FoDA L22

Clustering!

Assignment-based

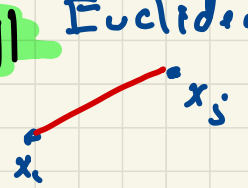
+ Voronoi Diagrams

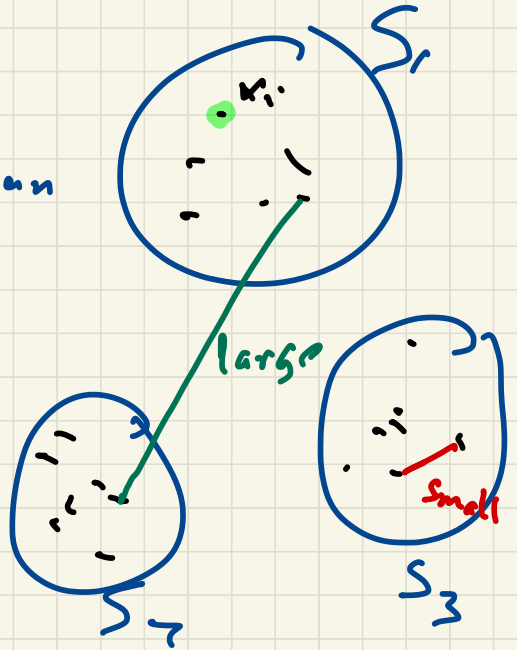
Clustering

Input $X \subset \mathbb{R}^d$ $\quad X = \{x_1, \dots x_n\}$

distance $d: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

$d(x_i, x_j) = \|x_i - x_j\|$ Euclidean



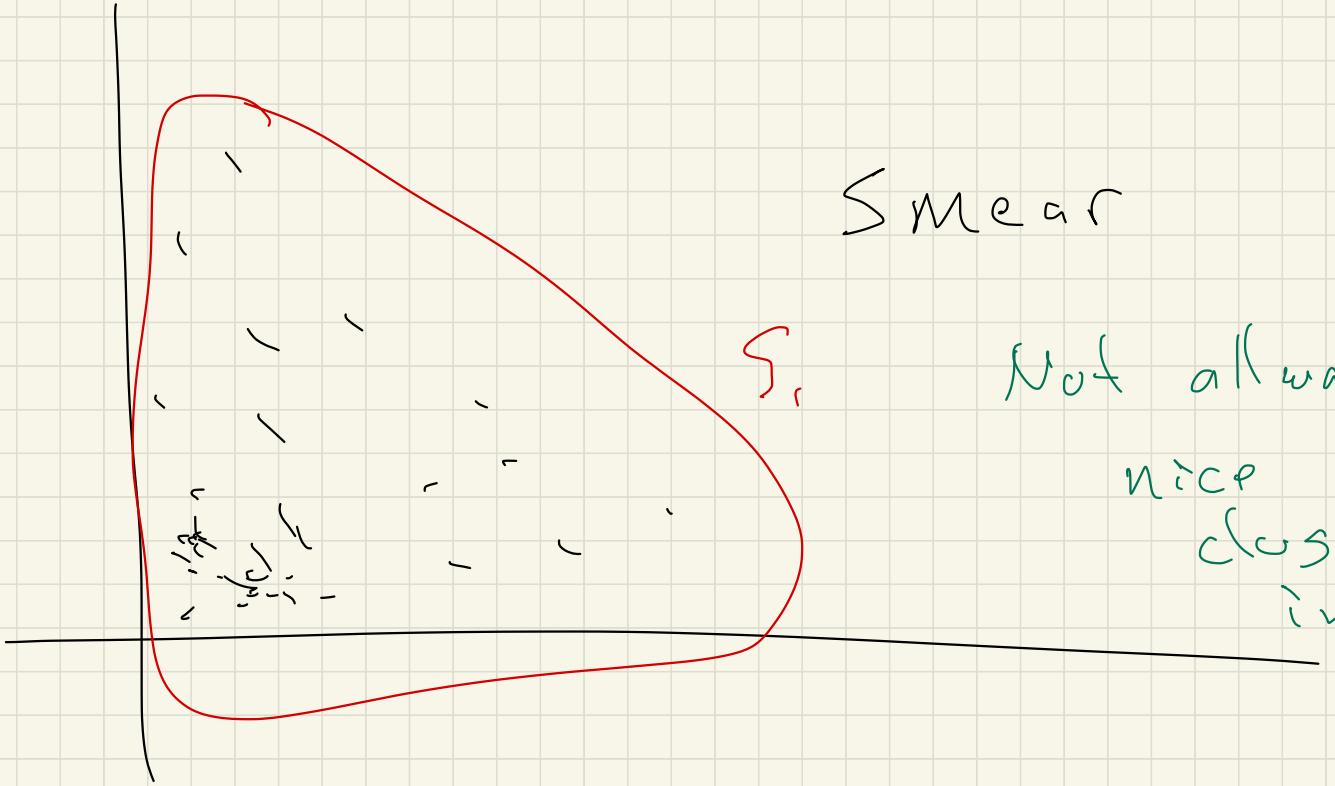


Goal Group data points X into $S_1, S_2 \dots S_k \subset X$

s.t. $x_i, x_i' \in S_i$ $d(x_i, x_i')$ small

$x_i \in S_i$ $x_j \notin S_j$ $d(x_i, x_j)$ large

often in high-dim



Smear

ζ_i

Not always nice clusters in data!

Assignment-based Clustering

Each cluster $S_i \in \{S_1, S_2, \ldots S_k\}$ is associated w/ a point $c_i \in \mathbb{R}^d$

$C = \{c_1, c_2, \ldots c_k\}$

Find C centers to minimize

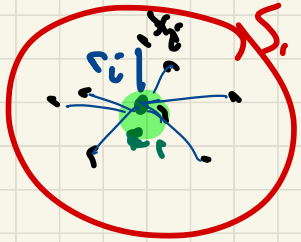
$r_i = d(x_i, c_1)$

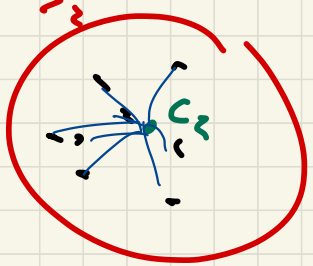
$\phi_C : \mathbb{R}^d \to C$

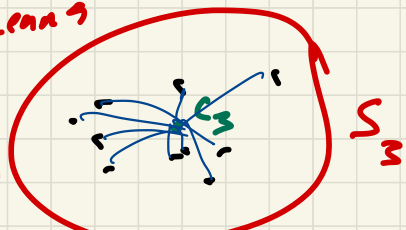
$\phi_C(x) = \underset{c_i \in C}{\operatorname{argmin}}\, d(x, c_i)$

