Hierarchical Clustering

CSCI 347

Adiesha Liyana Ralalage

Hierarchical clustering

- K-means clustering requires us to pre-specify the number of clusters **K** .
- **DBSCAN** requires approximating appropriate values for ϵ and **minpt.**
- Hierarchical clustering is an alternative approach which does not require that we commit to a particular choice of **K** and don't require estimates for parameters.
- The goal of hierarchical clustering is to create a sequence of nested partitions, which can be conveniently visualized via a tree or hierarchy of clusters, also called the cluster dendrogram.

Hierarchical clustering

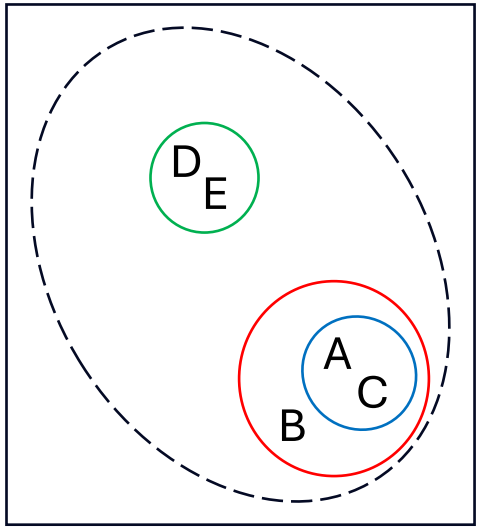
- The clusters in the hierarchy range from the fine-grained to the coarse-grained –the lowest level of the tree (the leaves) consists of each point in its own cluster, whereas the highest level (the root) consists of all points in one cluster.
- At some intermediate level, we may find meaningful clusters.
  - If the user provides the number of clusters k, we can choose the level at which there are k clusters
- In this lecture, we discuss bottom-up or agglomerative clustering. This is the most common type of hierarchical clustering.

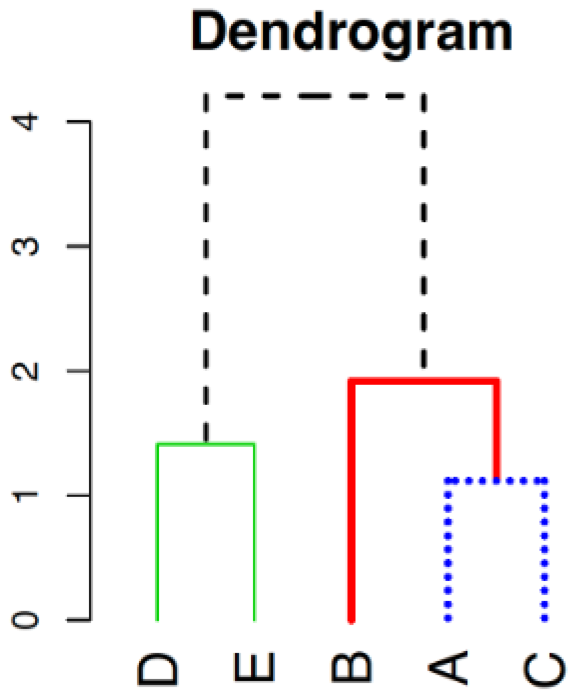
Hierarchical clustering

- There are two main approaches in Hierarchical clustering.
  - Agglomerative clustering
  - Divisive clustering
- Agglomerative strategies work in a bottom-up manner.
  - We start with each n points in a separate cluster and repeatedly merge if they are similar until all points are members of the same cluster.
- Divisive strategy works completely opposite, it starts with all points in the same cluster and then recursively split the clusters until all points are in separate clusters.

Hierarchical clustering algorithm

- Higher level idea:
  - Start each point in its own cluster.
  - Identify closest two clusters and merge them.
  - Repeat.
  - Ends when all points are in a single cluster.



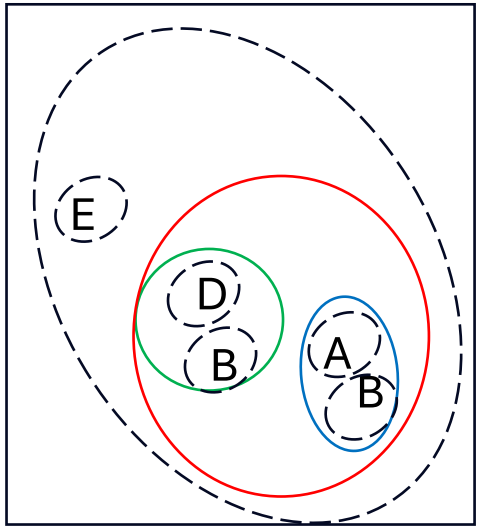


Hierarchical clustering algorithm

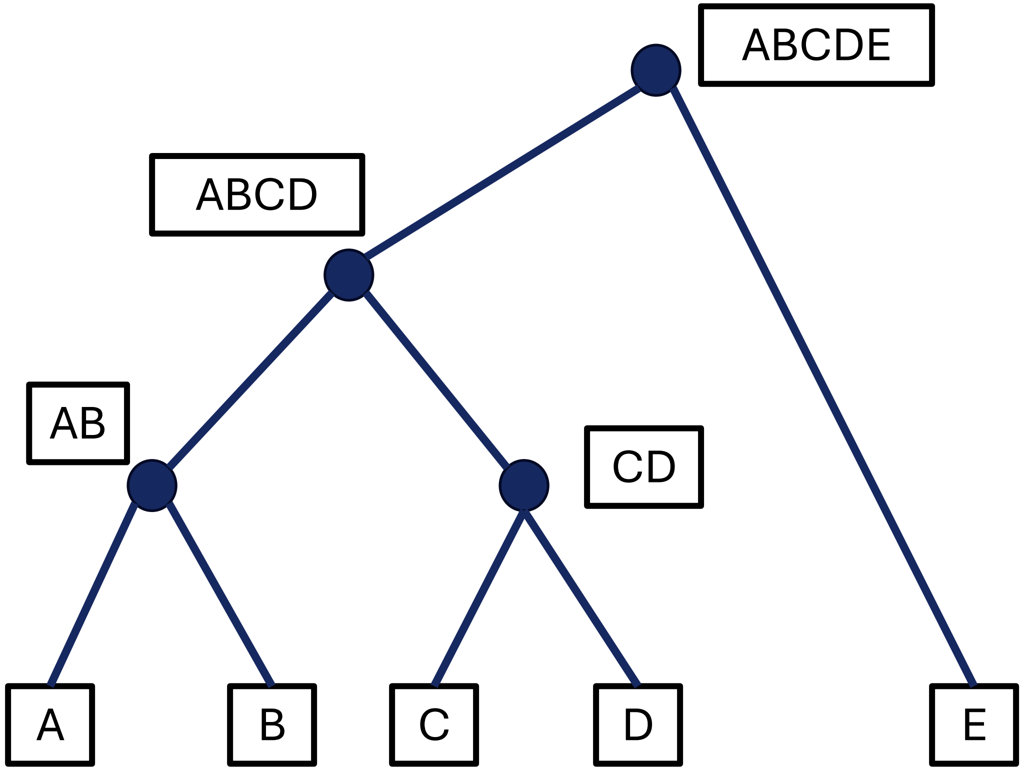
- Some notations we are going to use in this lecture.
- $D = \{x_1, x_2, x_3, \dots, x_n\}$ is the dataset, where $x_i \in \mathbb{R}^d$
- A clustering $C = \{C_1, C_2, C_3, \dots, C_k\}$ is a partition of D.
  - Each cluster is a set of points $C_i \subseteq D$, such that the clusters are pairwise disjoint $C_i \cap C_j = \emptyset \; for\ all\ i \neq j$ and $\cup\, C_i = D$.
- A clustering $\mathcal{A} = \{A_1, A_2, A_3, \dots, A_r\}$ is said to be nested in another clustering $\mathcal{B} = \{B_1, B_2, B_3, \dots, B_s\}$ if and only if $r > s$, and for each cluster $\forall A_i \in \mathcal{A}: \exists B_j \in \mathcal{B}: A_i \subseteq B_j$.
- Hierarchical clustering yields a sequence of n nested partitions $\mathcal{C}_1, \dots, \mathcal{C}_n$ ranging from the trivial clustering $\mathcal{C}_1 = \{\{x_1\}, \dots, \{x_n\}\}$ where each point is a separate cluster, to the trivial clustering $\mathcal{C}_n = \{\{x_1, \dots, x_n\}\}$, where all points are in the same cluster

Hierarchical clustering algorithm

- Cluster dendrogram is basically represents the hierarchy of clusters.



Clustering	Clusters
$\mathcal{C}_1$	$\{A\},\{B\},\{C\},\{D\},\{E\}$
$\mathcal{C}_2$	$\{AB\},\{C\},\{D\},\{E\}$
$\mathcal{C}_3$	$\{AB\},\{CD\},\{E\}$
$\mathcal{C}_4$	$\{ABCD\},\{E\}$
$\mathcal{C}_5$	$\{ABCDE\}$



Agglomerative clustering algorithm

AgglomerativeClustering(D, k)

1. $\mathcal{C} \leftarrow \{C_i = \{x_i\} | x_i \in D\}$
2. $\Delta = \{\delta(x_i, x_j) : x_i, x_j \in D\}$
3. Repeat:
   1. Find the closest pair of clusters $C_i, C_j \in \mathcal{C}$
   2. $C_{i,j} = C_i \cup C_j$
   3. $\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{ij}\}$
   4. Update the distance matrix Δ to reflect new clustering.
4. Until $|\mathcal{C}| = k$

Agglomerative clustering algorithm

- When it comes to computing distances between two clusters, we can employ several strategies.
    - Single Link
    - Complete Link
    - Group average
    - Mean distance

How to calculate the distance between clusters?

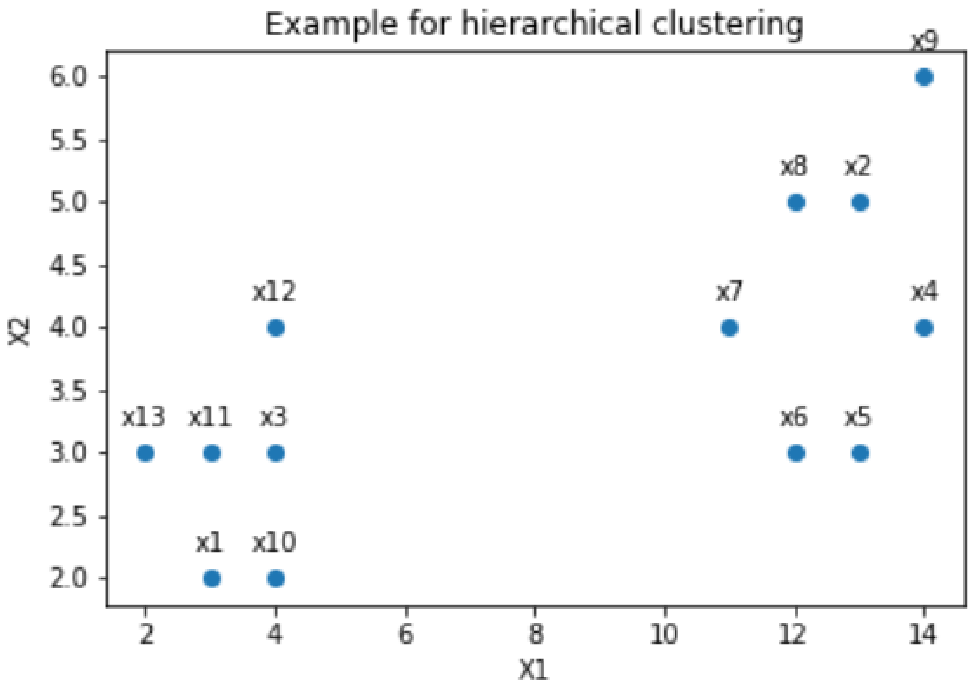
- Single Link:
  - $\delta(C_i, C_j) = \min\{\delta(x, y) \mid x \in C_i, y \in C_j\}$
  - Distance between two clusters is defined as the minimum distance between a point in C_i and a point in C_j
- Complete Link:
  - $\delta(C_i, C_j) = \max\{\delta(x, y) \mid x \in C_i, y \in C_j\}$
  - Distance between two clusters is defined as the maximum distance between a point in C_i and a point in C_j.
- Group average
  - $\delta(C_i, C_j) = \frac{\sum_{x \in C_i} \sum_{y \in C_j} \delta(x,y)}{n_i \cdot n_j}$
  - Distance is defined as the average pairwise distance between points in C_i and C_j

