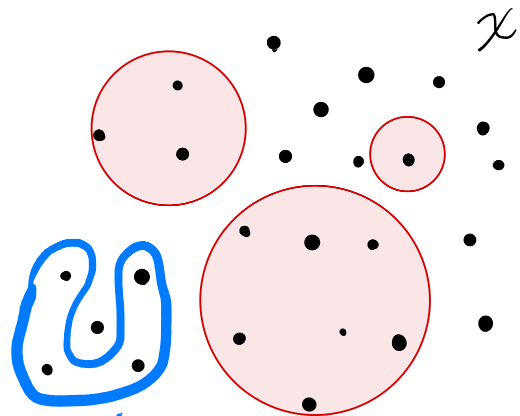
LECTURE 37

Applications of Sauer-Shelah lemma



These three subsets are cut by circles.
This subset is not.

① Consider any n-point set $\mathcal{X} \subset \mathbb{R}^2$.
How many subsets of $\mathcal{X}$ are cut by circles?

$\boxed{\textbf{Prop} \text{ At most } \sum_{k=0}^{3} \binom{n}{k}.}$ $\approx \frac{n^3}{6}$ for large n

Proof: Consider the class of functions

$$\mathcal{F} = \{ \mathbb{1}_C : C \text{ is a circle in } \mathbb{R}^2 \}$$

$vc(\mathcal{F}) = 3$.

Restrict the domain of each function $f \in \mathcal{F}$ to $\mathcal{X}$ $\Rightarrow$ $\mathcal{F}|_{\mathcal{X}}$

$\mathcal{F}|_{\mathcal{X}} = \{ \mathbb{1}_{C \cap \mathcal{X}} : C \text{ is a circle} \}$

indicators of the subsets of $\mathcal{X}$ cut by circles

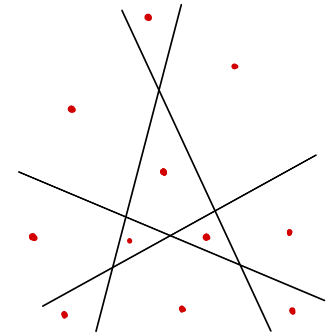
$vc(\mathcal{F}|_{\mathcal{X}}) \le vc(\mathcal{F}) \le 3$. Sauer-Shelah lemma $\Rightarrow$

$$|\mathcal{F}|_{\mathcal{X}}| \le \sum_{k=0}^{3} \binom{n}{k}.$$

Remark Prop is optimal for most sets $\mathcal{X}$
(whenever no 4 pts lie on the same circle)

(2) Consider $\forall$ n lines in $\mathbb{R}^2$.

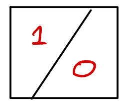
How many regions do they partition $\mathbb{R}^2$ into?



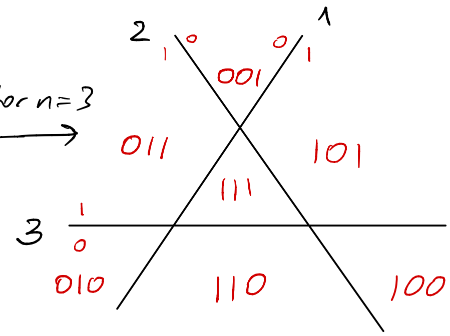
These four lines break $\mathbb{R}^2$ into 11 regions

Prop At most $\sum_{k=0}^{2} \binom{n}{k} = 1 + \frac{n(n+1)}{2}$.

Ex: $n=4$ lines $\Rightarrow 1 + \frac{4 \cdot 5}{2} = 11$ regions

Proof • Choose the orientation of each line $\to$ 

$\Rightarrow$ regions are encoded as binary strings of length n $\xrightarrow{\text{e.g. for } n=3}$



• Identify a binary string with a Boolean function on $\{1, \dots, n\}$

e.g. $0101 \longrightarrow f(1)=0,\ f(2)=1,\ f(3)=0,\ f(4)=1$.