Cost: $\sum_{x_i \in X} \| x_i - \phi_C(x_i) \|^2$ ← closest center

k-means





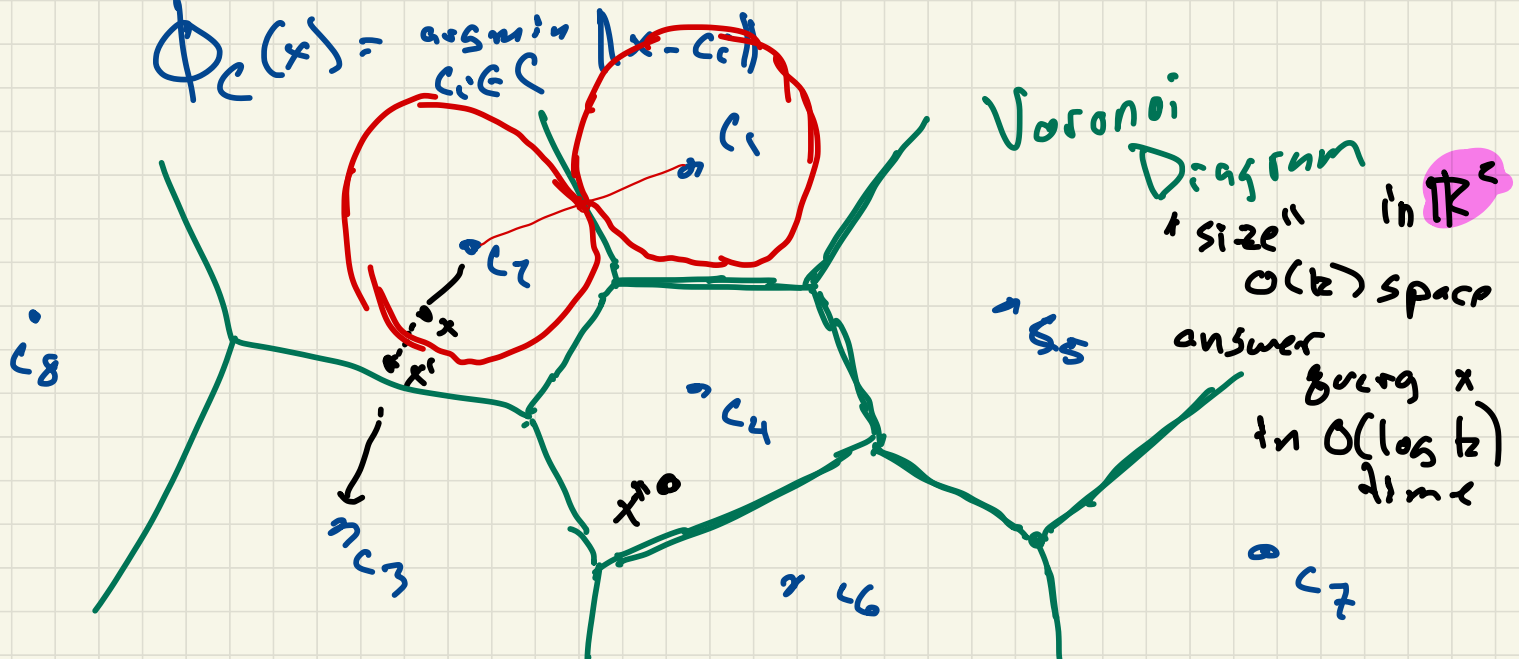


Voronoi Diagrams / Post Office Problem

Nearest Neighbor Problem

Given set k centers/points $C = \{c_1, \dots c_k\} \subset \mathbb{R}^d$

$\Phi_C(x) = \underset{c_i \in C}{\operatorname{argmin}} \|x - c_i\|$

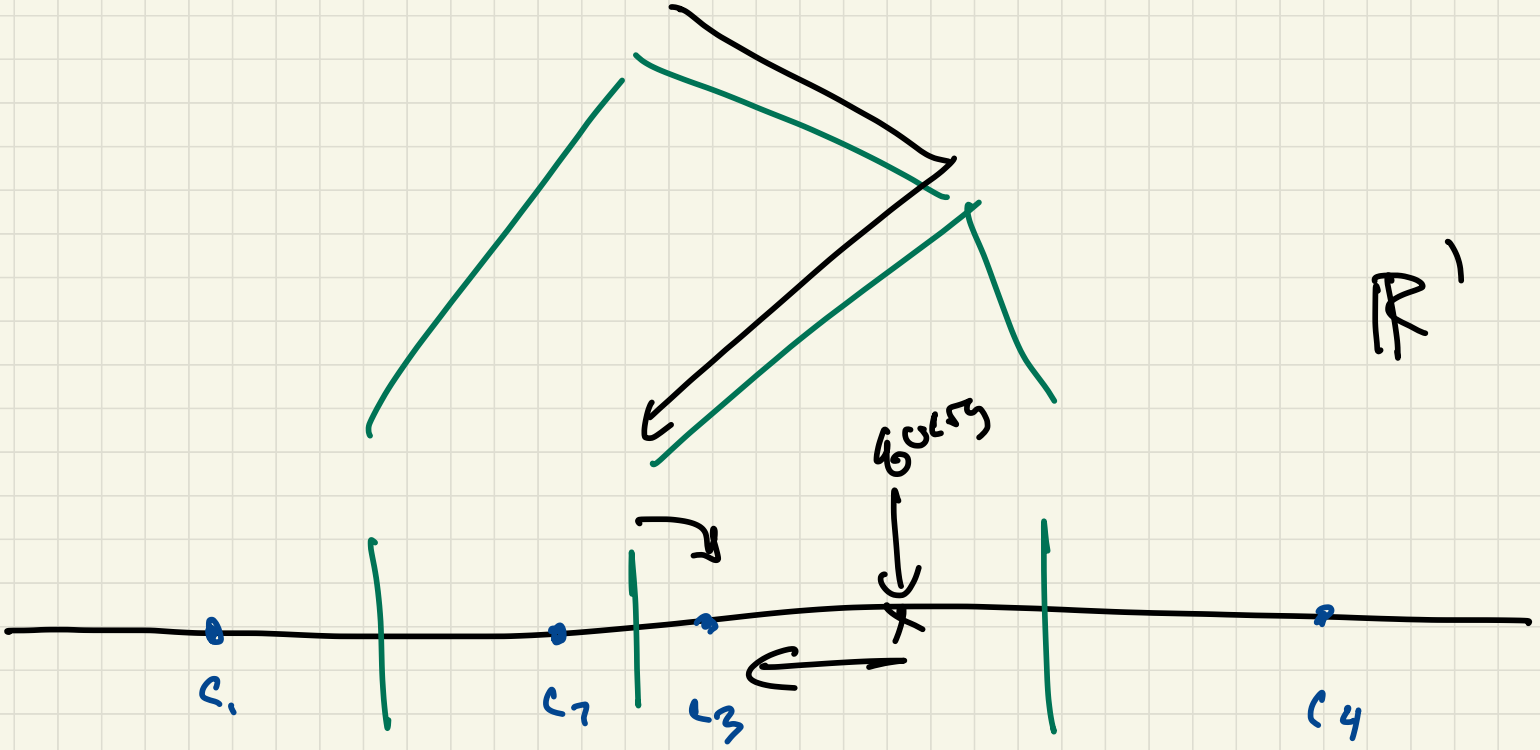


Voronoi Diagram

