k-centre Problem

Frank

December 13, 2021

1 k-centre Problem

1.1 Problem

- **Input:**
  - Set S of "sites" and $k \in \mathbb{Z}^+$
  - $|S| = n$
  - $S \subseteq M$ metric space with distance function d with properties:
    1. $d(x, y) = 0 \iff x = y$
    2. $d(x, y) = d(y, x)$ [symmetry]
    3. $d(x, y) \leq d(x, z) + d(z, y)$ [triangle inequality]

    Elements of M are $(x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$

    And that

    $$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_d - y_d)^2}$$

  - C = set of "centres" i.e. elements of M

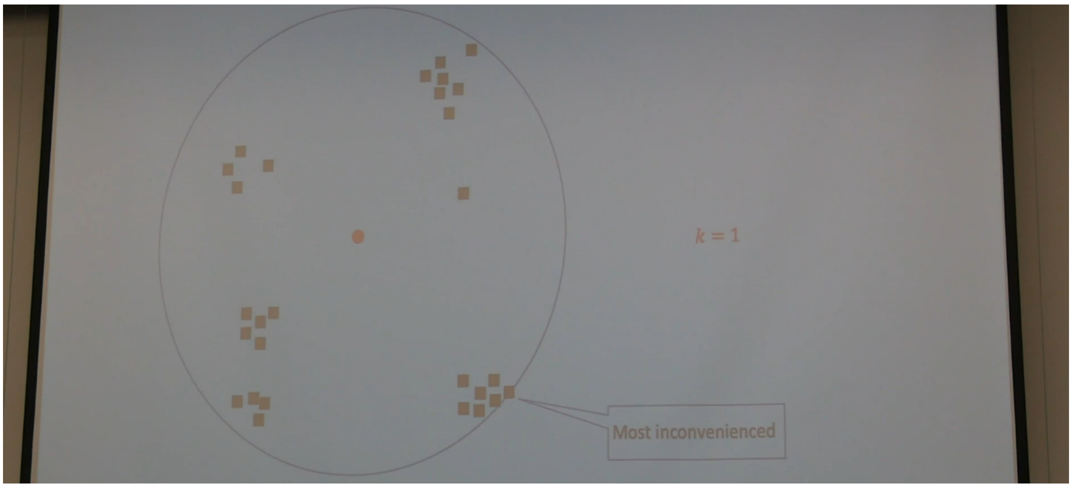
  - Choose k centres so as to minimize the max distance from a site to its closest centre in C

    i.e.

    $$d(s, C) = min_{s' \in C}\ d(s, s')$$
    $$r(C) = max_{s \in S}\ d(s, C)$$

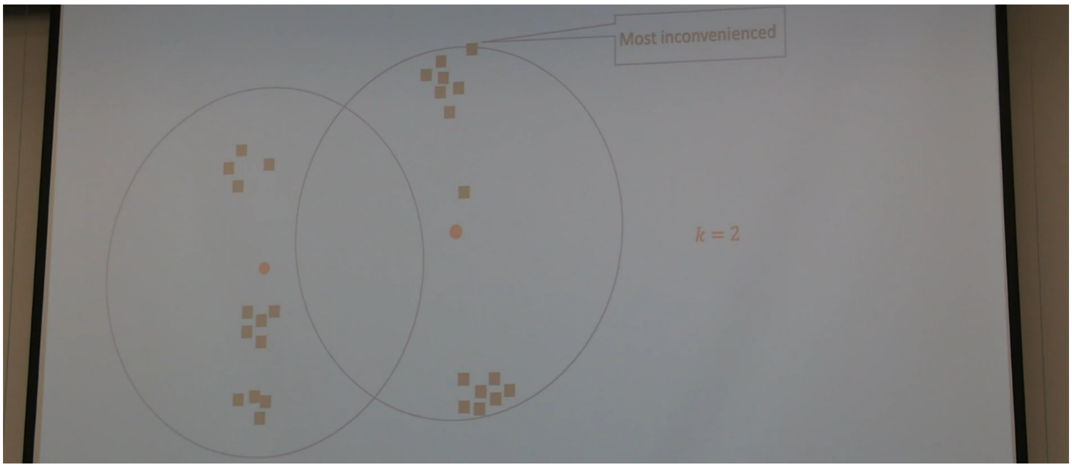
- **Output**:
  - A set C of k centres with minimum radius $r(C)$

- Example with $k = 1$:



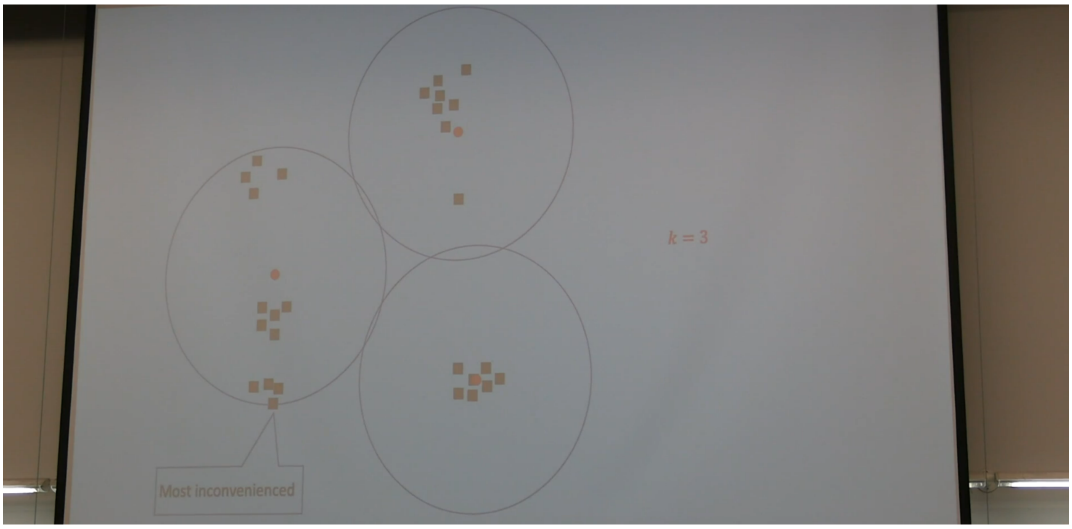
The most inconvenienced is the maximum radius

- Another example with $k = 2$:

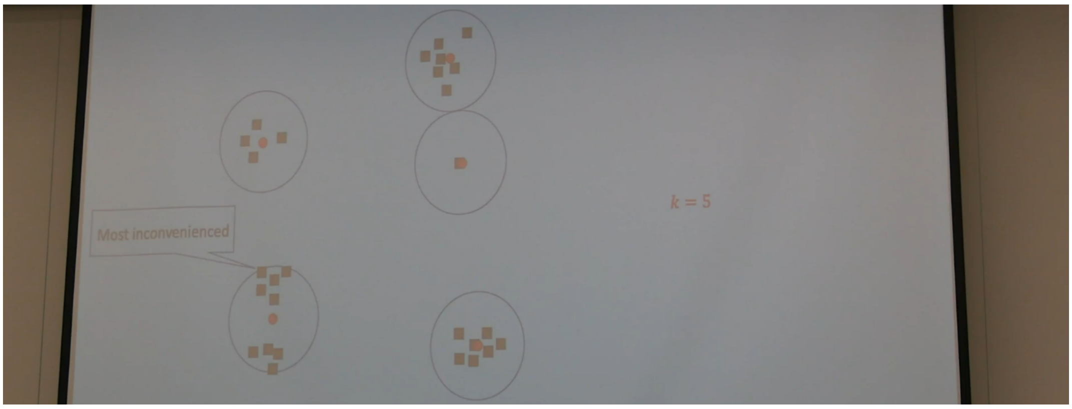


Notice how the radius shrunk.

- Example with $k = 3$:



- Example with $k = 5$:



So we can conclude that as k gets bigger, the radius gets smaller.

