance in the distance matrix D = $\{d_{ij}\}$ U and V, the two data are first grouped to form a cluster (UV).
- The distance between two clusters is defined as the squared Euclid distance between the two centroids

$$d(Gi, G_j) = ||\boldsymbol{c}_i - \boldsymbol{c}_j||^2$$

- The center of the new cluster is the weighted average;

$$\boldsymbol{c} = \frac{n_i \boldsymbol{c}_i + n_j \boldsymbol{c}_j}{n_i + n_j}$$

8.2 Hierarchical clustering model

❖ Centroid linkage

[Example 8.2.4] The five observed data for two variables x_1 and x_2 and the matrix of squared Euclid distances between these data are as follows. Create a hierarchical cluster using the centroid linkage method.

Table 8.2.1 Five observed data and the matrix of squared Euclid distances

		Distance/th>				
Data	(x_1, x_2)	A	B	C	D	E
A	(1, 5)	0				
B	(2, 4)	2	0			
C	(4, 6)	10	8	0		
D	(4, 3)	13	5	9	0	
E	(5, 3)	20	10	10	1	0

8.2 Hierarchical clustering model

<Answer of Example 8.2.4>

$$d((DE),A) = (4.5-1)^2 + (3-5)^2 = 16.25$$
$$d((DE),B) = (4.5-2)^2 + (3-4)^2 = 7.25$$
$$d((DE),C) = (4.5-4)^2 + (3-6)^2 = 9.25$$

Table 8.2.11 Modified distance matrix with cluster (DE) using the centroid linkage

	Distance/th>			
Cluster	A	B	C	(DE)
A	0			
B	2	0		
C	10	8	0	
(DE)	16.25	7.25	9.25	0

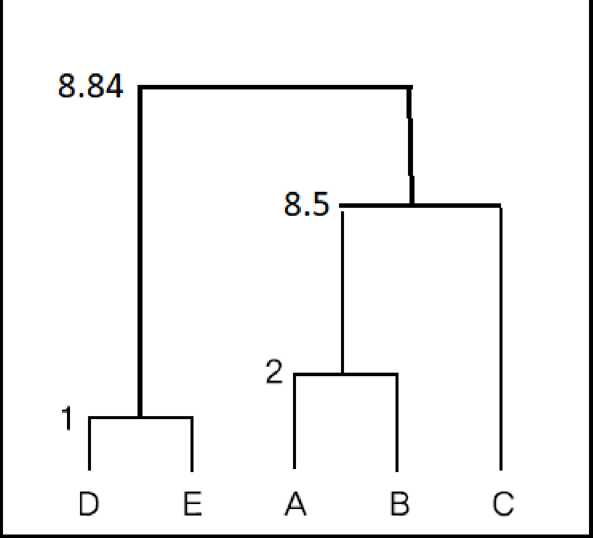
8.2 Hierarchical clustering model

<Answer of Example 8.2.4>

$$d((AB),C) = (1.5-4)^2 + (4.5-6)^2 = 8.5$$
$$d((AB),(DE)) = (1.5-4.5)^2 + (4.5-3)^2 = 11.25$$

Table 8.2.12 Modified distance matrix with cluster (AB) using the centroid linkage

	Distance/th>		
Cluster	(AB)	C	(DE)
(AB)	0		
C	8.5	0	
(DE)	11.25	9.5	0



$(AB)C$ becomes the next cluster and the center of the cluster $\frac{2\times(1.5,\ 4.5)+1\times(4,\ 6)}{2+1}$ = (2.3, 5)

$$d((AB)C,(DE)) = (2.3-4.5)^2 + (5-3)^2 = 8.84$$

8.2 Hierarchical clustering model

❖ Ward linkage

- Ward linkage measures the information loss caused by grouping data into a single cluster at each stage by the error sum of squares (ESS).

$$ESS_i = \sum_{j=1}^{n_i} \sum_{k=1}^{m} (x_{ijk} - \bar{x}_{ik})^2$$

$$ESS = \sum_{i=1}^{K} ESS_i = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{k=1}^{m} (x_{ijk} - \bar{x}_{ik})^2$$

- Clusters are merged to create a new cluster so that the increment of ESS due to the merging is minimized.

$$d(G_i, G_j) = \frac{||\mathbf{c}_i - \mathbf{c}_j||^2}{\frac{1}{n_i} + \frac{1}{n_j}}$$

8.2 Hierarchical clustering model

❖ Ward linkage

[Example 8.2.5] The five observed data for two variables $x_1$and x_2 and the matrix of squared Euclid distances between these data are as follows. Create a hierarchical cluster using the Ward linkage method.

Table 8.2.1 Five observed data and the matrix of squared Euclid distances

		Distance/th>				
Data	(x_1, x_2)	A	B	C	D	E
A	(1, 5)	0				
B	(2, 4)	2	0			
C	(4, 6)	10	8	0		
D	(4, 3)	13	5	9	0	
E	(5, 3)	20	10	10	1	0

8.2 Hierarchical clustering model

<Answer of Example 8.2.5>

$$d((DE),A) = \frac{(4.5-1)^2+(3-5)^2)}{\frac{1}{2}+\frac{1}{1}} = 11.17$$

$$d((DE),B) = \frac{(4.5-2)^2+(3-4)^2)}{\frac{1}{2}+\frac{1}{1}} = 4.83$$

$$d((DE),C) = \frac{(4.5-4)^2+(3-6)^2)}{\frac{1}{2}+\frac{1}{1}} = 6.17$$

Table 8.2.14 Modified distance matrix with cluster (DE) using the Ward linkage

	Distance/th>			
Cluster	A	B	C	(DE)
A	0			
B	2	0		
C	10	8	0	
(DE)	11.17	4.83	6.17	0

