Artificial Intelligence I 2022/2023

Week 10 Tutorial and Additional Exercises

K-means & Hierarchical Clustering

School of Computer Science

May 6, 2023

In this tutorial...

In this tutorial we will be covering

- K-means Clustering.
- Hierarchical Clustering.
- Cutting the Dendrogram.

k-means clustering

Recall the formal algorithm of *K-means clustering*:

Algorithm 1: K-means clustering.

Input: $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$: data set, K: number of clusters, $\mathbf{c}_1, \ldots, \mathbf{c}_K$: initial centroids.

Output: Clusters.

repeat

Assignment step: For all $i = 1, \ldots, N$, form clusters by assigning

$$Cluster(\mathbf{x}^{(i)}) \leftarrow \arg \min_{k=1,\ldots,K} Dist(\mathbf{x}^{(i)}, \mathbf{c}_k)$$

where $Dist(\cdot)$ is some distance function, e.g. squared Euclidean distance;

Refitting step: For all $k = 1, \ldots, K$ compute the centroid of the obtained clusters as

$$\mathbf{c}_k \leftarrow \frac{1}{n_k} \sum_{\{i : Cluster(\mathbf{x}^{(i)}) = k\}} \mathbf{x}^{(i)}$$

where n_k is the number of examples in the k-th cluster.

until $\mathbf{c}_1, \ldots, \mathbf{c}_K$ *"stop changing"*;

return *Clusters.*

Exercise 1

- Consider the following data set of 8 examples, each consisting of 2 features

$$\mathbf{x}^{(1)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mathbf{x}^{(2)} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \mathbf{x}^{(3)} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \mathbf{x}^{(4)} = \begin{bmatrix} 5 \\ 2 \end{bmatrix},$$

$$\mathbf{x}^{(5)} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \mathbf{x}^{(6)} = \begin{bmatrix} 7 \\ 4 \end{bmatrix}, \mathbf{x}^{(7)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{x}^{(8)} = \begin{bmatrix} 8 \\ 0 \end{bmatrix}.$$

- Use K-means in algorithm 1 to cluster these examples.
- Use $K = 2$, the squared Euclidean distance as the distance function, and the following initial centroids:

$$\mathbf{c}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \text{ and } \mathbf{c}_2 = \begin{bmatrix} 5 \\ 7 \end{bmatrix}.$$

- Find the clusters and the new centroids at the end of one iteration of algorithm 1 (lines 2 and 3).

Exercise 1: Solution

- We first find the squared Euclidean distance of each example from the two centroids:

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$	$\mathbf{x}^{(7)}$	$\mathbf{x}^{(8)}$
$L^2(\cdot, \mathbf{c}_1)^2$	1	5	5	8	10	32	4	25
$L^2(\cdot, \mathbf{c}_2)^2$	40	36	26	25	25	13	65	58

- The assignment therefore is the following:

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$	$\mathbf{x}^{(7)}$	$\mathbf{x}^{(8)}$
$Cluster(\cdot)$	1	1	1	1	1	2	1	1

- The new centroids are the following:

$$\mathbf{c}_1 = \begin{bmatrix} 4 \\ 9/7 \end{bmatrix}, \text{ and } \mathbf{c}_2 = \begin{bmatrix} 7 \\ 4 \end{bmatrix}.$$

- Run a second iteration of algorithm 1. Find the new clusters and the new centroids.

Exercise 2

- Reconsider the following set with 8 examples and 2 variables

$$\mathbf{x}^{(1)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mathbf{x}^{(2)} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \mathbf{x}^{(3)} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \mathbf{x}^{(4)} = \begin{bmatrix} 5 \\ 2 \end{bmatrix},$$

$$\mathbf{x}^{(5)} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \mathbf{x}^{(6)} = \begin{bmatrix} 7 \\ 4 \end{bmatrix}, \mathbf{x}^{(7)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{x}^{(8)} = \begin{bmatrix} 8 \\ 0 \end{bmatrix}.$$

- Use K-means in algorithm 1 to cluster these examples.
- Use $K = 2$, the squared Euclidean distance as the distance function, but the initial centroids are now:

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{ and } \mathbf{c}_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}.$$

- Find the clusters and the new centroids at the end of one iteration of algorithm 1 (lines 2 and 3).