Call the class of these functions $|\mathcal{F}|$.

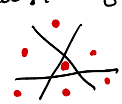
• Then $$|\text{regions}| = |\mathcal{F}| \underset{\substack{\uparrow \\ \text{Sauer-Shelah}}}{\leq} \sum_{k=0}^{d} \binom{n}{k} \quad \text{where } d = vc(\mathcal{F})$$

• $\boxed{\text{Claim: } vc(\mathcal{F}) \leq 2}$. Assume the contrary: $vc(\mathcal{F}) \geq 3$

$\Rightarrow \exists$ a 3-element subset of $\{1, \dots, n\}$ shattered by $\mathcal{F}$

$\Leftrightarrow \exists$ 3 lines that realize all 8 binary strings of length 8

$\Rightarrow$ these 3 lines break $\mathbb{R}^3$ into at least 8 regions.

But this is impossible: max # regions = 7:  QED.

(0) $\exists$ a simple inductive proof of Prop.

Remarks (1) More generally, $\forall$ $n \geq d$ hyperplanes partition $\mathbb{R}^d$ into at most $\sum_{k=1}^{d} \binom{n}{k}$ regions [Buck 1943]. The proof is similar.

(2) The bound is optimal, attained $\forall$ hyperplanes in general position.

(3) If not in general position, see enumerative combinatorics / hyperplane arrangements / Möbius function / Zaslavsky formula.

EMPIRICAL PROCESSES

- Let X be a r.v. taking values in $\forall$ set $\mathcal{X}$, $X_1, \dots, X_n$ be independent copies of X.
- Law of large numbers $\Rightarrow$ $\forall$ Boolean function $f : \mathcal{X} \to \{0,1\}$:

$$\frac{1}{n} \sum_1^n f(X_i) \xrightarrow{a.s.} \mathbb{E} f(X) \quad \text{as } n \to \infty$$

- Deviation:

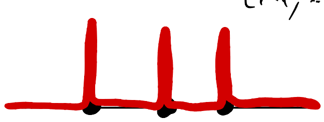
$$\mathbb{E} \left| \frac{1}{n} \sum_1^n f(X_i) - \mathbb{E} f(X) \right| \overset{L^1 \le L^2}{\le} \left(\mathbb{E} | \cdots |^2 \right)^{1/2} = \operatorname{Var}\left(\frac{1}{n} \sum_1^n f(X_i) \right)^{1/2}$$

$$= \Big[\frac{1}{n^2} \sum_1^n \underbrace{\operatorname{Var}(f(X_i))}_{\le 1 \ (f \text{ Boolean})} \Big]^{1/2} = \frac{1}{\sqrt{n}} . \qquad (\text{"Weak LLN"})$$

- Is this true uniformly over all Boolean functions f?

$$\mathbb{E} \sup_{\forall f} \left| \frac{1}{n} \sum_1^n f(X_i) - \mathbb{E} f(X) \right| \to 0 \quad \text{as } n \to \infty \ ?$$

NO For $f = \mathbb{1}_{\{X_1, \dots, X_n\}}$, $\quad \frac{1}{n} \sum_1^n f(X_i) = 1$ but $\mathbb{E} f(X) = 0$



- But it is true uniformly over $f \in \mathcal{F}$ whenever $vc(\mathcal{F}) < \infty$:

THM (Uniform Law of Large Numbers)

$$\boxed{\text{If } \mathcal{F} \text{ is a Boolean class with } d = vc(\mathcal{F}) < \infty, \text{ then } \quad \mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E} f(X) \right| \le C \sqrt{\frac{d \log n}{n}} .}$$

Remarks ① Same rate $O(1/\sqrt{n})$ as in WLLN!

② $\left(\frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E} f(X) \right)_{f \in \mathcal{F}}$ is called an empirical process.

The proof uses a new tool: