Lecture 10

- Growth function
- Sauer - Shelah - Perles Lemma
- finite VC dim $\Rightarrow$ Uniform Convergence

Recall:

+ Uniform convergence. (UC)

Class C has the uniform convergence property if $\forall\ \varepsilon, \delta \in (0,1)$, dist D $\exists\ m$ (as a function of $\varepsilon, \delta, \mathcal{H}$, but not D since we don't know D). s.t. for a training set of size m:

$$\Pr_{T \sim D^m}\left[\forall c \in C : \left|\hat{err}_T(c) - err(c)\right| \leq \varepsilon\right] \geq 1-\delta$$

+ 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.

+ Restriction of C to S

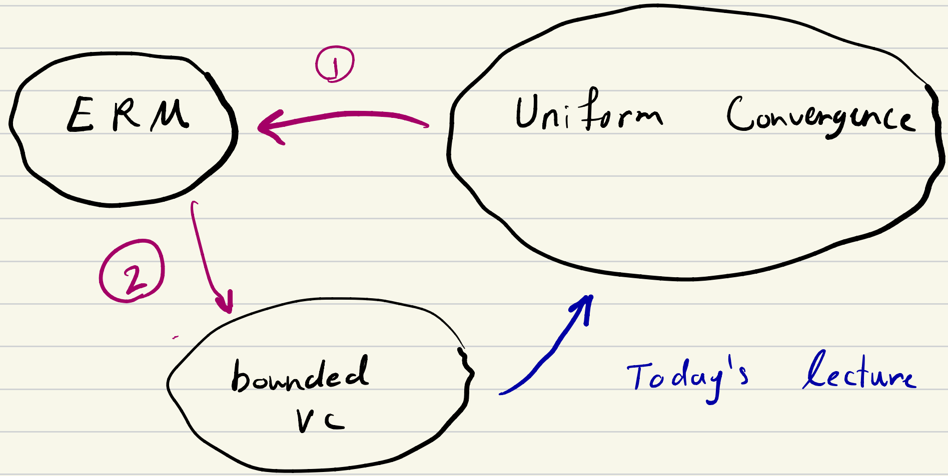
Let S be a set of m points in domain X. $S = \{x_1, \ldots, x_m\}$

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

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

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

Overview of today's lecture :

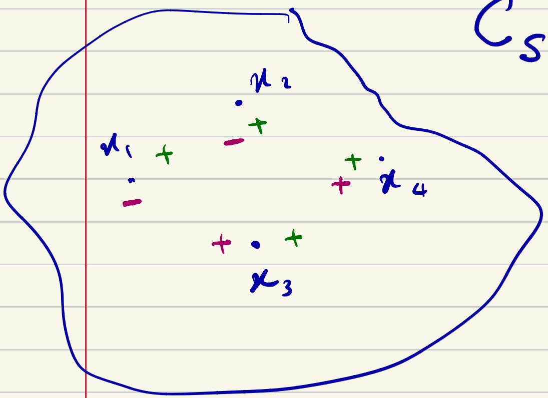


2 VC $\gg$ samples $\rightarrow$ ERM doesn't work

ERM work $\Rightarrow$ VC $<$ m
with m samples

growth function

$C = \{c_1, c_2\}$



$$C_S = \left\{ \begin{matrix} c_1 & (+,+,+,+) \\ c_2 & (-,-,+,+) \end{matrix} \right\}$$

if $|S| = m$

$\Rightarrow \quad |C_S| = 2^m$

Set S of size $m=4$

VC dim $(C) = 1$

C cannot shatter any two points

if $|S| = m$

$$\Rightarrow \quad |\mathcal{C}_S| = 2^m$$

$$\max_{\substack{S \subseteq X \\ |S| = m}} |C_S| \le \tau_S(m)$$

① Sauer's Lemma:

If $\quad VCdim(C) \le d$:

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

② $|S| = m$

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

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

+ 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 $\{0,1\}$ that can be obtained by $c \in C$.

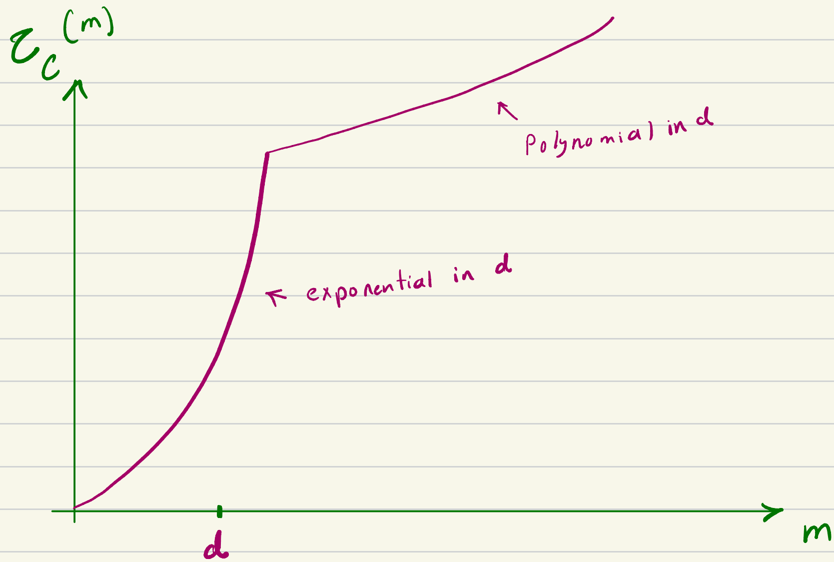
- With no assumption, we know $|C_S|$ is bounded by $2^{|S|} = 2^m$