Exercise 2: Solution

- We first find the squared Euclidean distances of each example from the two centroids:

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$	$\mathbf{x}^{(7)}$	$\mathbf{x}^{(8)}$
$L^2(\cdot, \mathbf{c}_1)^2$	5	17	13	20	10	52	0	49
$L^2(\cdot, \mathbf{c}_2)^2$	2	2	4	5	13	25	9	16

- The assignment therefore is the following:

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$	$\mathbf{x}^{(7)}$	$\mathbf{x}^{(8)}$
$Cluster(\cdot)$	2	2	2	2	1	2	1	2

- The new centroids are the following:

$$\mathbf{c}_1 = \begin{bmatrix} 3/2 \\ 3/2 \end{bmatrix}, \text{ and } \mathbf{c}_2 = \begin{bmatrix} 16/3 \\ 5/3 \end{bmatrix}.$$

- Run a second iteration of algorithm 1. Find the new clusters and the new centroids.

Inter-Cluster Dissimilarity Metrics Revisited

- Distance metrics can be generalised for clusters to define inter-cluster dissimilarity measures. Let C_1 and C_2 be clusters containing n_{C_1} and n_{C_2} examples respectively. Some examples of distance metrics between C_1 and C_2 are:
    1. *Single linkage* is defined as
    $$d_{SL}(C_1, C_2) := \min_{\mathbf{x}^{(1)} \in C_1, \mathbf{x}^{(2)} \in C_2} Dist(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}).$$
    2. *Complete linkage* is defined as
    $$d_{CL}(C_1, C_2) := \max_{\mathbf{x}^{(1)} \in C_1, \mathbf{x}^{(2)} \in C_2} Dist(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}).$$
    3. *Group linkage* is defined as
    $$d_{GL}(C_1, C_2) := \frac{1}{n_{C_1} n_{C_2}} \sum_{\mathbf{x}^{(1)} \in C_1} \sum_{\mathbf{x}^{(2)} \in C_2} Dist(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}).$$
- $Dist(\cdot)$ can be any distance function between vectors.

Hierarchical Clustering

Recall the formal algorithm of *Hierarchical Clustering*.

Algorithm 2: Hierarchical clustering.

Input: Distance matrix corresponding to the data set $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$
Output: Dendrogram.

```
repeat
    Find two clusters C_1, C_2 with the smallest inter-cluster dissimilarity. That is,
        arg min_{C_1,C_2} d_A(C_1, C_2)
    where A ∈ {SL, CL, GL} denotes single linkage (SL), complete linkage (CL)
    or group linkage (GL);
    Merge together C_1, C_2 into a single cluster;
    Note the clusters merged and their corresponding linkage d_A(·, ·) in a
    dendrogram.
until Only one cluster remains.;
return Dendrogram.
```

Exercise 3

- Consider the set with 6 examples and the following distance matrix (for some choice of distance function).

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$
$\mathbf{x}^{(1)}$	0	0.20	0.15	0.76	0.54	0.31
$\mathbf{x}^{(2)}$	0.20	0	0.89	0.18	0.66	0.27
$\mathbf{x}^{(3)}$	0.15	0.89	0	0.82	0.73	0.56
$\mathbf{x}^{(4)}$	0.76	0.18	0.82	0	0.42	0.39
$\mathbf{x}^{(5)}$	0.54	0.66	0.73	0.42	0	0.51
$\mathbf{x}^{(6)}$	0.31	0.27	0.56	0.39	0.51	0

- Use Hierarchical clustering in algorithm 2 to merge all examples into a single cluster.
- Use **single linkage** as the inter-cluster dissimilarity metric.
- Sketch the dendrogram you found along the way clearly depicting the height at which two clusters fuse.

Exercise 3: Solution

- We start with each example in its own cluster and calculate the distance matrix for these clusters.

