L19: Approx III

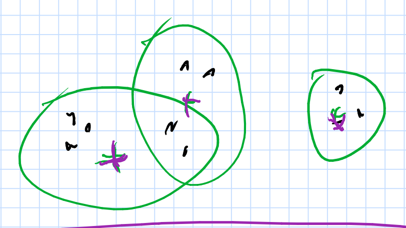
10/28/2021

n items
k: parameter
Partition the items into k groups of similar items.

k-center problem

P: set of n points in the plane or $\mathbb{R}^d$
Euclidean distance

Compute k disks $d_1, d_2, \dots d_k$
s.t.
- $P \subseteq \bigcup d_i$
- $\text{price}(d_1, \dots, d_k) = \max(\text{radius}(d_1), \text{radius}(d_2), \dots \text{radius}(d_k))$.



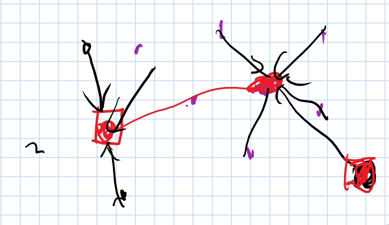
Pick set $C \subseteq P$ of k points s.t.

$r_{opt}(P, k) = \max_{p \in P} \left(\min_{c \in C} \|p - c\| \right)$ ≡ ℓ_∞

Known: Hard to approximate within factor 1.8.
$\leq 1.8\, r_{opt}(P, k)$

2-approx in $O(nk)$ time

- c_1 = arbitrary point P



- c_1 arbit point of P
- for $i = 2$ to k do

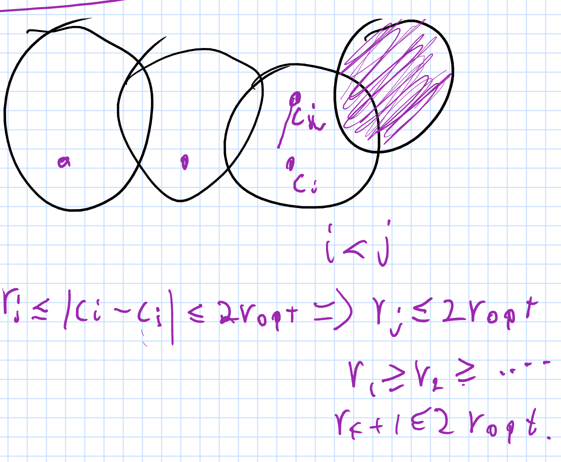
old (for $i = 1$ to n do $p_1 \dots p_n$
  ($\ell_i := \min(p_i, d(c_1, c_2, \dots, c_{i-1}))$ "$O(1)$"
($r_i = \max(\ell_1, \dots, \ell_n)$
($c_i \leftarrow \arg\max_{p \in P} \ell_p$ ← c_i is the point realizing ℓ_i

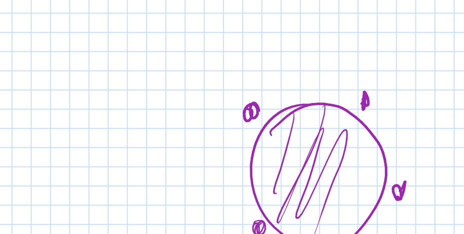
$O(nk)$.

Claim $r_1 \geq r_2 \geq r_3 \dots$

Claim $r_{k+1} \leq 2 r_{opt}(P, k)$

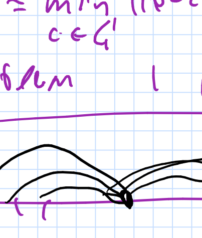
proof





$i < j$

$r_j \leq |c_i - c_j| \leq 2r_{opt} \Rightarrow r_j \leq 2r_{opt}$
$r_1 \geq r_2 \geq \dots \geq r_{k+1}$
$r_{k+1} \leq 2r_{opt}$. ∎



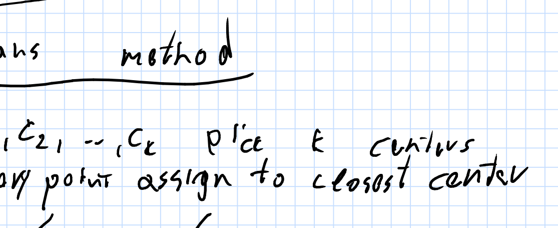
Gonzalez

k-median clustering

C $\text{price}(C, P) = \sum_{p \in P} d(p, C)$

$d(p, C) = \min_{c \in C} \|p - c\|$

1-median problem 1 point.