"size" in $\mathbb{R}^2$

$O(k)$ space

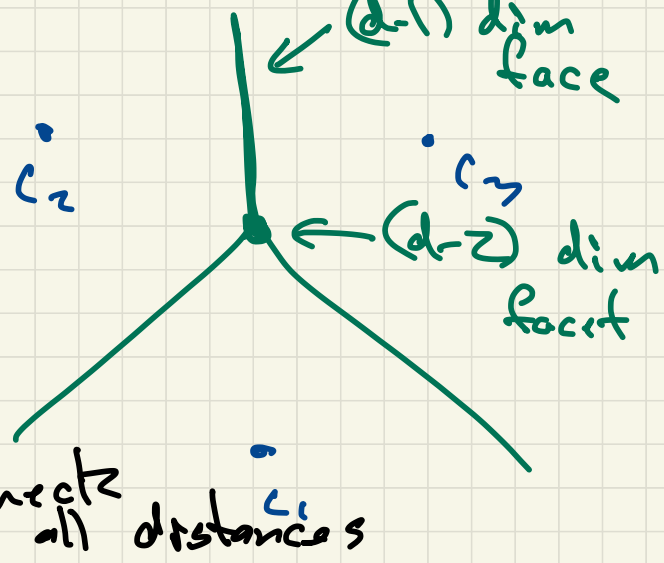
answer query x in $O(\log k)$ time


$\mathbb{R}^1$
c_1
c_2
c_3
c_4

Voronoi Diagrams in $\mathbb{R}^d$

Store Voronoi Diagram
- all boundaries

space $O\left(k^{\lceil d/2 \rceil}\right)$



NN search

d	time
1	$\log k$
2	$\log k$
3	k
d	k

d	space
1	$O(k)$
2	$O(k)$
3	$O(k^2)$
4	$O(k^2)$
5	$O(k^3)$