	$\{\mathbf{x}^{(1)}\}$	$\{\mathbf{x}^{(2)}\}$	$\{\mathbf{x}^{(3)}\}$	$\{\mathbf{x}^{(4)}\}$	$\{\mathbf{x}^{(5)}\}$	$\{\mathbf{x}^{(6)}\}$
$\{\mathbf{x}^{(1)}\}$	0	0.20	0.15	0.76	0.54	0.31
$\{\mathbf{x}^{(2)}\}$	0.20	0	0.89	0.18	0.66	0.27
$\{\mathbf{x}^{(3)}\}$	0.15	0.89	0	0.82	0.73	0.56
$\{\mathbf{x}^{(4)}\}$	0.76	0.18	0.82	0	0.42	0.39
$\{\mathbf{x}^{(5)}\}$	0.54	0.66	0.73	0.42	0	0.51
$\{\mathbf{x}^{(6)}\}$	0.31	0.27	0.56	0.39	0.51	0

- The closest clusters are $\{\mathbf{x}^{(1)}\}$ and $\{\mathbf{x}^{(3)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}, \{\mathbf{x}^{(2)}\}, \{\mathbf{x}^{(4)}\}, \{\mathbf{x}^{(5)}\}, \{\mathbf{x}^{(6)}\}.$$

Exercise 3: Solution (continued)

- We then recalculate the distance matrix for the new clusters.

	$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	$\{\mathbf{x}^{(2)}\}$	$\{\mathbf{x}^{(4)}\}$	$\{\mathbf{x}^{(5)}\}$	$\{\mathbf{x}^{(6)}\}$
$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	0	0.20	0.76	0.54	0.31
$\{\mathbf{x}^{(2)}\}$	0.20	0	0.18	0.66	0.27
$\{\mathbf{x}^{(4)}\}$	0.76	0.18	0	0.42	0.39
$\{\mathbf{x}^{(5)}\}$	0.54	0.66	0.42	0	0.51
$\{\mathbf{x}^{(6)}\}$	0.31	0.27	0.39	0.51	0

- The closest clusters are $\{\mathbf{x}^{(2)}\}$ and $\{\mathbf{x}^{(4)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}, \{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\}, \{\mathbf{x}^{(5)}\}, \{\mathbf{x}^{(6)}\}.$$

Exercise 3: Solution (continued)

- We then recalculate the distance matrix for the new clusters.

	$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	$\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\}$	$\{\mathbf{x}^{(5)}\}$	$\{\mathbf{x}^{(6)}\}$
$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	0	0.20	0.54	0.31
$\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\}$	0.20	0	0.42	0.27
$\{\mathbf{x}^{(5)}\}$	0.54	0.42	0	0.51
$\{\mathbf{x}^{(6)}\}$	0.31	0.27	0.51	0

- The closest clusters are $\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$ and $\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}\}, \{\mathbf{x}^{(5)}\}, \{\mathbf{x}^{(6)}\}.$$

Exercise 3: Solution (continued)

- We then recalculate the distance matrix for the new clusters.

	$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}\}$	$\{\mathbf{x}^{(5)}\}$	$\{\mathbf{x}^{(6)}\}$
$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}\}$	0	0.42	0.27
$\{\mathbf{x}^{(5)}\}$	0.42	0	0.51
$\{\mathbf{x}^{(6)}\}$	0.27	0.51	0

- The closest clusters are $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}\}$ and $\{\mathbf{x}^{(6)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}, \{\mathbf{x}^{(5)}\}.$$

Exercise 3: Solution (continued)

- We then recalculate the distance matrix for the new clusters.

