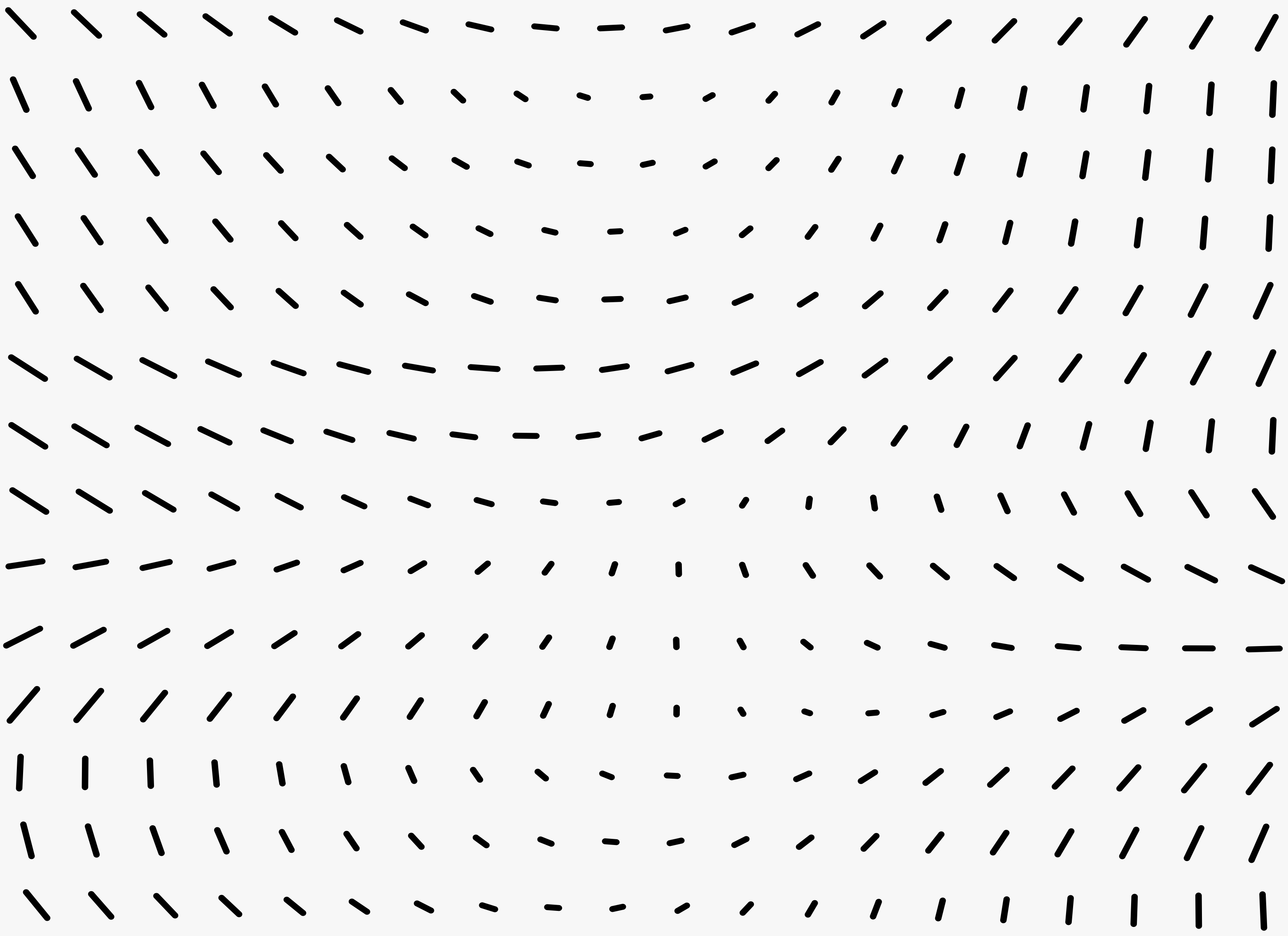
Lecture 28

- VC dim
- ε-net thm via sampling



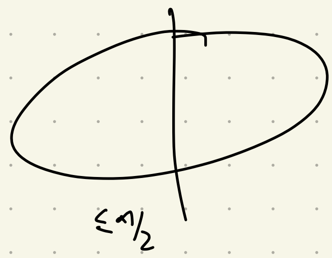
Lecture 28

<u>Def</u>: Given a set system $\mathcal{F}$ on ground set X, $\mathcal{F} \subseteq 2^X$, **the transversal number** of $\mathcal{F}$, denoted by $\tau(\mathcal{F})$, is the min # of elements needed to intersect all sets in $\mathcal{F}$.

