# Lecture 16

definitions $\left.\begin{array}{l}\text{Restrictions} \\ \text{Growth function} \\ \text{VC dimension}\end{array}\right\}$

- finite VC dim $\Longrightarrow$ Uniform Convergence (Part 1)
- Sauer - Shelah - Perles Lemma

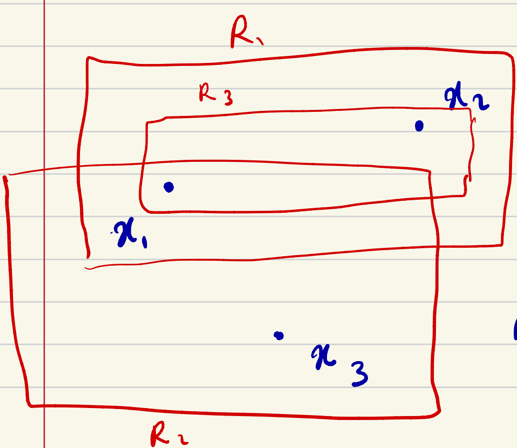
## Def. Restriction of $C$ to $S$.

Let $S$ be a set of $m$ points in domain $X$. $S = \{x_1, \dots, x_m\}$

The restriction of $C$ to $S$ is the set of functions from $S$ to $\{0,1\}$ that can be derived from $C$.

$$C_S : \{ (c(x_1), c(x_2), \dots, c(x_m)) \mid c \in C \}$$

where we represent each function from $S$ to $\{0,1\}$ as a vector in $\{0,1\}^{|S|}$ or $\{0,1\}^m$




$C = \{R_1, R_2, R_3\}$

assign positive label to points inside the rectangle

$$\text{Restrictions}: \left\{ \begin{array}{cccc} & x_1 & x_2 & x_3 \\ & (+, & +, & -) \\ & (+, & -, & +) \end{array} \right.$$

while $C$ might have infinitely many hypotheses, its "effective size" is small

def. growth function

Let $C$ be a concept class. Then, the growth function of $C$, denoted $\tau_C : \mathbb{N} \to \mathbb{N}$, is defined as:

$$\tau_C(m) = \max_{S \subset X : |S| = m} |C_S|$$

$\tau_C(m) \approx$ number of functions from $S$ to $\{0,1\}^m$ that can be obtained by $c \in C$.

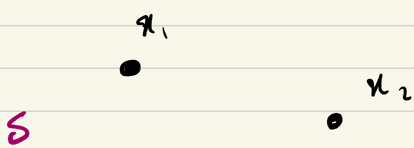
- With no assumption, we know $|C_S|$ is bounded by $2^{|S|} = 2^m$

## def. shattering

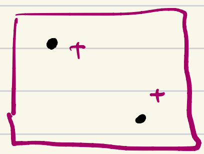
A Class $C$ shatters a finite set $S$ if the restriction of $C$ to $S$ is the set of all functions from $C$ to $\{0, 1\}$. That is $|C_S| = 2^{|S|} = 2^m$

---

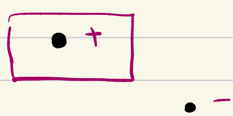
Example $C$ = axis-aligned rectangles




$(+, +)$

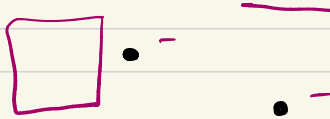



$(+, -)$

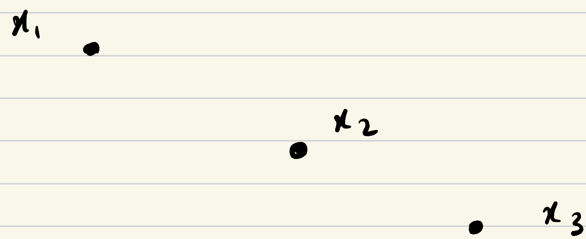



$(-, +)$


$(-, -)$



How about 3 points ?



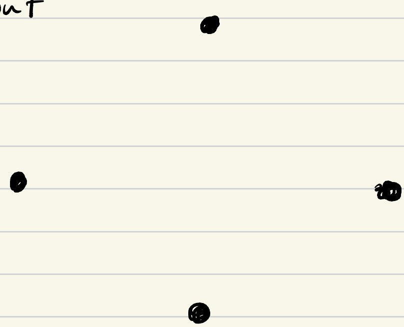

Can you label them with

$(+, -, +)$

$C$ does not shatter this $S$.

How about

4 points?



---

what we have shown earlier indicates:

if $C$ shatters $S$, we <u>cannot</u>

learn with $|S|/2 = m/2$ samples.

# Def. VC Dimension

The VC dimension of a concept class $C$, denoted by $VCdim(C)$, is the maximal size of a set $S$ that can be shattered by $C$.