	$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}$	$\{\mathbf{x}^{(5)}\}$
$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}$	0	0.42
$\{\mathbf{x}^{(5)}\}$	0.42	0

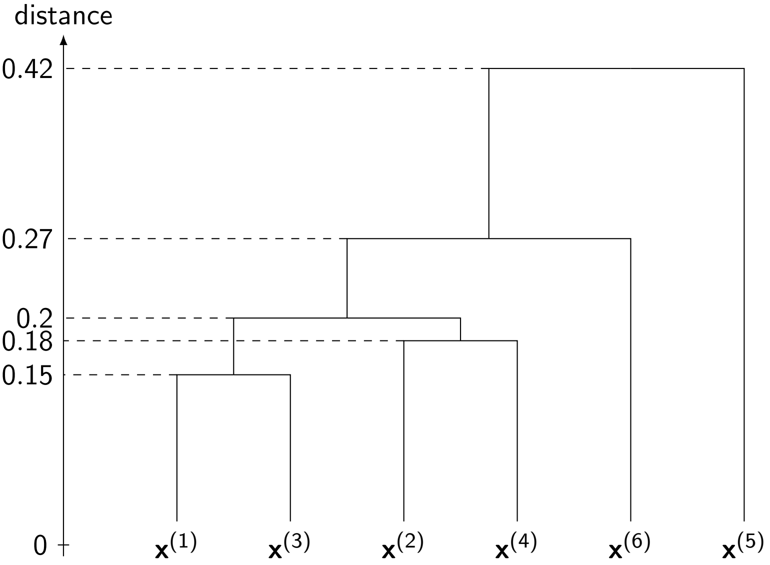
- The closest clusters are $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}$ and $\{\mathbf{x}^{(5)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)}\}.$$

- Finally, we construct the dendrogram.

Exercise 3: Solution (continued)

The dendrogram is the following:



Exercise 4

- Reconsider the set with 6 examples and the following distance matrix (for some choice of distance function).

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$
$\mathbf{x}^{(1)}$	0	0.20	0.15	0.76	0.54	0.31
$\mathbf{x}^{(2)}$	0.20	0	0.89	0.18	0.66	0.27
$\mathbf{x}^{(3)}$	0.15	0.89	0	0.82	0.73	0.56
$\mathbf{x}^{(4)}$	0.76	0.18	0.82	0	0.42	0.39
$\mathbf{x}^{(5)}$	0.54	0.66	0.73	0.42	0	0.51
$\mathbf{x}^{(6)}$	0.31	0.27	0.56	0.39	0.51	0

- Use Hierarchical clustering in algorithm 2 to merge all examples into a single cluster.
- Use **complete linkage** as the inter-cluster dissimilarity metric.
- Sketch the dendrogram you found along the way.

Exercise 4: Solution

- We start with each example in its own cluster and calculate the distance matrix for these clusters.

	$\{\mathbf{x}^{(1)}\}$	$\{\mathbf{x}^{(2)}\}$	$\{\mathbf{x}^{(3)}\}$	$\{\mathbf{x}^{(4)}\}$	$\{\mathbf{x}^{(5)}\}$	$\{\mathbf{x}^{(6)}\}$
$\{\mathbf{x}^{(1)}\}$	0	0.20	0.15	0.76	0.54	0.31
$\{\mathbf{x}^{(2)}\}$	0.20	0	0.89	0.18	0.66	0.27
$\{\mathbf{x}^{(3)}\}$	0.15	0.89	0	0.82	0.73	0.56
$\{\mathbf{x}^{(4)}\}$	0.76	0.18	0.82	0	0.42	0.39
$\{\mathbf{x}^{(5)}\}$	0.54	0.66	0.73	0.42	0	0.51
$\{\mathbf{x}^{(6)}\}$	0.31	0.27	0.56	0.39	0.51	0

- The closest clusters are $\{\mathbf{x}^{(1)}\}$ and $\{\mathbf{x}^{(3)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}, \{\mathbf{x}^{(2)}\}, \{\mathbf{x}^{(4)}\}, \{\mathbf{x}^{(5)}\}, \{\mathbf{x}^{(6)}\}.$$

Exercise 4: Solution (continued)

- We then recalculate the distance matrix for the new clusters.