How to calculate the distance between clusters?

- Mean distance:
  - $\delta(C_i, C_j) = \delta(\mu_i, \mu_j)$
  - $\mu_i = \frac{1}{n}\sum_{x \in C_i} x$
  - Distance between two clusters is defined as the distance between the means or centroids of the two clusters
- There are several other strategies as well.

Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3



Agglomerative clustering algorithm

AgglomerativeClustering(D, k)

1. $\mathcal{C} \leftarrow \{C_i = \{x_i\} | x_i \in D\}$
2. $\Delta = \{\delta(x_i, x_j) : x_i, x_j \in D\}$
3. Repeat:
   1. Find the closest pair of clusters $C_i, C_j \in \mathcal{C}$
   2. $C_{i,j} = C_i \cup C_j$
   3. $\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{i,j}\}$
   4. Update the distance matrix Δ to reflect new clustering.
4. Until $|\mathcal{C}| = k$

Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

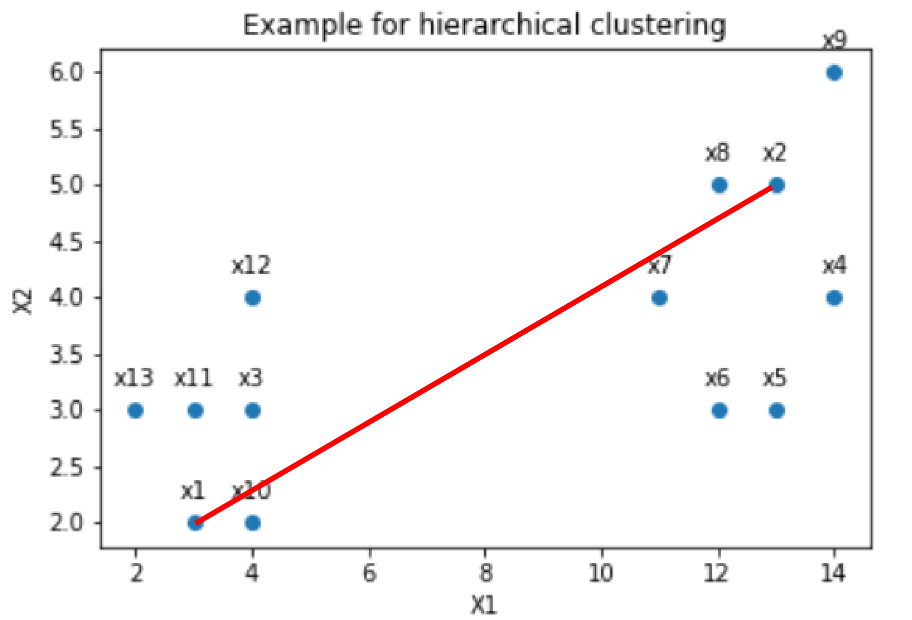
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
x_1	0												
x_2		0											
x_3			0										
x_4				0									
x_5					0								
x_6						0							
x_7							0						
x_8								0					
x_9									0				
x_{10}										0			
x_{11}											0		
x_{12}												0	
x_{13}													0

MONTANA STATE UNIVERSITY

Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

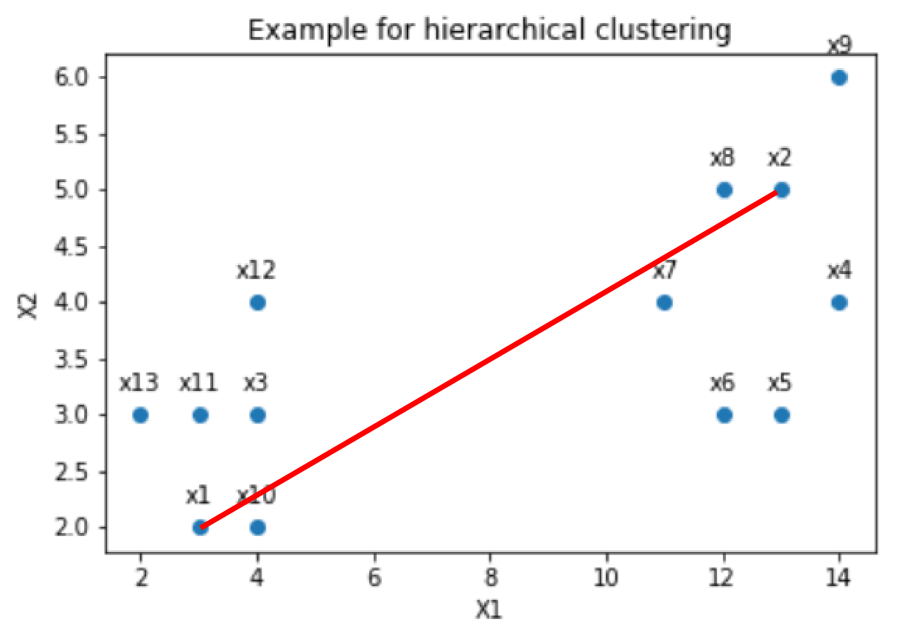
	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
$\boldsymbol{x_1}$	0												
$\boldsymbol{x_2}$		0											
x_3			0										
x_4				0									
x_5													
x_6													
x_7													
x_8													
x_9													
x_{10}										0			
x_{11}											0		
x_{12}												0	
x_{13}													0



Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
$\boldsymbol{x_1}$	0												
$\boldsymbol{x_2}$		0											
x_3			0										
x_4				0									
x_5													
x_6													
x_7													
x_8													
x_9													
x_{10}										0			
x_{11}											0		
x_{12}												0	
x_{13}													0

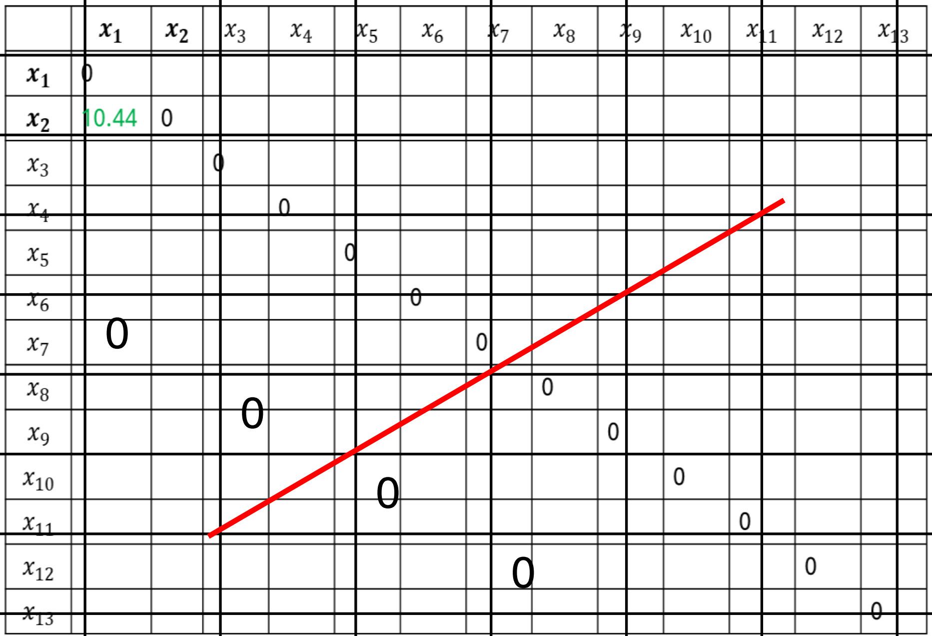


$$\delta(x_1, x_2) = \sqrt{(3-13)^2 + (2-5)^2} = \sqrt{109} = 10.44$$

Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
$\boldsymbol{x_1}$	0												
$\boldsymbol{x_2}$	10.44	0											
x_3			0										
x_4				0									
x_5					0								
x_6						0							
x_7							0						
x_8								0					
x_9									0				
x_{10}										0			
x_{11}											0		
x_{12}												0	
x_{13}													0



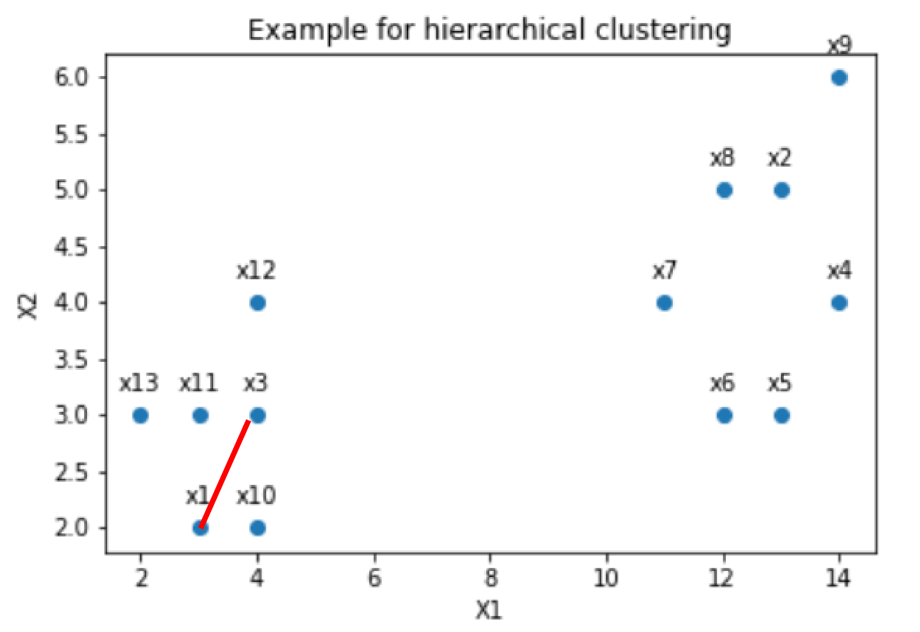
	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
$\boldsymbol{x_1}$	0												
$\boldsymbol{x_2}$	10.44	0											
x_3			0										
x_4				0									
x_5					0								
x_6						0							
x_7							0						
x_8								0					
x_9									0				
x_{10}										0			
x_{11}											0		
x_{12}												0	
x_{13}													0