- Note that this is an NP-hard problem
- The above examples show centres that are not necessarily sites, but our algorithm is a 2-approximation algorithm that chooses centres from the set of sites.

1.2 Algorithm: Greedy Approach

1. Choose a random site as the centre
2. Pick the most inconvenienced site within the radius of all the centres (pick the radius that is guaranteed to be larger than all other radiuses) and assign that site as a new centre. Shrink the radius according to the new inconvenienced site out of all the centres.
3. Repeat until we have selected k centres

---

procedure K-CENTRE
    $s_1 \leftarrow$ any site in S
    $C_1 \leftarrow \{s_1\}$
    for $(i = 2..k)$ **do**
        $s_i \leftarrow$ site $s \in S$ that maximizes $d(s, C_{i-1})$
        $C_i \leftarrow C_{i-1} \cup \{s_i\}$
    return C_k

1.3 Error Analysis

- Let $C^\star =$ an optimal set of k centres
- **Thm**: $r(C_k) \leq 2 \cdot r(C^\star)$

  Proof:

  - **Claim**: $\forall$ iteration $i \geq 2$, $d(s_i, C_{i-1}) \leq d(s, s')$ $\forall s, s' \in C_i$ s.t. $s \neq s'$

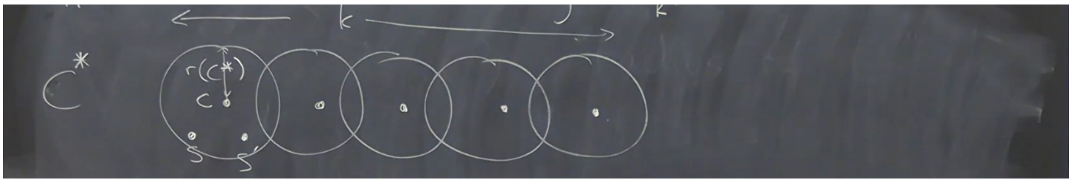
    Proof of claim:

    * True if $s_i = s$ or $s_i = s'$ by default (picked s or s')
    * For $s_i = s$: $d(s, C_{i-1}) = min_{s'' \in C_{i-1}}\ d(s, s'')$ by definition
      Thus, $d(s, C_{i-1}) \leq d(s, s')$ because $s' \in C_{i-1}$
      Similar results for $s_i = s'$
    * So consider the case where $s, s' \in C_{i-1}$
    * Suppose s was added to C in iteration $j \leq i-1$ and s' was added to C in iteration $j' \leq i-1$. So $j \neq j'$, and WLOG, $j' < j \leq i-1$
    * $d(s_i, C_{i-1}) \leq d(s_i, C_{j-1})$ (by decreasing radius as we choose more sites)
    * Now consider that $d(s_i, C_{j-1}) \leq d(s_j, C_{j-1})$ since we added s_j in iteration $j-1$ and thus it is more inconvenienced (more distance) compared to s_i.
    * Now $d(s_j, C_{j-1}) \leq d(s_j, s_{j'}) = d(s, s')$ because j' was added earlier
      $\implies d(s_i, C_{i-1}) \leq d(s, s')$, as wanted

    Now we can use our claim.

  - Imagine we don't stop the algorithm at iteration k but go to iteration $k+1$. So we get $k+1$ sites.
  - Then $s_{k+1} =$ most inconvenienced by C_k

- Consider the optimal set of centres



  Let c be the centre in $C^\star$ whose circle of radius $r(C^\star)$ contains two distinct centres $s, s' \in C_{k+1}$
- $r(C_k) = d(s_{k+1}, C_k) \le d(s, s')$ by claim

  $r(C_k) \le d(s,c) + d(c,s')$ by triangle inequality

  $r(C_k) \le r(C^\star) + r(C^\star) = 2 \cdot r(C^\star)$ since $d(s,c) \le r(C^\star)$ and $d(c,s') \le r(C^\star)$
- Fact: If $P \neq NP$, then there is no polytime approximation algorithm for k-centre with ratio < 2