	$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	$\{\mathbf{x}^{(2)}\}$	$\{\mathbf{x}^{(4)}\}$	$\{\mathbf{x}^{(5)}\}$	$\{\mathbf{x}^{(6)}\}$
$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	0	0.89	0.82	0.73	0.56
$\{\mathbf{x}^{(2)}\}$	0.89	0	0.18	0.66	0.27
$\{\mathbf{x}^{(4)}\}$	0.82	0.18	0	0.42	0.39
$\{\mathbf{x}^{(5)}\}$	0.73	0.66	0.42	0	0.51
$\{\mathbf{x}^{(6)}\}$	0.56	0.27	0.39	0.51	0

- The closest clusters are $\{\mathbf{x}^{(2)}\}$ and $\{\mathbf{x}^{(4)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}, \{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\}, \{\mathbf{x}^{(5)}\}, \{\mathbf{x}^{(6)}\}.$$

Exercise 4: Solution (continued)

- We then recalculate the distance matrix for the new clusters.

	$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	$\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\}$	$\{\mathbf{x}^{(5)}\}$	$\{\mathbf{x}^{(6)}\}$
$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	0	0.89	0.73	0.56
$\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\}$	0.89	0	0.66	0.39
$\{\mathbf{x}^{(5)}\}$	0.73	0.66	0	0.51
$\{\mathbf{x}^{(6)}\}$	0.56	0.39	0.51	0

- The closest clusters are $\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}\}$ and $\{\mathbf{x}^{(6)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}, \{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}, \{\mathbf{x}^{(5)}\}.$$

Exercise 4: Solution (continued)

- We then recalculate the distance matrix for the new clusters.

	$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	$\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}$	$\{\mathbf{x}^{(5)}\}$
$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	0	0.89	0.73
$\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}$	0.89	0	0.66
$\{\mathbf{x}^{(5)}\}$	0.73	0.66	0

- The closest clusters are $\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}$ and $\{\mathbf{x}^{(5)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}, \{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)}\}.$$

Exercise 4: Solution (continued)

- We then recalculate the distance matrix for the new clusters.

	$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	$\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)}\}$
$\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$	0	0.89
$\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)}\}$	0.89	0

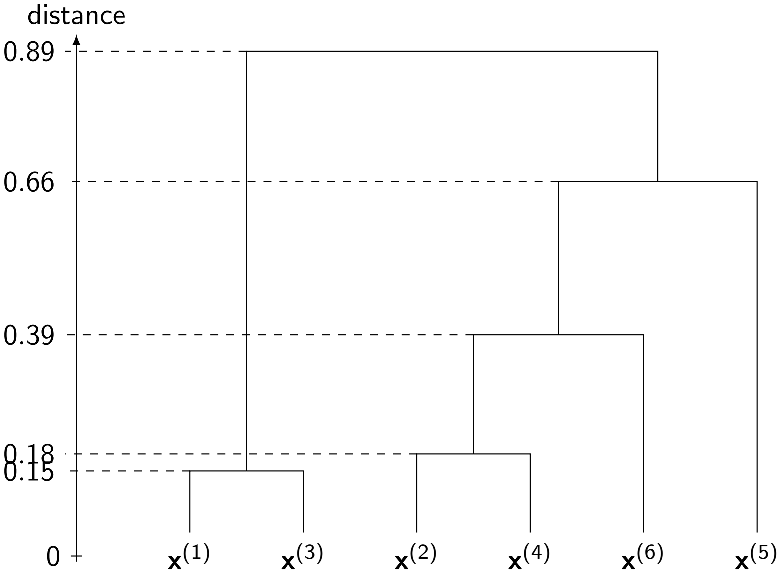
- The closest clusters are $\{\mathbf{x}^{(1)}, \mathbf{x}^{(3)}\}$ and $\{\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)}\}$.
- The new clusters are

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)}\}.$$

- Finally, we construct the dendrogram.

Exercise 4: Solution (continued)

The dendrogram is the following:



Cutting the Dendrogram

- In Hierarchical Clustering, we can impose a threshold on the inter-cluster distance or on the number of clusters.
- When this threshold is surpassed, the algorithm terminates without forming any further clusters, and returns the clusters formed so far.
- Different thresholds can result in different clusters.

Exercise 5

- Reconsider the set with 6 examples and the following distance matrix (for some choice of distance function).

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$
$\mathbf{x}^{(1)}$	0	0.20	0.15	0.76	0.54	0.31
$\mathbf{x}^{(2)}$	0.20	0	0.89	0.18	0.66	0.27
$\mathbf{x}^{(3)}$	0.15	0.89	0	0.82	0.73	0.56
$\mathbf{x}^{(4)}$	0.76	0.18	0.82	0	0.42	0.39
$\mathbf{x}^{(5)}$	0.54	0.66	0.73	0.42	0	0.51
$\mathbf{x}^{(6)}$	0.31	0.27	0.56	0.39	0.51	0

- Use Hierarchical clustering but impose a threshold of 0.35 on the inter-cluster distance. Write the final clusters.
- Use Hierarchical clustering but impose a threshold of 3 on the number of clusters. Write the final clusters.
- Use **single linkage** as the inter-cluster dissimilarity metric.

Exercise 5: Solution

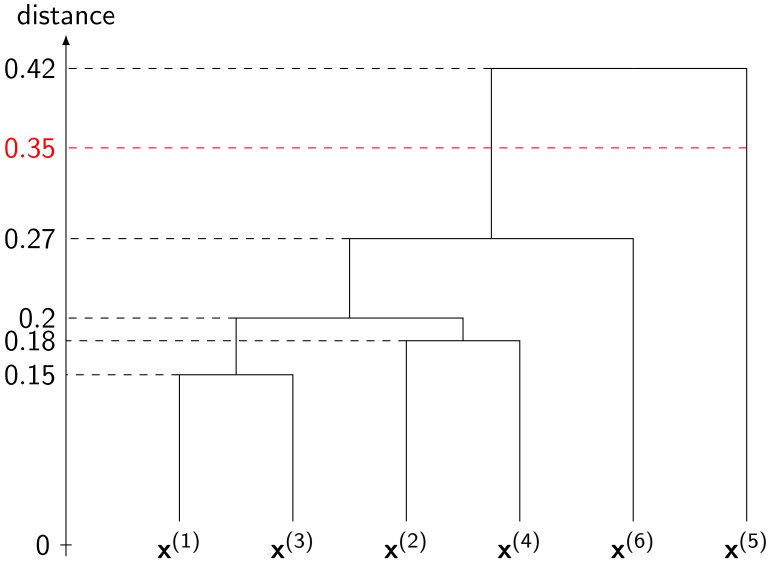
- If we impose a threshold of 0.35 on the inter-cluster distance, the final clusters will be

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}\}, \{\mathbf{x}^{(5)}\}.$$

- We next draw the dendrogram with the horizontal cut.

Exercise 5: Solution (continued)

The dendrogram with the horizontal cut is the following:



Exercise 5: Solution (continued)

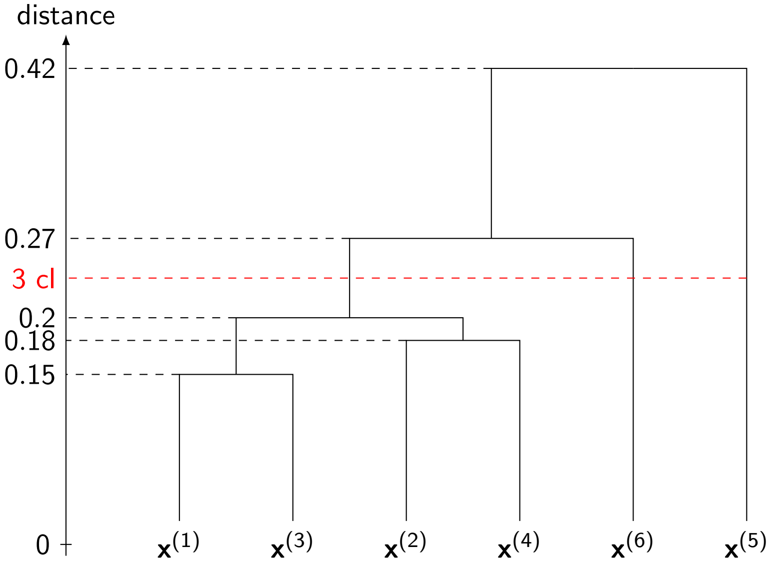
- If we impose a threshold of 3 on the number of clusters, the final clusters will be

$$\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}\}, \{\mathbf{x}^{(5)}\}, \{\mathbf{x}^{(6)}\}.$$

- We next draw the dendrogram with the horizontal cut.