$$\delta(x_1, x_2) = \sqrt{(3-13)^2 + (2-5)^2} = \sqrt{109} = 10.44$$

Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

	$\boldsymbol{x_1}$	x_2	$\boldsymbol{x_3}$	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
$\boldsymbol{x_1}$	0												
x_2	10.44	0											
$\boldsymbol{x_3}$	1.41		0										
x_4				0									
x_5					0								
x_6													
x_7													
x_8													
x_9													
x_{10}										0			
x_{11}											0		
x_{12}												0	
x_{13}													0



$$\delta(x_1, x_3) = \sqrt{(3-4)^2 + (2-3)^2} = \sqrt{2} = 1.41$$

MONTANA STATE UNIVERSITY

Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
x_1	0												
x_2	10.44	0											
x_3	1.41	9.21	0										
x_4	11.18	1.41	10.05	0									
x_5	10.05	2	9	1.41	0								
x_6	9.06	2.24	8	2.24	1	0							
x_7	8.25	2.24	7.07	3	2.24	1.41	0						
x_8	9.49	1	8.25	2.23	2.24	2	1.41	0					
x_9	11.70	1.41	10.44	2	3.16	3.61	3.61	2.24	0				
x_{10}	1	9.49	1	10.20	9.06	8.06	7.28	8.54	10.77	0			
x_{11}	1	10.20	1	11.05	10	9	8.06	7.28	8.54	10.77	0		
x_{12}	2.24	9.06	1	10	9.06	8.06	7	8.06	10.20	2	1.41	0	
x_{13}	1.41	11.18	2	12.04	11	10	9.06	10.20	12.37	2.24	1	2.24	0

Agglomerative clustering algorithm

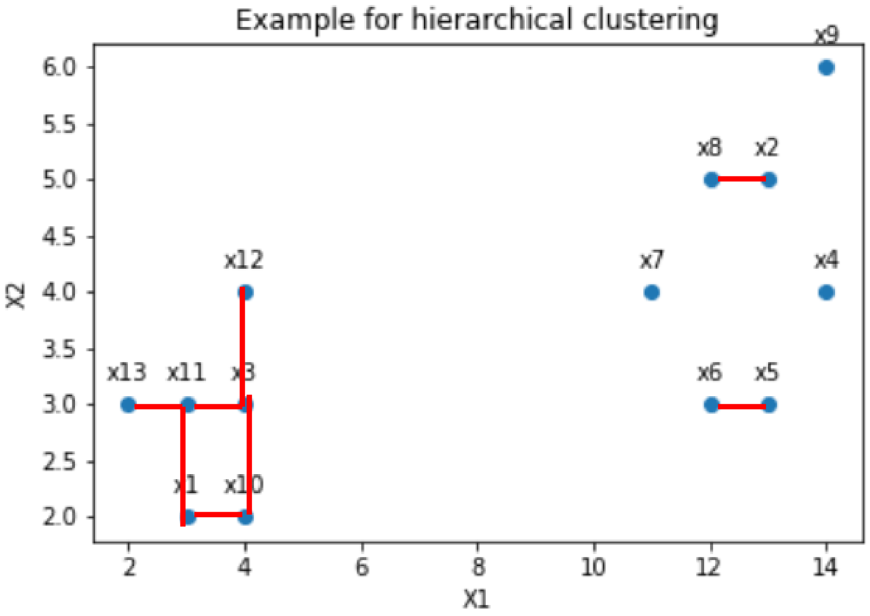
AgglomerativeClustering(D, k)

1. $\mathcal{C} \leftarrow \{C_i = \{x_i\} | x_i \in D\}$
2. $\Delta = \{\delta(x_i, x_j): x_i, x_j \in D\}$
3. Repeat:
    1. Find the closest pair of clusters $C_i, C_j \in \mathcal{C}$
    2. $C_{i,j} = C_i \cup C_j$
    3. $\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{i,j}\}$
    4. Update the distance matrix Δ to reflect new clustering.
4. Until $|\mathcal{C}| = k$

Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

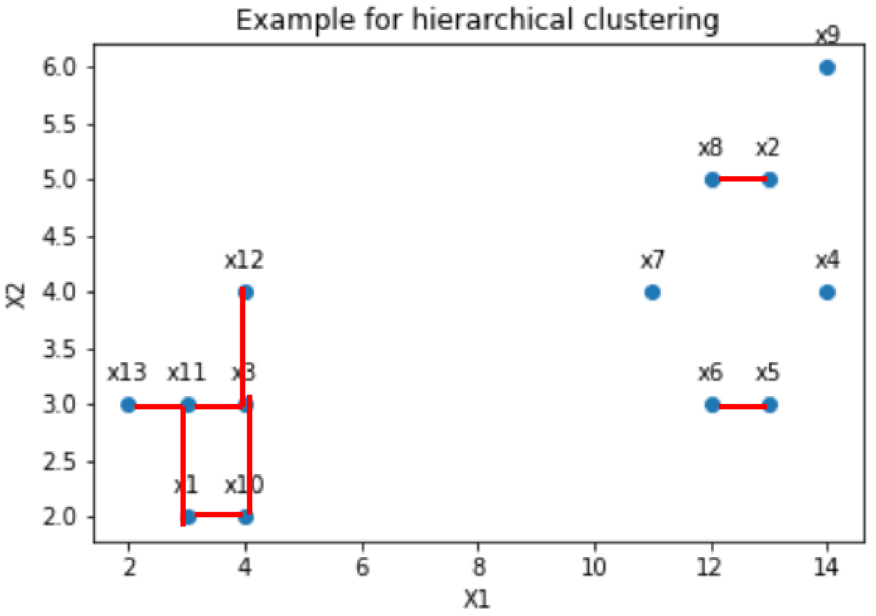
	x_1	x_2	x_3	x_4	x_5	x_6							x_{13}
x_1	0												
x_2	10.44	0											
x_3	1.41	9.21	0										
x_4	11.18	1.41	10.05	0									
x_5	10.05	2	9	1.41	0								
x_6	9.06	2.24	8	2.24	1	0							
x_7	8.25	2.24	7.07	3	2.24	1.41	0						
x_8	9.49	1	8.25	2.23	2.24	2	1.41	0					
x_9	11.70	1.41	10.44	2	3.16	3.61	3.61	2.24	0				
x_{10}	1	9.49	1	10.20	9.06	8.06	7.28	8.54	10.77	0			
x_{11}	1	10.20	1	11.05	10	9	8.06	7.28	8.54	10.77	0		
x_{12}	2.24	9.06	1	10	9.06	8.06	7	8.06	10.20	2	1.41	0	
x_{13}	1.41	11.18	2	12.04	11	10	9.06	10.20	12.37	2.24	1	2.24	0



Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

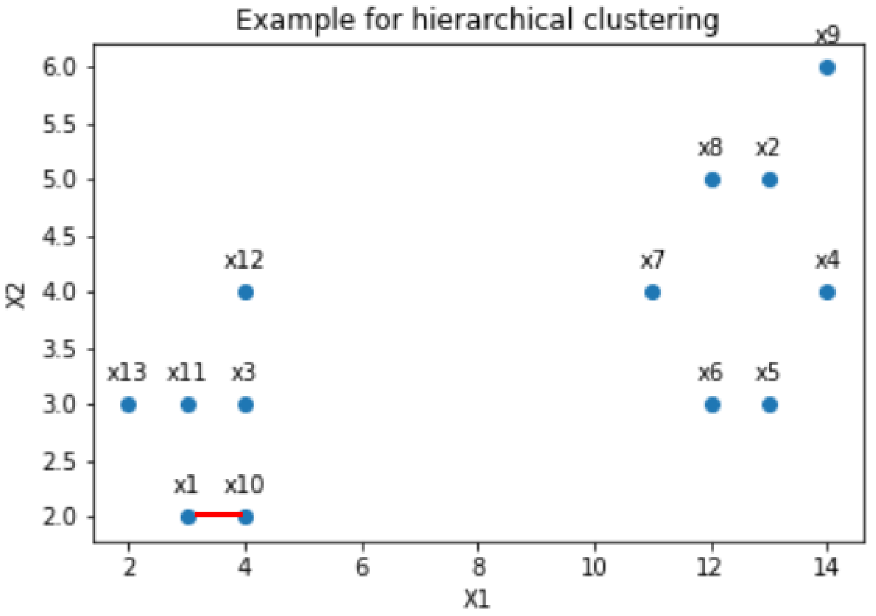
	x_1	x_2	x_3	x_4	x_5	x_6							x_{13}
x_1	0												
x_2	10.44	0											
x_3	1.41	9.21	0										
x_4	11.18	1.41	10.05	0									
x_5	10.05	2	9	1.41	0								
x_6	9.06	2.24	8	2.24	1	0							
x_7	8.25	2.24	7.07	3	2.24	1.41	0						
x_8	9.49	1	8.25	2.23	2.24	2	1.41	0					
x_9	11.70	1.41	10.44	2	3.16	3.61	3.61	2.24	0				
x_{10}	1	9.49	1	10.20	9.06	8.06	7.28	8.54	10.77	0			
x_{11}	1	10.20	1	11.05	10	9	8.06	7.28	8.54	10.77	0		
x_{12}	2.24	9.06	1	10	9.06	8.06	7	8.06	10.20	2	1.41	0	
x_{13}	1.41	11.18	2	12.04	11	10	9.06	10.20	12.37	2.24	1	2.24	0



Example

	X_1	X_2
x_1	3	2
x_2	13	5
x_3	4	3
x_4	14	4
x_5	13	3
x_6	12	3
x_7	11	4
x_8	12	5
x_9	14	6
x_{10}	4	2
x_{11}	3	3
x_{12}	4	4
x_{13}	2	3

	x_1	x_2	x_3	x_4	x_5	x_6							x_{13}
x_1	0												
x_2	10.44	0											
x_3	1.41	9.21	0										
x_4	11.18	1.41	10.05	0									
x_5	10.05	2	9	1.41	0								
x_6	9.06	2.24	8	2.24	1	0							
x_7	8.25	2.24	7.07	3	2.24	1.41	0						
x_8	9.49	1	8.25	2.23	2.24	2	1.41	0					
x_9	11.70	1.41	10.44	2	3.16	3.61	3.61	2.24	0				
x_{10}	1	9.49	1	10.20	9.06	8.06	7.28	8.54	10.77	0			
x_{11}	1	10.20	1	11.05	10	9	8.06	7.28	8.54	10.77	0		
x_{12}	2.24	9.06	1	10	9.06	8.06	7	8.06	10.20	2	1.41	0	
x_{13}	1.41	11.18	2	12.04	11	10	9.06	10.20	12.37	2.24	1	2.24	0

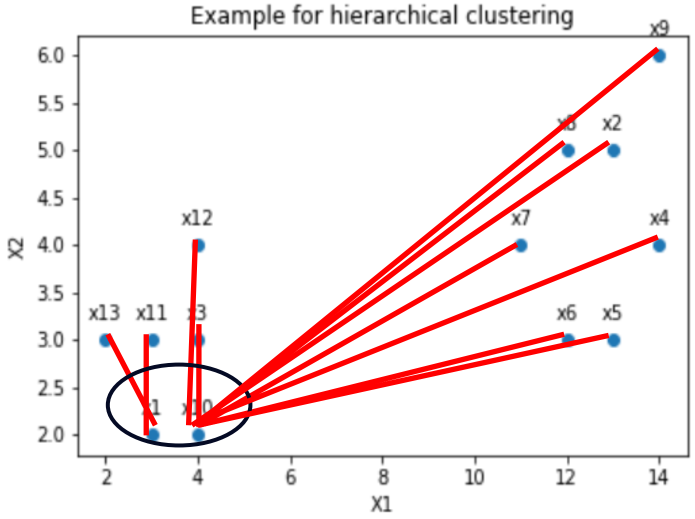


Agglomerative clustering algorithm

AgglomerativeClustering(D, k)

