CPSC 516 — ε-nets and simplex range queries ← by Haussler & Welzl — 15 Nov 2022

Simplex range query: Given a set A of points in $\mathbb{R}^d$, build a query structure that, given a query simplex S, returns $S \cap A$.

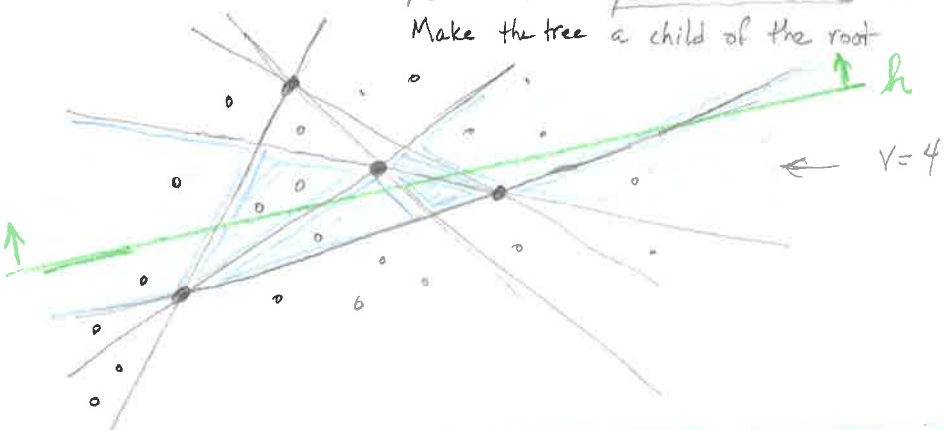
For query $\boxed{\text{half-spaces}}$ h (rather than simplices) one can build a partition tree recursively:

(0) Create a root

(1) Choose a random subset N of A of size v

(2) Form the line arrangement of lines through pairs of points in N

(3) For each cell of this arrangement with $> v$ points of A
Form the (recursively) partition tree for points of A in the cell
Make the tree a child of the root



To search: visit child cells intersected by $\boxed{h}$

To perform efficiently

(1) total number of points in cells intersected by any line is $< \varepsilon \cdot m$ where m is the # points in the current cell

hyperplane bounding h

(2) total number of cells intersected by any line is reasonably small (it's $O(v^2)$ in any case)

constant independent of m

Choose $v = \left\lfloor \frac{c}{\varepsilon}\left(\log \frac{1}{\varepsilon} + \log \frac{1}{\delta}\right) \right\rfloor$ for constant c yields (1) with prob $\geq 1-\delta$

ε-nets & simplex range queries,

A range space is a pair (X, R) where X is a set (of points) and R is a set of ranges $\equiv$ subsets of X

For example $X = \mathbb{R}$ and $R = H_1$

$\mathbb{R}$: real line

H_1: half-lines $(-\infty, a]$ or $[a, -\infty)$ for $a \in \mathbb{R}$

$X = \mathbb{R}$ and $R = I$

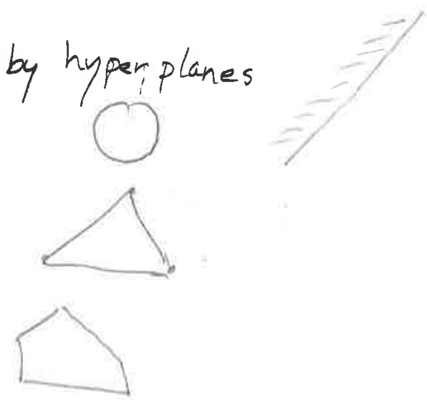
I: intervals $[a, b]$ for $a \leq b \in \mathbb{R}$

$X = \mathbb{R}^d$ $R = H_d$ half spaces in $\mathbb{R}^d$ bounded by hyperplanes

$X = \mathbb{R}^d$ $R = B_d$ balls in $\mathbb{R}^d$

$X = \mathbb{R}^d$ $R = S_d$ d-dim simplices in $\mathbb{R}^d$

$X = \mathbb{R}^d$ $R = C_d$ convex sets in $\mathbb{R}^d$



An ε-net with respect to range space (X, R) for a finite point set $A \subseteq X$ is a set of points $N \subseteq A$ such that N contains a point in r for every $r \in R$ with $\frac{|A \cap r|}{|A|} > \varepsilon$.

i.e. If a range contains a large $(> \varepsilon)$ fraction of the points in A then the range must contain a point in the ε-net N.

Example For halfspaces in $\mathbb{R}^d$, the smallest 0-net for A is the set of extreme points in A.

But for $\varepsilon > 0$ $\exists$ ε-nets for A of size $\left\lceil \frac{8(d+1)}{\varepsilon} \log_2 \frac{8(d+1)}{\varepsilon} \right\rceil$

ε-nets and Vapnik-Chervonenkis dimension

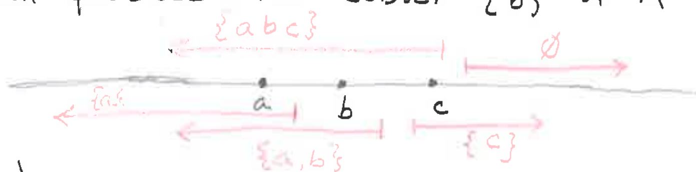
VC-dimension of a range space (X, R) is the size of the largest subset $A \subseteq X$ that can be shattered by R (or ∞ if no largest exists)

all subsets of A

$$\{A \cap r : r \in R\} = 2^A$$

Ex 1 $(\mathbb{R}, H_1)$ has VC-dim = 2

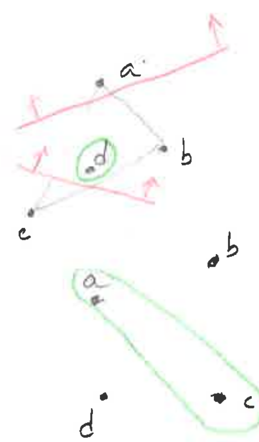
since for any set A with 3 points in $\mathbb{R}$ $a \le b \le c$ no $r \in R$ can produce the subset $\{b\}$ of A



Ex 2 $(\mathbb{R}^2, H_2)$ has VC-dim = 3 !

any set A with 4 points in $\mathbb{R}^2$

- either has one (d) inside Convex Hull
  no $r \in R$ can produce $\{d\}$
- or has 4 on convex Hull
  no $r \in R$ can produce $\{ac\}$ or $\{bd\}$



What is VC($\mathbb{R}^2$, $\check{H}_2$)? = 2

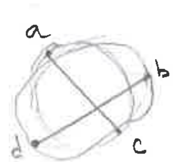
half spaces below hyperplanes nonvertical

Ex 3 $(\mathbb{R}^2, B_2)$ has VC-dim = 3

any set A with 4 points in $\mathbb{R}^2$

- either has one (d) inside Convex Hull
  no $r \in R$ can produce $\{d\}$
- or has 4 on convex hull
  $\Rightarrow$ at least one of $\{ac\}$ and $\{bd\}$ cannot be produced by $r \in R$

why?



two circles would have to intersect 4 times.

use lifting map to relate $VC(\mathbb{R}^d, B_d)$ to $VC(\mathbb{R}^{d+1}, \check{H}_{d+1})$

Ex 4 $(\mathbb{R}^2, \mathcal{C}_2)$ has VC-dim $= \infty$

For any n, a set A with n points in convex position can be shattered: For any subset $B \subseteq A$

① $CH(B) \in \mathcal{C}_2$ and

② $CH(B) \cap A = B$