If $C$ can shatter sets of arbitrary large size, we say $VCdim(C) = \infty$

---

Example 1:

$VCdim$ ( Axis-aligned rectangle ) $= 4$

We need to show:

- there is a set of size 4 that is shattered.

- No set of size 5 is shattered.

Example 2: finite classes:

$$|C_S| \leq |C| = 2^{\log |C|}$$

C cannot shatter any set of size larger than $\log |C|$

$$VCdim(|C|) \leq \log |C|$$

$\longleftrightarrow$

If $VCdim(C) = d$

$$\forall\ m \leq d \Rightarrow \tau_C(m) \leq 2^m$$

$$\forall\ m > d \Rightarrow \tau_C(m) < 2^m$$

# VC dimension

- infinite classes can still be PAC-learnable.

$\Rightarrow$ size is <u>not</u> determinant of learnability.

So, what is then?

$\boxed{\text{VC-dim of } C \text{ characterizes its learnability!}}$

The fundamental theorem of PAC learning

For a concept class $C$ of $c: X \to \{-1, +1\}$ with 0-1 loss function, the following are equivalent:

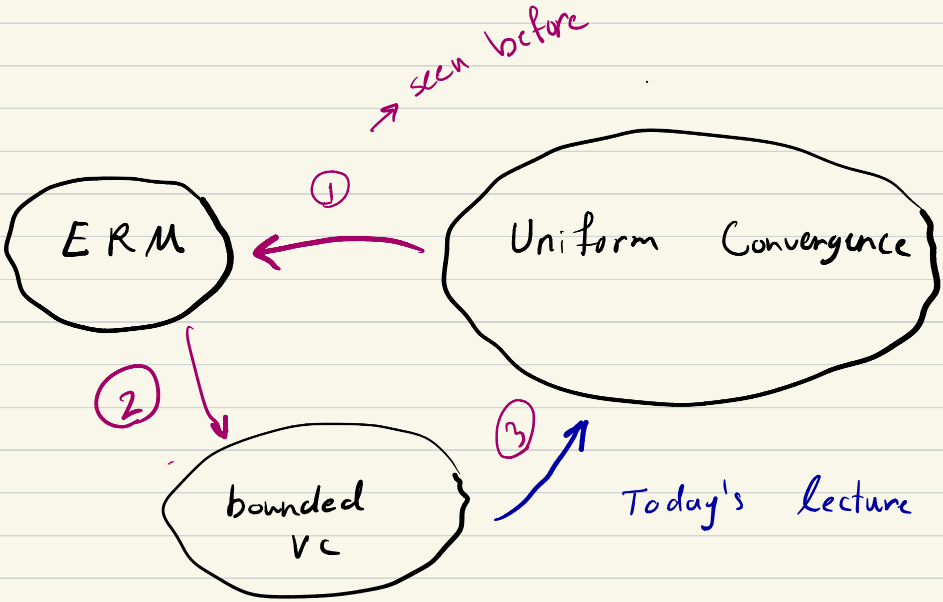
- $C$ has uniform convergence.

- Any ERM is a successful agnostic PAC learner

- $H$ has a finite VC dim.

what have left to show is :

finite VCdim $\Rightarrow$ Uniform convergence.




2 last time we have shown if $VC > 2m$ ER does not work with m samples.

ERM work $\Rightarrow$ $VC < m$

with m samples

Proof of 3 has two steps

1 Sauer's Lemma:

If $\text{VCdim}(C) \le d$ :

$$\tau_C(m) \le m^d$$

2 $|S| = m$

$c \in C$ : $\left| err(c) - err(c) \right| \approx \sqrt{\dfrac{\log(\tau_C(2m))}{2m}}$

$m \approx \dfrac{d}{\varepsilon^2} \implies$ uniform convergence

# Sauer–Shelah–Perles Lemma

$\boxed{1}$ If $VCdim(C) \leq d < \infty$, then

$$\forall m \quad \tau_C(m) \leq \sum_{i=0}^{d} \binom{m}{i}$$

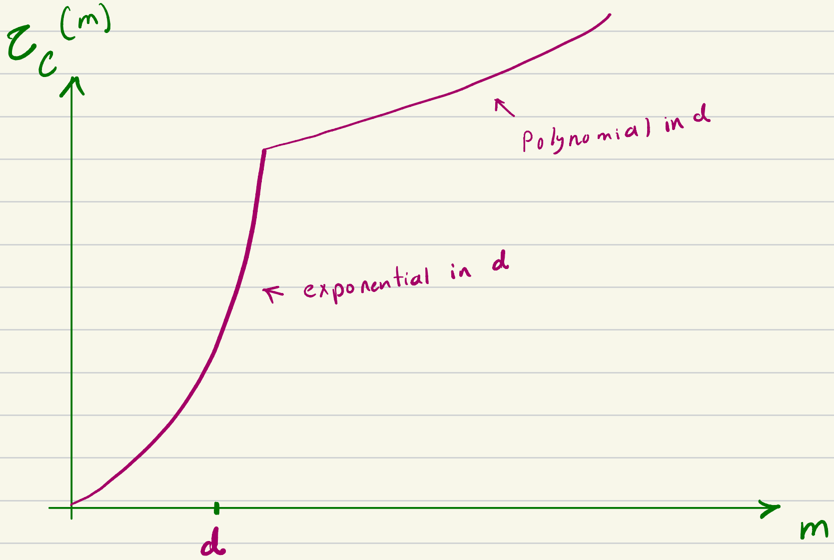
$\boxed{2}$ In particular, if $m \geq d+1$,

$$\tau_C(m) \leq \left(\frac{em}{d}\right)^d$$

why is this interesting?

- better than what we can naively imply from VC: for $m \geq d$ $\tau_C(m) < 2^m$

- As the number of samples increases the size of the restriction of $C$ to $S$ (the sample set) grows polynomially not exponentially $(2^{|S|})$.



$\tau_c^{(m)}$
Polynomial in d
exponential in d
d
m

# Proof of SSP

Here we focus on the proof of $\boxed{1}$

Part $\boxed{2}$ can be proven via part 1 and induction on d.

Proof. It suffices to show

i.e. $|C_T| = 2^{|T|}$

$$\ast \quad \forall S \quad |C_S| \le \left|\{T \subseteq S \mid C \text{ shatters } T\}\right|$$

$\emptyset$ is always shattered

By definition of VC dim. C does not shatter any set of size > d.

A set S has $\sum_{i=0}^{d} \binom{|S|}{i}$ subsets of size $\le$ d.

Hence, $\ast \Rightarrow \tau_C(m) \le \sum_{i=0}^{d} \binom{m}{i}$

Now, we focus on proving $\star$ by an inductive argument on the size of $S$: $|S| = m$.

Base case: $m = 1$

$S$ has one element $\rightsquigarrow$ $S$ has two subsets: $\emptyset, S$

two possible restriction: $(0)$, $(1)$

if $|C_S| = 2 \Rightarrow$ both $S$ and $\emptyset$ are shattered

$\star$ : $2 = 2$ ✓

if $|C_S| = 1 \Rightarrow$ $\emptyset$ is shattered

$S$ is not shattered

$\star$ : $1 = 1$ ✓

## inductive step

Assume $*$ holds for any set of size $<m$

we want to prove $*$ for $m$.

Consider $S = \{x_1, x_2, \ldots, x_m\}$

Let $S'$ denote $\{x_2, x_3, \ldots, x_m\}$.

$$Y_1 := \{ (y_2, y_3, \ldots, y_m) \mid$$

$$(0, y_2, \ldots, y_m) \in C_S \lor (1, y_2, \ldots, y_m) \in C_S \}$$

$$Y_0 = \{ (y_2, \ldots, y_m) \mid$$

$$(0, y_2, \ldots, y_m) \land (1, y_2, \ldots, y_m) \in C_S \}$$

Observe $|C_S| = |Y_0| + |Y_1|$

Now, we want to relate $|Y_0|$ and $|Y_1|$ to the # subsets that $C$ can shatter

By induction assumption:

$$|Y_1| = |C_{S'}| \leq |\{T \subseteq S' \mid C \text{ shatters } T\}|$$

$$= |\{T \subseteq S \mid x_1 \notin T \text{ and } C \text{ shatters } T\}|$$

---

$\forall \ (y_2, \ldots, y_m) \in Y_0$

$\exists$ a pair of concepts $c_1, c_2$ s.t

$$c_1(x_1) = 1, \quad c_1(x_2) = y_2, \ldots, c_1(x_m) = y_m$$

$$\underbrace{c_2(x_1) = 0}_{\text{differ only in } x_1}, \quad c_2(x_2) = y_2, \ldots, c_2(x_m) = y_m$$

Let $C'$ be the set of all of these pairs.

$$|Y_0| = |C'_{S'}| = |\{T \subseteq S' \mid C' \text{ shatters } T\}|$$

$C'$ can also shatters $T \cup \{x_1\}$

$$= |\{T \subseteq S \mid x_1 \in T \text{ and } C' \text{ shatters } T\}|$$

$$\leq |\{T \subseteq S \mid x_1 \in T \text{ and } C \text{ shatters } T\}|$$

$$|C_S| = \overline{|Y_0| + |Y_1|}$$

$$= |\{T \subseteq S \mid x_1 \in T \text{ and } C \text{ shatters } T\}|$$

$$+ |\{T \subseteq S \mid x_1 \notin T \text{ and } C \text{ shatters } T\}|$$

$$= |\{T \subseteq S \mid C \text{ shatters } T\}|$$

$\square$