1. $\mathcal{C} \leftarrow \{C_i = \{x_i\} | x_i \in D\}$
2. $\Delta = \{\delta(x_i, x_j) : x_i, x_j \in D\}$
3. Repeat:
   1. Find the closest pair of clusters $C_i, C_j \in \mathcal{C}$
   2. $C_{ij} = C_i \cup C_j$
   3. $\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{i,j}\}$
   4. Update the distance matrix Δ to reflect new clustering.
4. Until $|\mathcal{C}| = k$

Example



$C_{1,10}$

	$\{x_1, x_{10}\}$	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{11}	x_{12}	x_{13}
$\{x_1, x_{10}\}$	0											
x_2	9.49	0										
x_3	1	9.21	0									
x_4	10.20	1.41	10.05	0								
x_5	9.06	2	9	1.41	0							
x_6	3.61	2.24	8	2.24	1	0						
x_7	7.28	2.24	7.07	3	2.24	1.41	0					
x_8	8.54	1	8.25	2.23	2.24	2	1.41	0				
x_9	10.77	1.41	10.44	2	3.16	3.61	3.61	2.24	0			
x_{11}	1	10.20	1	11.05	10	9	8.06	9.22	11.40			
x_{12}	2	9.06	1	10	9.06	8.06	7	8.06	10.20	1.41	0	
x_{13}	1.41	11.18	2	12.04	11	10	9.06	10.20	12.37	1	2.24	0

1. $C_{i,j} = C_i \cup C_j$
2. $\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{i,j}\}$
3. Update the distance matrix Δ to reflect new clustering.

We pick the single linkage strategy to compute the distance between two clusters.

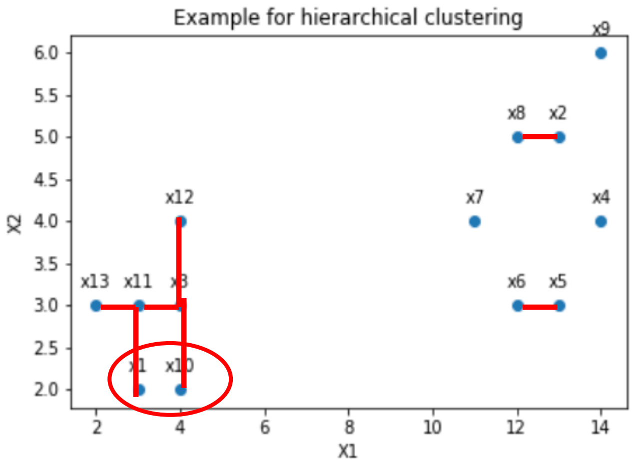
Agglomerative clustering algorithm

AgglomerativeClustering(D, k)

1. $\mathcal{C} \leftarrow \{C_i = \{x_i\} | x_i \in D\}$
2. $\Delta = \{\delta(x_i, x_j): x_i, x_j \in D\}$
3. Repeat:
    1. Find the closest pair of clusters $C_i, C_j \in \mathcal{C}$
    2. $C_{i,j} = C_i \cup C_j$
    3. $\mathcal{C} \leftarrow (\mathcal{C} \setminus \{C_i, C_j\}) \cup \{C_{i,j}\}$
    4. Update the distance matrix Δ to reflect new clustering.
4. Until $|\mathcal{C}| = k$

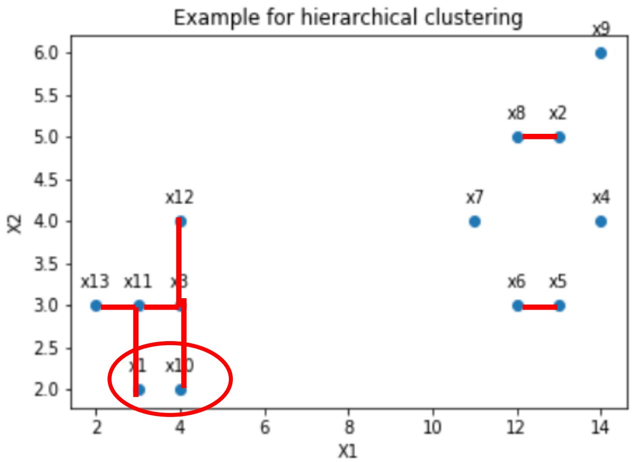
Example

	$\{x_1, x_{10}\}$	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{11}	x_{12}	x_{13}
$\{x_1, x_{10}\}$	0											
x_2	9.49	0										
x_3	1	9.21	0									
x_4	10.20	1.41	10.05	0								
x_5	9.06	2	9	1.41	0							
x_6	3.61	2.24	8	2.24	1	0						
x_7	7.28	2.24	7.07	3	2.24	1.41	0					
x_8	8.54	1	8.25	2.23	2.24	2	1.41	0				
x_9	10.77	1.41	10.44	2	3.16	3.61	3.61	2.24	0			
x_{11}	1	10.20	1	11.05	10	9	8.06	9.22	11.40			
x_{12}	2	9.06	1	10	9.06	8.06	7	8.06	10.20	1.41	0	
x_{13}	2	11.18	2	12.04	11	10	9.06	10.20	12.37	1	2.24	0



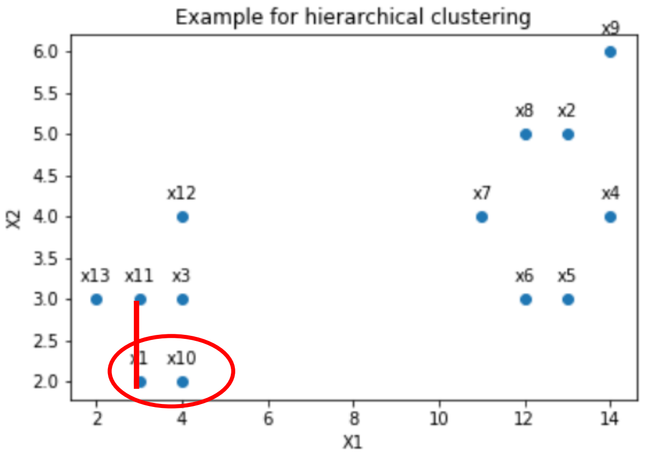
Example

	$\{x_1, x_{10}\}$	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{11}	x_{12}	x_{13}
$\{x_1, x_{10}\}$	0											
x_2	9.49	0										
x_3	1	9.21	0									
x_4	10.20	1.41	10.05	0								
x_5	9.06	2	9	1.41	0							
x_6	3.61	2.24	8	2.24	1	0						
x_7	7.28	2.24	7.07	3	2.24	1.41	0					
x_8	8.54	1	8.25	2.23	2.24	2	1.41	0				
x_9	10.77	1.41	10.44	2	3.16	3.61	3.61	2.24	0			
x_{11}	1	10.20	1	11.05	10	9	8.06	9.22	11.40			
x_{12}	2	9.06	1	10	9.06	8.06	7	8.06	10.20	1.41	0	
x_{13}	2	11.18	2	12.04	11	10	9.06	10.20	12.37	1	2.24	0



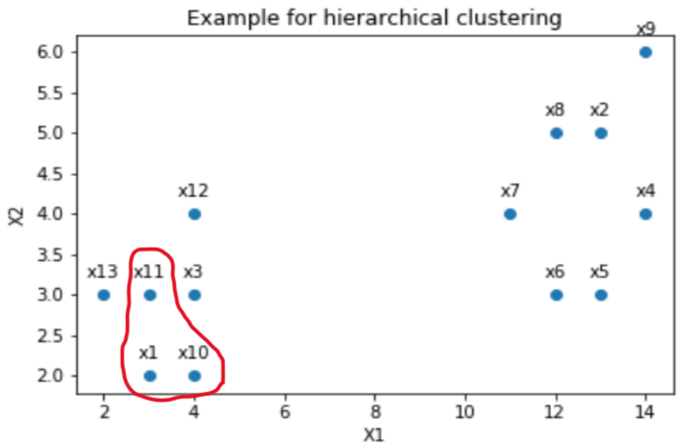
Example

	$\{x_1, x_{10}\}$	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{11}	x_{12}	x_{13}
$\{x_1, x_{10}\}$	0											
x_2	9.49	0										
x_3	1	9.21	0									
x_4	10.20	1.41	10.05	0								
x_5	9.06	2	9	1.41	0							
x_6	3.61	2.24	8	2.24	1	0						
x_7	7.28	2.24	7.07	3	2.24	1.41	0					
x_8	8.54	1	8.25	2.23	2.24	2	1.41	0				
x_9	10.77	1.41	10.44	2	3.16	3.61	3.61	2.24	0			
x_{11}	1	10.20	1	11.05	10	9	8.06	9.22	11.40			
x_{12}	2	9.06	1	10	9.06	8.06	7	8.06	10.20	1.41	0	
x_{13}	2	11.18	2	12.04	11	10	9.06	10.20	12.37	1	2.24	0



Example

	$\{x_1, x_{10}, x_{11}\}$	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{12}	x_{13}
$\{x_1, x_{10}, x_{11}\}$	0										
x_2	9.49	0									
x_3	1	9.21	0								
x_4	10.20	1.41	10.05	0							
x_5	9.06	2	9	1.41	0						
x_6	3.61	2.24	8	2.24	1	0					
x_7	7.28	2.24	7.07	3	2.24	1.41	0				
x_8	8.54	1	8.25	2.23	2.24	2	1.41	0			
x_9	10.77	1.41	10.44	2	3.16	3.61	3.61	2.24	0		
x_{12}	1.41	9.06	1	10	9.06	8.06	7	8.06	10.20	0	
x_{13}	1	11.18	2	12.04	11	10	9.06	10.20	12.37	2.24	0



Different distance measures will affect results

Linkage	Description
Complete	Maximal inter-cluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B, and record the largest of these similarities
Single	Minimal inter-cluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B and record the smallest of these dissimilarities.
Average	Mean inter-cluster dissimilarity. Compute all pairwise dissimilarities between the observations in cluster A and the observations in cluster B and record the average of these dissimilarities.
Centroid	Dissimilarity between the centroid for cluster A (a mean vector of length p) and the centroid for cluster B. Centroid linkage can result in undesirable inversions.

MONTANA
STATE UNIVERSITY

Computation complexity of agglomerative clustering

- Initially it takes $O(n^2)$ time to create the pairwise distance matrix.
- At each merge step, the distance from the merge cluster to all other clusters needs to be recomputed.
  - Distance between the other clusters remain unchanged.
  - In step t, we need to compute $O(n-t)$ distances, we can do this in $O(n)$
- Other operation is to find the closest point in the distance matrix.
  - We have $O(n^2)$ distances in the matrix.
  - If we try to naively find the min, it will take $O(n^2)$.
  - We can improve this by having a min heap.
  - Creating the heap takes $O(n^2)$, finding the min distance takes $O(1)$, deleting and updating takes $O(\log n^2) = O(\log n)$
  - Total time for all merge steps takes $O(n^2 \log n)$

MONTANA STATE UNIVERSITY