

Data Mining & Big Data

Exercise 1 hierarchical algorithms

The following table shows the dissimilarities between 5 objects.

	a	b	c	d	e
a	0				
b	1	0			
c	9	2	0		
d	6	5	3	0	
e	8	6	14	1	0

1. Apply a single-link algorithm. Show the resulting dendrogram.
2. Same question with a complete-link algorithm. Find a cut-off.

Exercise 2 k-means

Let $I = \{1, 2, 18, 20, 31\}$ be a set of elements in a 1-D space. We are looking for a partition of I in 3 clusters.

1. Apply k-means with the centroids 1, 2 and 18. What is the final partition ?
2. Apply k-means with the centroids 18, 20 and 31.
3. Compare the two partitions. How can we measure their quality ?