$$\min t : \exists\, T \in \binom{X}{t} \text{ s.t. } \forall F \in \mathcal{F},\ T \cap F \neq \emptyset.$$

<u>Ex</u> $\mathcal{F} = \binom{[n]}{n/2}$, $\tau(\mathcal{F}) = \frac{n}{2} + 1$

$\longrightarrow \infty$ as $n \to \infty$



<u>Def</u>: $\mathcal{F} \subseteq 2^X$, X finite, $\varepsilon \in [0,1]$

Say $N \subseteq X$ is an **ε-net** for $\mathcal{F}$ if $N \cap F \neq \emptyset$ $\forall F \in \mathcal{F}$ with $|F| \geq \varepsilon |X|$.

Informally: ε-net cares only large fish.
is a transversal for large sets

<u>Ex</u> $\mathcal{F} = \binom{[n]}{n/2}$ no $1/2$-net of size $n/2$

<u>Q</u>: When can we have bdd size ε-net?

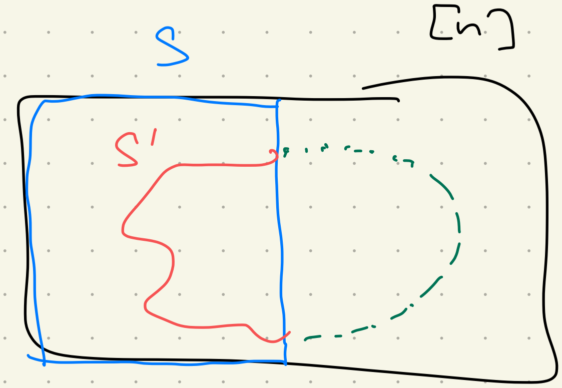
Problematic example: has large 'shattered set'.

Def: Given $\mathcal{F} \subseteq 2^X$, a set $S \subseteq X$ is shattered by $\mathcal{F}$ if $\mathcal{F}\big|_S = 2^S$

i.e. $\forall\, S' \subseteq S$, $\exists\, F \in \mathcal{F}$ s.t. $F \cap S = S'$.

Def: The VC dim of $\mathcal{F} \subseteq 2^X$ is the cardinality of the largest shattered set.

Ex. $\mathcal{F} = \binom{[n]}{n/2}$



- S largest shattered set

  $|S| < n/2 + 1$

  as every set in $\mathcal{F}$ has size $= n/2$

- every $n/2$-set is shattered by $\mathcal{F}$

  $VC(\mathcal{F}) = n/2$

As it turns out a large shattered set is the ONLY obstruction.

Thm [Haussler - Welzl 87]

$\forall\ \mathcal{F} \subseteq 2^X$, VC dim $= d \geq 2$, $\forall\ 0 < \varepsilon < 1/2$

$\Longrightarrow \exists$ an ε-net for $\mathcal{F}$ of size

$$\leq 20 \cdot d \cdot \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} = m$$

May assume that $\forall F \in \mathcal{F}$, $|F| \geq \varepsilon |X|$

Rmk Komlós - Pach - Woeginger : $\varepsilon \to 0$, $\cdots \leq \underbrace{(1+o(1))}_{\text{optimal}} d \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$

Idea • Random sample m pts $N = (x_1, \dots, x_m)$ and show whp N is an ε-net for $\mathcal{F}$.

B : bad event "N not ε-net" $\quad Pr(B) \to 0$

$\qquad\qquad\qquad\qquad\qquad$ |||

$\qquad\qquad\qquad$ "$\exists F \in \mathcal{F}$, $F \cap N = \phi$"

[Hard to analyze directly]

Clever Trick : Take another random sample $M = (y_1, \dots, y_m)$

Consider instead $Pr($ N Bad & M typical $) \simeq Pr(B)$

Couple N & M $\Rightarrow$ localize bad events

& VC dim kicks Sauer-Shelah

Better union .

Pf : • Write $\frac{1}{\varepsilon} = r$, $m = 20 d \cdot r \cdot \log r$.

• Sample uniformly at random (w./ repetition allowed)

$N = (x_1, x_2, \ldots, x_m)$

WTS $\Pr(B) = \Pr(\exists F \in \mathcal{F},\ F \cap N = \phi) \to 0$

• Take $M = (y_1, \ldots, y_m)$ a new set of m random elements . Let $k = m/2r$ ← half of expectation

Define event

$$B' = \text{"} \exists F \in \mathcal{F} \text{ s.t. } \begin{cases} N \cap F = \phi \\ |M \cap F| \geq k \end{cases}$$

multiset

recall $|F| \geq \varepsilon |X| = |X|/r$

Claim $\Pr(B') \leq \Pr(B) \leq 2\Pr(B')$

Pf : • Shall show $\forall N$, $\Pr(B|N) \leq 2\Pr(B'|N)$

• If N is an $1/r$-net $\Rightarrow$ both $= 0$ ✓

• Thus we can fix a set $F \in \mathcal{F}$ s.t. $F \cap N = \phi$.

$$\Rightarrow \Pr(B'|N) \geq \Pr(\underbrace{|M \cap F|}_{} \geq k)$$

Sum of indep. Binom. w./ prob. $\geq 1/r$ so Chernoff

$\geq 1/2$ ⇐

- Suffices to show $Pr(B') \longrightarrow 0$

View sampling N & M differently:

1) a random seq $Z = (z_1, z_2, \cdots, z_{2m})$
2) randomly choose m positions of Z to be N
   rest $\Rightarrow M$.

STS Fix an arbitrary outcome of Z,

$$Pr(B' | Z) \longrightarrow 0$$

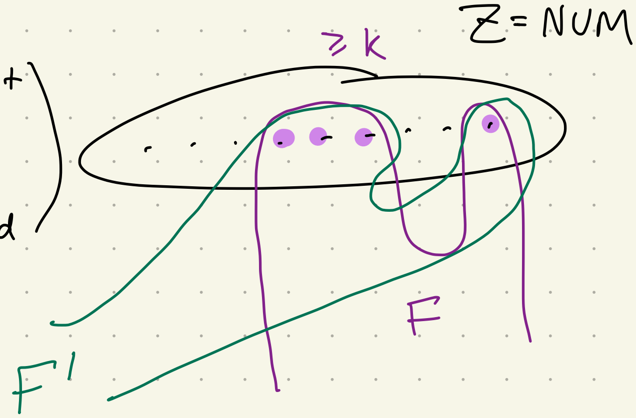
- Fix $F \in \mathcal{F}$, consider the cond. probability

$$p_F = Pr(N \cap F = \phi, |M \cap F| \geq k \mid Z)$$

- Recall $Z = N \cup M$, thus if $|Z \cap F| < k$
  then $p_F = 0$
- Suppose $|Z \cap F| \geq k$, we shall bound

$$p_F \leq Pr(N \cap F = \phi \mid Z)$$

$$= Pr\left(\begin{array}{l}\text{a random } m\text{-set} \\ \text{in } Z \text{ avoids} \geq \\ k \text{ positions occupied} \\ \text{by } F\end{array}\right)$$



$$\leq \frac{\binom{2m-k}{m}}{\binom{2m}{m}} \leq \left(1-\frac{k}{2m}\right)^m \leq e^{-k/2} \leq \varepsilon^{5d}$$

<u>Key obs.</u> $\forall F, F' \in \mathcal{F}$ if $F \cap Z = F' \cap Z$,

then "$N \cap F = \phi$ & $|M \cap F| \geq k$" is the same as

"$N \cap F' = \phi$ & $|M \cap F'| \geq k$"

i.e. B' depends only on trace $\mathcal{F}|_Z = \{F \cap Z : F \in \mathcal{F}\}$

Sauer-Shelah $\Rightarrow$ $|\mathcal{F}|_Z| \leq O(2m)^d \leftarrow O\left(\frac{e \cdot 2m}{d}\right)^d$

Apply union bd (not over $\mathcal{F}$) but over $\mathcal{F}|_Z$

$$\Rightarrow \quad \Pr(B' \mid Z) \leq |\mathcal{F}|_Z| \cdot \varepsilon^{5d}$$

$$= O(2m)^d \cdot \varepsilon^{5d} \longrightarrow 0$$

<u>Lem</u> (Sauer-Shelah) $\forall \mathcal{F}$, VC dim $= d$

$$\forall |Z| = n \Rightarrow |\mathcal{F}|_Z| \leq \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d}$$

<u>Rmk</u> Optimal

$$\{F \cap Z : F \in \mathcal{F}\}$$