Sauer's Lemma

Let C be a concept class with $\text{VCdim}(C) \leq d < \infty$. Then for all $m \in \mathbb{N}$, we have:

1. $\tau_C(m) \leq \sum_{i=1}^{d} \binom{m}{i}$

2. If $m > d+1 \Rightarrow \tau_C(m) \leq \left(\frac{em}{d}\right)^d$



Here we focus on the proof of part 1.

Part 2. can be proven via part 1 and induction on d.

Proof. It suffices to show

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

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

$\emptyset$ 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 < d.

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

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 siz $<m$

we want to prove * for m.

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

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

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

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

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

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

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

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

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

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 $T \cup \{x_1\}$

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

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

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

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

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

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

$\square$

Fundamental Theorem of PAC learning.

$$\boxed{\text{finite VC dim} \Rightarrow \text{Uniform Convergence}}$$

Realizable case:

$$O\left(\frac{\text{VC dim}(C)\ln(1/\varepsilon) + \ln(1/\delta)}{\varepsilon}\right) \text{ samples}$$

$\Rightarrow$ (ε, δ) - uniform convergence of C

Agnostic case:

$$O\left(\frac{\text{VC dim}(C) + \ln(1/\delta)}{\varepsilon^2}\right) \text{ samples}$$

$\Rightarrow$ (ε, δ) - uniform convergence of C

I this lecture, we prove an easier version.

We show: $d = \text{VCdim}(C)$

$$m = O\left(\frac{d}{(\delta\varepsilon)^2} \cdot \log \frac{d}{\varepsilon\delta} \right) \text{ samples}$$

$\Rightarrow (\varepsilon, \delta)$ - uniform convergence.

Lemma 1: Let C be a concept class with growth function $\tau_C(m)$. Then for every data distribution D, parameter $\delta \in (0,1]$, with probability at least $1-\delta$ over the choice of $S \sim D^m$, we have,

$$\forall c \in C : \left| err(c) - \hat{err}_S(c) \right| < \frac{4 + \sqrt{\log(\tau_C(2m))}}{\delta \sqrt{2m}}$$

**

for sufficiently large $m = O\left(\frac{d}{(\delta\varepsilon)^2} \log\left(\frac{d}{\varepsilon\delta}\right)\right)$

+ Sauer's Lemma

$$\tau_C(2m) \leq \left(\frac{2me}{d}\right)^d$$

$\Rightarrow$ right hand side of $**$ $\leq \varepsilon$

$\Rightarrow$ (ε, δ) Uniform convergence.

Proof of Lemma 1:

$$** \quad \forall c \in C : \left| err(c) - \hat{err}_S(c) \right| \leq \frac{4 + \sqrt{\tau_C(2m)}}{\delta \sqrt{2m}}$$

We show

$$\underset{S \sim D^m}{\mathbb{E}} \left[\sup_{c \in C} \left| err(c) - \hat{err}(c) \right| \right] \leq \frac{4 + \sqrt{\tau_C(2m)}}{\sqrt{2m}}$$

***

The above bound implies the lemma:

if $\mathbb{E}[X] \leq A$, then by Markov's ineq.

$$\Pr\left[X > \frac{1}{\delta} A \right] < \frac{\mathbb{E}[X]}{A/\delta} \leq \delta$$

$\Rightarrow$ Hence, with prob. $1-\delta$ $\quad X \leq \frac{A}{\delta}$

$$\mathop{\mathbb{E}}_{S\sim D^m}\left[\sup_{c\in C}\left|err(c)-\hat{err}_S(c)\right|\right]$$

$$err(c) = \mathbb{E}[\hat{err}_S(c)]$$

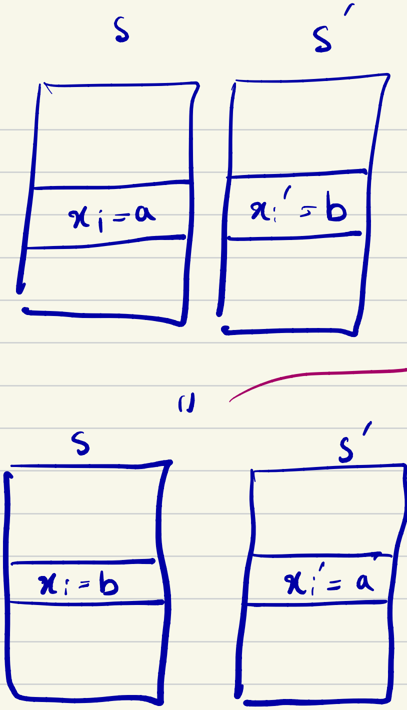
$$= \mathop{\mathbb{E}}_{S\sim D^m}\left[\sup_{c\in C}\left|\mathop{\mathbb{E}}_{S'\sim D^n}\left[\hat{err}_{S'}(c)\right]-\hat{err}_S(c)\right|\right]$$

