The k-center problem

Sunday, 23 April 2023 8:02 PM

Clustering is a very important problem in many different real-life tasks. The idea is to recognize similarities & dissimilarities in large amount of data.

Input :- Undirected complete graph $G = (V, E)$

$d_{ij} \geq 0$ between each $i, j \in V$

Distances obey "Metric" Rules

Integer k (positive)

Goal :- Find k clusters grouping together the vertices that are most similar to each other

$S \subseteq V$, $|S| = k$

and each vertex will assign itself to its closest cluster center.

For k-centers problem, goal is to minimize the maximum distance of a vertex to its cluster center.

Geometric Interpretation :-

Find the centers of k different balls of the same radius that cover all points so that the radius is as small as possible.

$d(i, S) = \min_{j \in S} d_{ij}$

Radius of S is equal to $\max_{i \in V} d(i, S)$

Find a set of size k of minimum radius

Algorithm :-

Pick arbitrary $i \in V$

$S \leftarrow \{i\}$

while $|S| < k$ do

$j \leftarrow \arg\max_{j \in V} d(j, S)$

$S \leftarrow S \cup \{j\}$

Claim :-

Above is a 2-approximation algorithm for the k-center problem.

Proof :- Let $S^* = \{j_1, \ldots, j_k\}$ denote the optimal solution and r^* denote its radius

The solution partitions the nodes V into clusters $V_1, \ldots, V_k$

How?

Each pair of points j and j' in same cluster V_i are at most $2r^*$ apart (Prove why?)

Now consider $S \subseteq V$ given by the algorithm.

If one center in S is selected from each cluster of S^*, then every point in V is clearly within $2r^*$ of some selected point in S.

But suppose algorithm selects two points within the same cluster

Say $j, j' \in V_i$ (i^{st} j selected then j')

Distance between these two points is at most $2r^*$.

Why does algo select j'?

Hence all points are within distance atmost $2r^*$ of some center already selected for S.