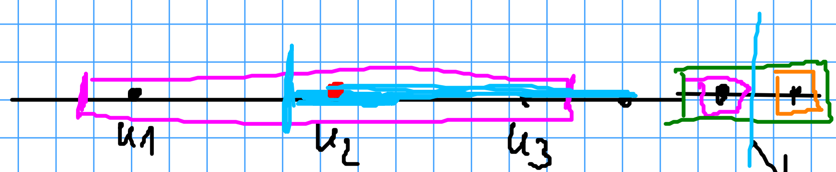
VC-dimension d

Example (Points and Intervals)

$U \subseteq \mathbb{R}$

$\mathcal{S}$ a collection of intervals



no 3 points can be shattered

2 points can be shattered

VC-dim is 2

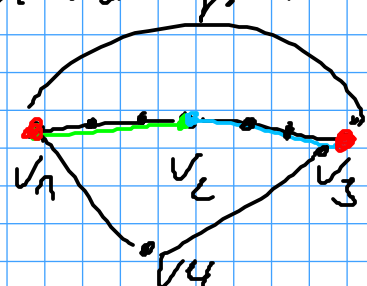
Example (Vertices and Unique Shortest Paths)

graph $G(V, E)$ undirected

$U = V$

$\mathcal{S}$ - collection of unique shortest paths in G

(between any two vertices at most one shortest path)

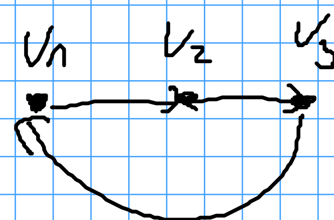


2 vertices can be shattered

3 never

VC-dim is 2

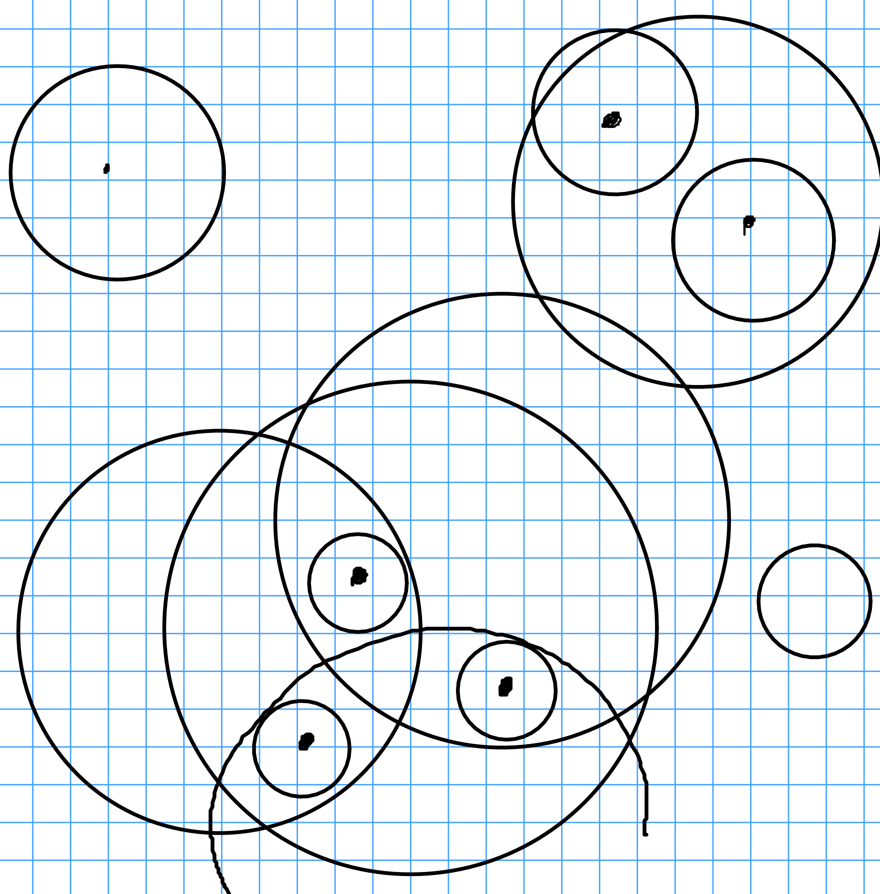
Exercise What is the VC-dim of a system of unique directed shortest paths?



Example (Points and circles)

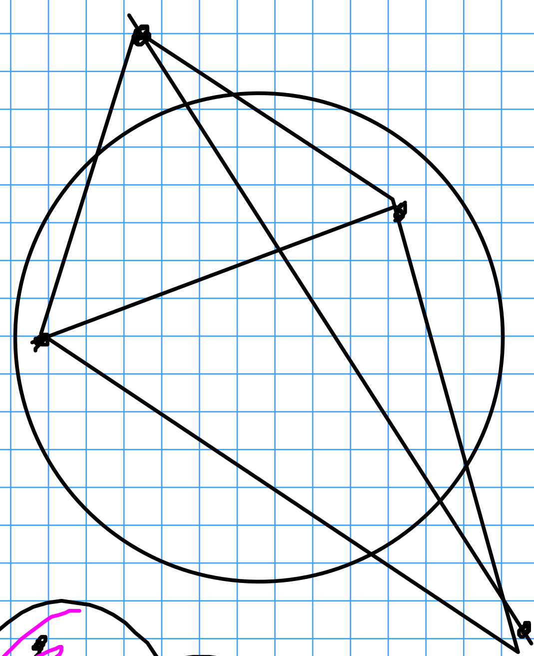
$U \subseteq \mathbb{R}^2$

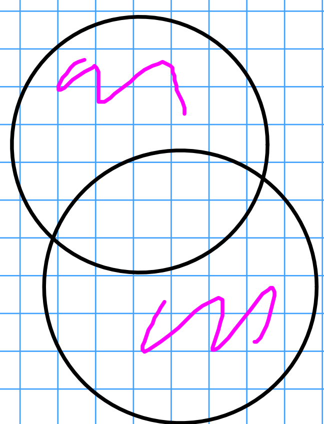
$\mathcal{S}$ a set of circles

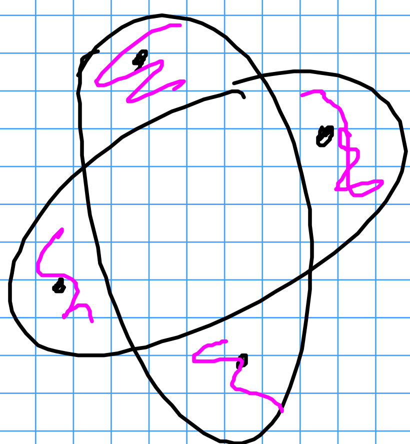


↳ case when points are inside CH is trivial for convex objects

↳ now case when all points are on the CH





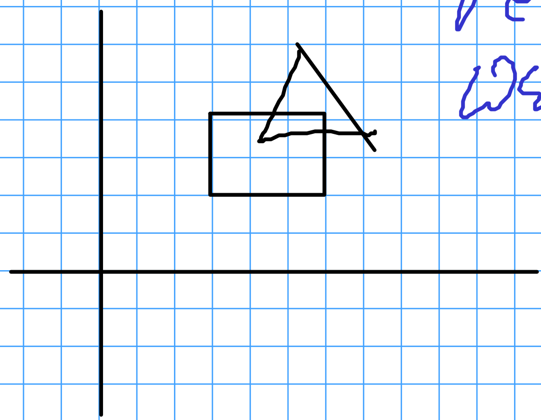


If one pair of diagonal points is separable and the other pair as well, the 2 circles would need to have a symmetric difference consisting of four regions, but circles can have at most 2 such regions $\Rightarrow$ VC-dim is 3

Exercises: What about axis-parallel rectangles?

What about triangles?

$\mathbb{R}^2$



Exercise: Think of a system with unbounded VC-dimension.

We show now: the complexity of a system upper bonds the number of elements in $\mathcal{S}$

trivial UB: $|\mathcal{S}| \leq 2^n$ → exponential in the size of the universe

Lemma (Shatter Lemma)

For a set system with VC-dimension at most d, a subset $U' \subseteq U$ of size m can intersect at most

$$\sum_{i=0}^{d} \binom{m}{i} \leq m^d \text{ sets in } \mathcal{S}.$$

Proof. We will prove the lemma by showing that a system on $|\mathcal{U}|=n$ contains at most n^d sets in $\mathcal{S}$.

Induction:

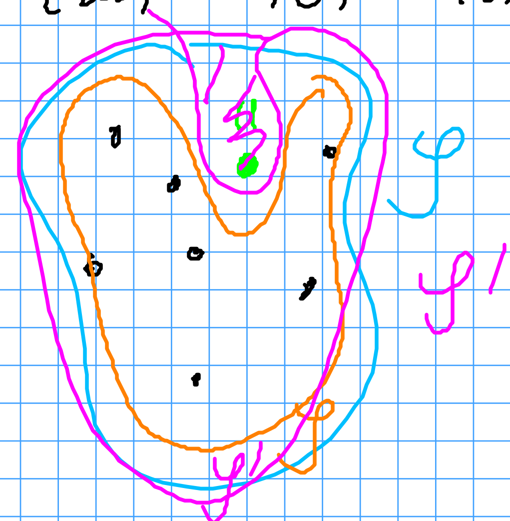
Consider $u \in \mathcal{U}$,

define new set system $(\mathcal{U}', \mathcal{S}')$

with $\mathcal{U}' = \mathcal{U} \setminus \{u\}$

and $\mathcal{S}' = \{ S \setminus \{u\} \mid S \in \mathcal{S} \}$

How much can $|\mathcal{S}|$ and $|\mathcal{S}'|$ differ?



The only way the removal of u triggers a smaller $\mathcal{S}'$ is by existence of $S_a, S_b \in \mathcal{S}$:

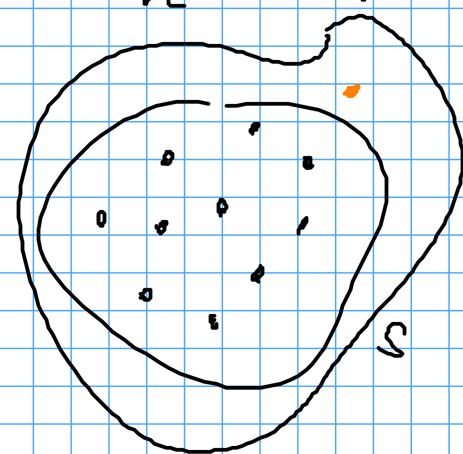
$$S_a = S_b \cup \{u\}$$

$\Rightarrow$ call the set containing all those sets of $\mathcal{S}$ $\mathcal{S}''$

$$\hookrightarrow \quad |\mathcal{Y}| = |\mathcal{Y}'| + |\mathcal{Y}''|$$

$\mathcal{Y}'$: $n-1$, d

$\mathcal{Y}''$: $n-1$, VC-dim?



if one can shatter a set of elements of size s with $\mathcal{Y}''$ then one can shatter those elements unified with u with $\mathcal{Y}$

VC-dim $\Rightarrow \leq d-1$

induction hypothesis

$$|\mathcal{Y}'| \leq \sum_{i=0}^{d} \binom{n-1}{i}$$

$$|\mathcal{Y}''| \leq \sum_{i=0}^{d-1} \binom{n-1}{i}$$

$$|\mathcal{Y}| \leq \sum_{i=0}^{d} \binom{n-1}{i} + \sum_{i=0}^{d-1} \binom{n-1}{i} = \sum_{i=0}^{d} \binom{n}{i}$$

//

⇒ if we have a set system with constant VC-dim, the number of sets in $\mathcal{S}$ is polynomial in the size of the Universe

⇒ In general this bound tight, but for certain systems the bound can be improved

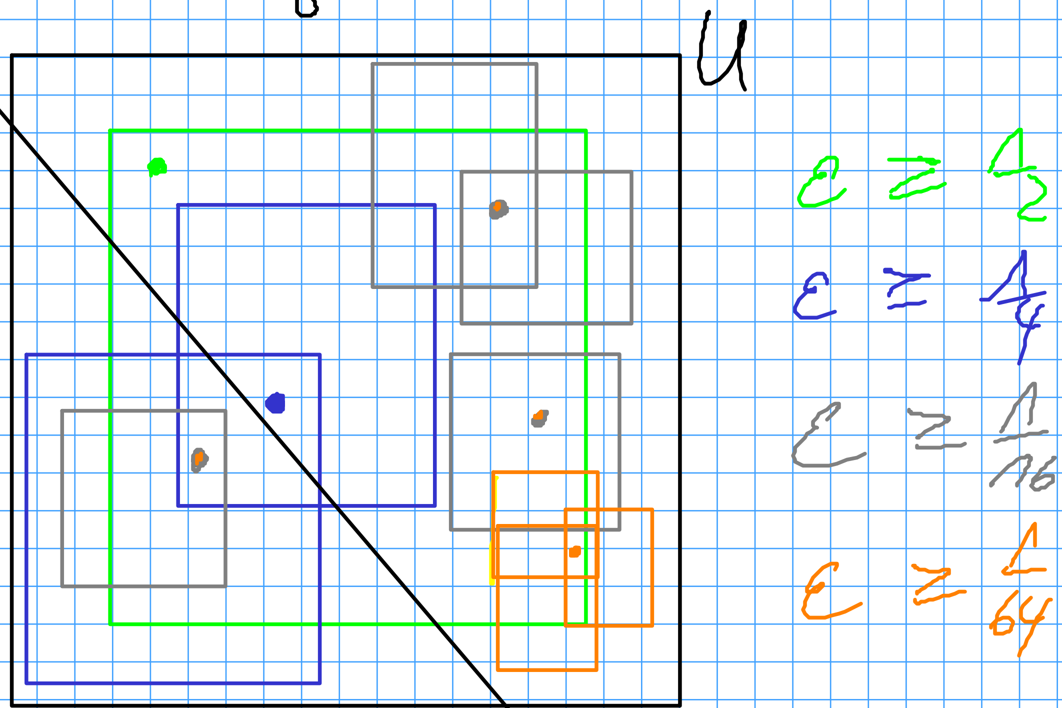
⇒ Brönnimann and Goodrich:
· Hitting sets of size $O(d \log(d \cdot OPT))$
$O(\log(OPT))$ - APX for constant VC-dim.
↳ already below inapproximability bound

2.5.4. ε-Nets

Def. (ε-Net) Let $(U, \mathcal{S})$ be a set system and $\varepsilon \in [0,1]$ a real number. A set $R \subseteq U$ is an ε-net for $(U, \mathcal{S})$, if $\forall S \in \mathcal{S}$ with $|S| \geq \varepsilon |U|$:
$$R \cap S \neq \emptyset.$$

↳ so a 0-net is a Hitting Set, every Hitting set obviously is an ε-net for arbitrary ε, but $\varepsilon \in]0,1[$ a ε-net does not need to be a HS

Example (geometric ε-net)



How to compute an ε-net for given ε?

Let $\mathcal{S}_\varepsilon := \{ S \in \mathcal{S} \mid |S| \geq \varepsilon |U| \}$.

If we draw a random sample R of U with $|R| = r$, the probability of a certain $S \in \mathcal{S}_\varepsilon$ not to be hit is at most $(1-\varepsilon)^r$ (compare basic sampling theorem). So the probability that some set $S \in \mathcal{S}_\varepsilon$ in not hit by R is upper bounded by $|\mathcal{S}_\varepsilon| (1-\varepsilon)^r$.

Accordingly the prob. that all relevant sets are hit is at least

$1 - |\mathcal{S}_\varepsilon| \, (1-\varepsilon)^r$. If we choose

$$r = \frac{(c + \log|\mathcal{S}_\varepsilon|)}{\varepsilon}$$ we get

a success probability of

$$P(R \text{ is } \varepsilon\text{-net}) \geq 1 - |\mathcal{S}_\varepsilon| \cdot (1-\varepsilon)^{\frac{c + \log|\mathcal{S}_\varepsilon|}{\varepsilon}}$$

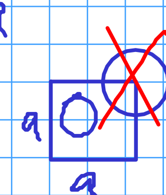
$$= 1 - |\mathcal{S}_\varepsilon| \Big(\underbrace{(1-\varepsilon)^{\frac{1}{\varepsilon}}}_{\text{converges to } \frac{1}{e}} \Big)^{c + \log|\mathcal{S}_\varepsilon|}$$

$$\geq 1 - |\mathcal{S}_\varepsilon| \frac{1}{e^{c + \log|\mathcal{S}_\varepsilon|}}$$

$$= 1 - \cancel{|\mathcal{S}_\varepsilon|} \frac{1}{e^c \cdot \cancel{|\mathcal{S}_\varepsilon|}} = 1 - e^{-c}$$

$\Rightarrow$ the larger we choose c the better

Exercise: Let U be points in a square of side length a

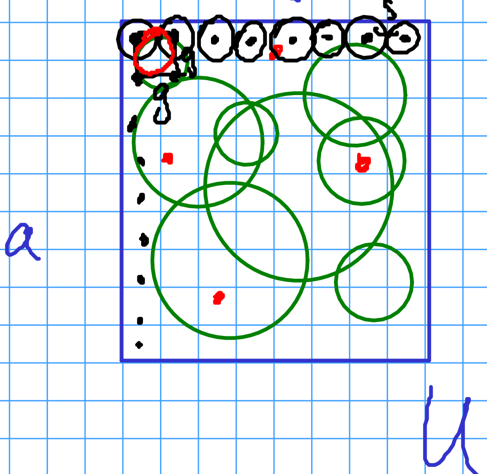


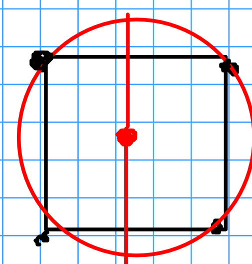
and $\mathcal{S}$ is a collection of

completely contained circles. Prove that for every $\varepsilon \le 1$ there exists an ε-net of size $\frac{\pi}{\varepsilon}$.

For every circle with at least an area of $\varepsilon \cdot a^2$ a hitter is present in a carefully chosen sample of size $\frac{\pi}{\varepsilon}$.

$a = \frac{100}{5}$





↳ grid with cell length of g, we hit all circles with radius $r \ge g$ for sure

↳ $\left(\frac{a}{g}\right)^2 = \frac{a^2}{g^2}$ points in the grid

$A(\text{rectangle}) = a^2$

$A(\text{circle}) \ge \varepsilon \cdot a^2$

$$\pi r^2 \ge \varepsilon \cdot a^2$$

$$r^2 \ge \frac{\varepsilon \cdot a^2}{\pi}$$

$$r \ge \sqrt{\frac{\varepsilon \cdot a^2}{\pi}}$$

for $r \ge \sqrt{\frac{\varepsilon \cdot a^2}{\pi}}$ ↳ insert $g := r$

we hit all circles with $\le \frac{a^2}{a^2 \cdot \frac{\varepsilon}{\pi}} = \frac{\pi}{\varepsilon}$

$\Rightarrow$ if for every $\varepsilon \in (0,1]$ there exists an ε-net of size $\frac{c}{\varepsilon}$ for constant c, We will see that there exists a constant APX for hitting set

Golden Exercise : Show that for a system of unique undirected shortest paths there exists always an ε-net of size $\frac{c}{\varepsilon}$. $(c = 4)$

$\hookrightarrow$ if you solve, no oral exam