Jensen's inequality : convex f

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(x)]$$

$$\leq \mathop{\mathbb{E}}_{S,S'\sim D^m}\left[\sup_{c\in C}\left|\hat{err}_{S'}(c)-\hat{err}_S(c)\right|\right]$$

$$= \mathop{\mathbb{E}}_{S,S'\sim D^m}\left[\sup_{c\in C}\frac{1}{m}\left|\sum_{i=1}^{m}\mathbb{1}(c(x_i')\neq y_i') - \mathbb{1}(c(x_i)\neq y_i)\right|\right]$$



same probability

Example $\quad E\left[x_i - x_i'^{\,2}\right] = \dfrac{a-b^{2}}{2} + \dfrac{b-a^{2}}{2}$

$$E\left[x_i' - x_i^{2}\right] = \frac{b-a^{2}}{2} + \frac{a-b^{2}}{2}$$

we can switch $x_i \longleftrightarrow x_i'$

$$\mathbb{1}(c(x_i) \neq y_i) - \mathbb{1}(c(x'_i) \neq y'_i)$$

$$\Updownarrow \quad \text{switch}$$

$$\mathbb{1}(c(x'_i) \neq y'_i) - \mathbb{1}(c(x_i) \neq y_i)$$

$$= -\left(\mathbb{1}(c(x_i) \neq y'_i) - \mathbb{1}(c(x'_i) \neq y'_i)\right)$$

$\forall \sigma = (\sigma_1, \dots, \sigma_m) \in \{+1, -1\}^m$

$$\sim \leq \mathbb{E}_{S,S' \sim D^m}\left[\sup_{c \in C} \frac{1}{m} \left| \sum_{i=1}^{m} \sigma_i \left(\mathbb{1}(c(x_i) \neq y_i) - \mathbb{1}(c(x'_i) \neq y'_i)\right)\right|\right]$$

$$\leq \mathbb{E}_{S,S'} \mathbb{E}_{\sigma \sim_U \{+1,-1\}^m}\left[\sup_{c \in C} \frac{1}{m} \left| \sum_{i=1}^{m} \sigma_i \left(\mathbb{1}(c(x_i) \neq y_i) - \mathbb{1}(c(x'_i) \neq y'_i)\right)\right|\right]$$

$C_{S \cup S'}$:

$$\left\{ (z_1, z_2, \ldots, z_m, z_1', \ldots, z_m') \;\middle|\; \exists c \in C : \forall i \in [m] \quad c(x_i) = z_i \text{ and } c(x_i') = z_i' \right\}$$

$$\leq \mathop{\mathbb{E}}_{S, S'} \mathop{\mathbb{E}}_{\sigma} \left[\sup_{Z \in C_{S \cup S'}} \frac{1}{m} \left| \sum_{i=1}^{m} \sigma_i \left(\mathbb{1}(z_i \neq y_i) - \mathbb{1}(z_i' \neq y_i') \right) \right| \right]$$

Fix S, S', and Z

Consider

$$A_Z(\sigma_i) = \sigma_i \left(\mathbb{1}(z_i \neq y_i) - \mathbb{1}(z_i' \neq y_i') \right)$$

source of randomness

$$\mathbb{E}_{\sigma_i \sim_U \{+1,-1\}} [A_Z (\sigma_i)] = 0$$

$$A_Z(\sigma_i) \in [-1, 1]$$

Hoeffding bound

$$\Pr_{\sigma} \left[\left| \frac{1}{m} \sum_{i=1}^{m} A_Z (\sigma_i) \right| > \rho \right]$$

$$\leq 2 \exp(-2m\rho^2)$$

Union Bound over $|C_{S \cup S'}|$ many

Z's:

$$\Pr_{\sigma} \left[\max_{Z \in C_{S \cup S'}} | A_Z (\sigma) | > \rho \right]$$

$$\leq 2 \, |C_{S \cup S'}| \cdot \exp(-2m\rho^2)$$

magic

$$\longrightarrow \mathbb{E}_{\sigma}\left[\max_{z \in C_{sus'}} |A_z(\sigma)|\right]$$

p

$$\int_0^\infty Pr[X \geq t]\,dt$$

$$\leq \frac{4 + \sqrt{\log |C_{sus'}|}}{\sqrt{2m}} \qquad \leq \tau(2m)$$

$$\Rightarrow \mathbb{E}_{S}\left[\sup_{c \in C} |err(c) - \hat{err}_S(c)|\right]$$

$$\leq \frac{4 + \sqrt{\log \tau_C(2m)}}{\sqrt{2m}}$$

$\Rightarrow$ ★★★ $\square$