Exercise 5: Solution (continued)

The dendrogram with the horizontal cut is the following:



Optional Material

Optional Exercise 1

- Let C_1, C_2 and C_3 be clusters. Prove the following:
  1. $d_{SL}(C_1, C_2 \cup C_3) = \min\{d_{SL}(C_1, C_2), d_{SL}(C_1, C_3)\}$.
  2. $d_{CL}(C_1, C_2 \cup C_3) = \max\{d_{CL}(C_1, C_2), d_{CL}(C_1, C_3)\}$.
- Recall that
  1. $C \cup C' = \{\mathbf{x} : \mathbf{x} \in C \vee \mathbf{x} \in C'\}$.
  2. $d_{SL}(C, C') = \min_{\mathbf{x} \in C, \mathbf{x}' \in C'} Dist(\mathbf{x}, \mathbf{x}')$.
  3. $d_{CL}(C, C') = \max_{\mathbf{x} \in C, \mathbf{x}' \in C'} Dist(\mathbf{x}, \mathbf{x}')$.
  4. $Dist(\cdot, \cdot)$ is some distance function for vectors.

Optional Exercise 1: Solution

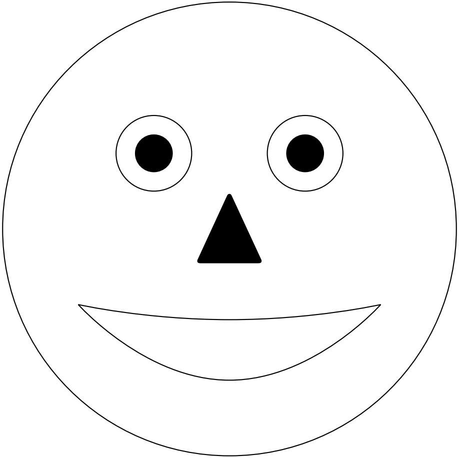
- We proceed as follows

1. $d_{SL}(C_1, C_2 \cup C_3) = \min_{\mathbf{x} \in C_1, \mathbf{x}' \in C_2 \cup C_3} Dist(\mathbf{x}, \mathbf{x}') =$
$\min_{\mathbf{x} \in C_1, \mathbf{x}' \in C_2 \vee \mathbf{x}' \in C_3} Dist(\mathbf{x}, \mathbf{x}') =$
$\min\{\min_{\mathbf{x} \in C_1, \mathbf{x}' \in C_2} Dist(\mathbf{x}, \mathbf{x}'), \min_{\mathbf{x} \in C_1, \mathbf{x}' \in C_3} Dist(\mathbf{x}, \mathbf{x}')\} =$
$\min\{d_{SL}(C_1, C_2), d_{SL}(C_1, C_3)\}$.
2. $d_{CL}(C_1, C_2 \cup C_3) = \max_{\mathbf{x} \in C_1, \mathbf{x}' \in C_2 \cup C_3} Dist(\mathbf{x}, \mathbf{x}') =$
$\max_{\mathbf{x} \in C_1, \mathbf{x}' \in C_2 \vee \mathbf{x}' \in C_3} Dist(\mathbf{x}, \mathbf{x}') =$
$\max\{\max_{\mathbf{x} \in C_1, \mathbf{x}' \in C_2} Dist(\mathbf{x}, \mathbf{x}'), \max_{\mathbf{x} \in C_1, \mathbf{x}' \in C_3} Dist(\mathbf{x}, \mathbf{x}')\} =$
$\max\{d_{CL}(C_1, C_2), d_{CL}(C_1, C_3)\}$.

Q&A

Any questions?

Until the next time...



Thank you for your attention!