

T-61.5060 Algorithmic methods in data mining

Exercises November 22, 2007

1. Consider the following table.

	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}_3	\mathcal{C}_4
v_1	1	2	1	1
v_2	2	2	2	2
v_3	2	1	1	1
v_4	2	2	2	4
v_5	3	3	3	3
v_6	3	1	3	2

Apply the BESTCLUSTERING, AGGLOMERATIVE, and BALLS algorithms to this data.

2. Given a 0-1 dataset D with attributes U . Consider the task of finding a subset X of U with minimum cardinality such that for each row t in the data there is at least one attribute $A \in X$ with $t(A) = 1$.

If D contains random bits, each 1 with probability 0.5, give a rough bound for the number of elements that X will have?

3. Consider the following problem. Given a dataset D (from some metric space), a bound c , and an integer k . The task is to find a set $X \subset D$ with $|X| = k$ such that as many points $y \in D$ as possible satisfy

$$\min_{x \in X} d(y, x) < c.$$

Show how to apply the greedy method to this problem. (This is easy.)
What is the relationship of this problem with clustering problems?