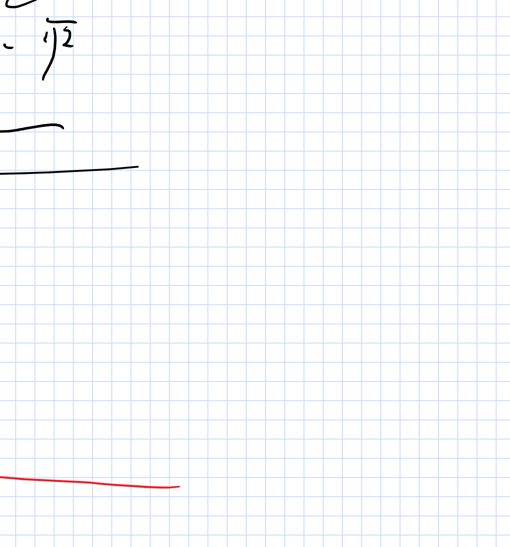
Fermat-Weber point.

k-mean clustering

$C \subseteq \mathbb{R}^d$ $\text{price}(C, P) = \sum_{p \in P} (d(p, C))^2$

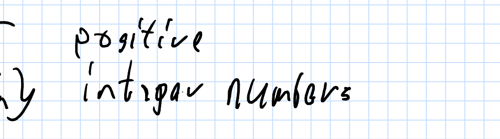
k-means method

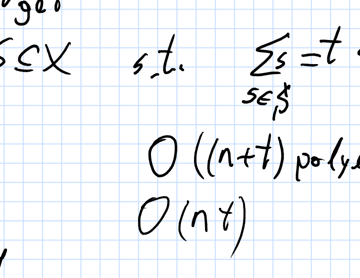
$c_1, c_2, \dots, c_k$ pick k centers
every point assign to closest center
$C_1, C_2, \dots, C_k$
$c_i \leftarrow \text{centroid}(C_i)$



(x, y) — (x', y')

$\sum (x - x')^2 + (y - \ )^2$





Subset sum

$X = \{x_1, x_2, \dots, x_n\}$ positive integer numbers
t: target
Q: $\exists S \subseteq X$ s.t. $\sum_{s \in S} = t$?

$O((n + t)\,\text{polylog})$ FFT
$O(nt)$

$T_0 \leftarrow \{0\}$
For $i = 1$ to n do
  $T_i \leftarrow T_{i-1} \cup (T_{i-1} + x_i)$ $O(tn)$
  $\{s + x_i \mid s \in T_{i-1}\}$
  remove from T_i all numbers bigger than t.
return $t \in T_n$ $t = \Theta(2^n)$

Task: Get a subset sum as close to t as possible, but smaller than t.

Trim(L, δ) L: list of numbers in sorted order (integers).
  $O \leftarrow L$
  $O[1 \dots m]$ sorted list of numbers
  $c \leftarrow O[1]$
  For $i = 2$ to m do
    if $O[i] < (1 + \delta)c$
      continue
    else
      output $O[i]$
      $c \leftarrow O[i]$
[also throw away numbers bigger than t]

Claim Trim(L) output a list of size $O\left(\frac{\ln t}{\delta}\right)$. ←

L_i ≡ Exact list (Computed by the exact algorithm)
$\cup$
T_i: trimmed list

$T_i \leftarrow \text{trim}(T_{i-1} \cup (T_{i-1} + x_i))$

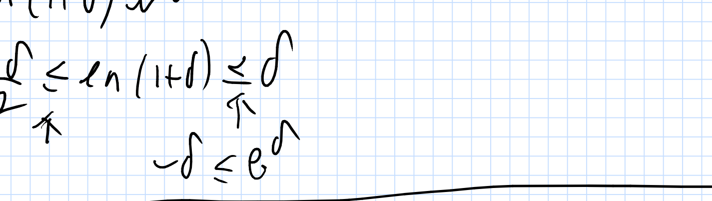
Claim $x \in L_i \Rightarrow y \in T_i$ $y \leq x \leq (1 + \delta)^i y$

$x \in L_n$ $\exists y$ $y \leq x \leq (1 + \delta)^n y$
$1 + \delta \leq \exp(\delta)$ $\leq \exp(\delta n) y$
$\delta = \frac{\varepsilon}{2n}$ $\leq (1 + \varepsilon) y$

$O\left(\frac{\ln t}{\delta}\right) = O\left(\frac{n \ln t}{\varepsilon}\right)$ ←

$O\left(\frac{n^2 \ln t}{\varepsilon}\right)$ RT. $(1 + \varepsilon)$-approx

$u \leq t \leq (1 + \varepsilon) u$ $\varepsilon = 0.01$



$x_i > (1 + \delta) x_{i-1} \geq (1 + \delta)^i$
$(1 + \delta)^i \leq t$ $i \ln(1 + \delta) \leq \ln t$
$i \leq \frac{\ln t}{\ln(1 + \delta)}$

$\ln(1 + \delta) \approx \delta$
$\frac{\delta}{2} \leq \ln(1 + \delta) \leq \delta$
$\delta \leq e^\delta$