8.2 Hierarchical clustering model

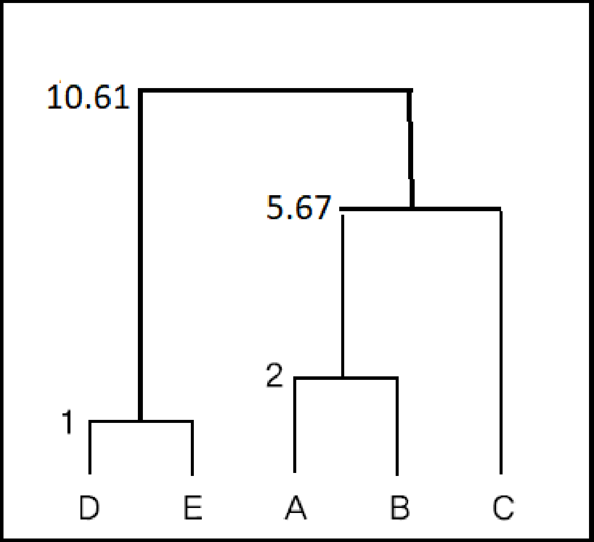
<Answer of Example 8.2.5>

$$d((AB),C) = \frac{(1.5-4)^2+(4.5-6)^2)}{\frac{1}{2}+\frac{1}{1}} = 5.67$$

$$d((AB),(DE)) = \frac{(1.5-4.5)^2+(4.5-3)^2)}{\frac{1}{2}+\frac{1}{1}} = 11.25$$

Table 8.2.15 Modified distance matrix with cluster (AB) using the Ward linkage

	Distance/th>		
Cluster	(AB)	C	(DE)
(AB)	0		
C	5.67	0	
(DE)	11.25	6.17	0



$(AB)C$ becomes the next cluster

$$\frac{2\times(1.5,4.5)+1\times(4,6)}{2+1} = (2.3,\ 5)$$

$$d((AB)C,(DE)) = \frac{(2.3-4.5)^2+(5-3)^2)}{\frac{1}{3}+\frac{1}{2}} = 10.61$$

8.3 K-means clustering model

- K-means clustering is for continuous data, and the mean and median can be used as the centers of the clusters,
- A similar concept can be applied to discrete data by defining the centroid of the discrete data.
- The K-means clustering determines the number of clusters K first and calculates the center of each cluster
- Classifies each data into a cluster with the closest cluster center and recalculates the center of each cluster
- The same method is repeated until there is no change in the cluster center.

8.3 K-means clustering model

K-means clustering algorithm

Step 1	Determine the number of clusters K you want.
Step 2	Select the initial center of each cluster.
Step 3	repeat
Step 4	Classify each data into the cluster with the closest cluster center.
Step 5	Recalculate the center of each cluster.
Step 6	until (there is little change in the cluster center

8.3 K-means clustering model

[Example 8.3.1] For the two variables x_1 and x_2, four data were observed as follows. Find two clusters using the 2-means clustering algorithm with the squared Euclid distance between the data.

Table 8.3.1 Data for the 2-means clustering algorithm

Data	(x_1, x_2)
A	(3, 4)
B	(-1, 2)
C	(-2, -3)
D	(1, -2)

8.3 K-means clustering model

<Answer of Example 8.3.1>

Let the center of cluster 1 be data A=(3,4) and the center of cluster 2 be data C=(-2,-3). The distances from each data to the centers of the two clusters are as follows.

Table 8.3.2 Distance between data and the center of cluster

Data	Cluster 1 Distance to center (3, 4)	Cluster 2 Distance to center (-2, -3)
A	0	74
B	20	26
C	74	0
D	40	10

8.3 K-means clustering model

<Answer of Example 8.3.1>

If each data is classified by the nearest cluster center, data A and B are classified into cluster 1, data C and D are classified into cluster 2, and the center of the new cluster 1 by the average is (1,3), and the center of cluster 2 is (-0.5,-2.5). The distances from each data to the centers of the two new clusters are as follows.

Table 8.3.3 Modified distance between data and the center of cluster

Data	Cluster 1 Distance to center (1, 3)	Cluster 2 Distance to center (-0.5, -2.5)
A	5	54.5
B	5	20.5
C	45	2.5
D	25	2.5

8.3 K-means clustering model

❖ Theoretical background of K-means clustering

- A measure of clustering performance can be defined as the sum of distances from all data to each center.

$$(\text{Performance measure of clustering}) = \sum_{i=1}^{K} \sum_{\boldsymbol{x} \in G_i} d(\boldsymbol{c}_i, \boldsymbol{x})$$

- The performance measure of a clustering is the within sum of squares (WSS) in the case of the Euclidean distance.

$$\text{WSS} = \sum_{i=1}^{K} \sum_{x \in G_i} (c_i - x)^2$$

8.3 K-means clustering model

❖ Theoretical background of K-means clustering

- The solution that minimizes the within sum of squares is as follows.

$$\frac{\partial}{\partial c_i}\mathrm{SSE} = \sum_{x \in G_i} 2(c_i - x) = 0$$

$$c_i = \frac{1}{n_i}\sum_{x \in G_i} x$$

- The solution that minimizes the within sum of squares is the mean of clusters.

Summary

- Basic concept of clustering:
  - Classify data whose group is unknown into homogeneous groups.
  - Cohesion and separation are clustering measures.
- Hierarchical clustering:
  - Agglomerative clustering starts from one data and groups the closest clusters.
  - Single, complete, average, centroid, Ward linkage.
- K-means clustering:
  - Determines the number of clusters K and selects the initial center of each cluster.
  - Each data is classified into a cluster with the closest cluster center.
  - The center of each cluster is recalculated.

Thank you !!!