Lecture 22

Last time

- Radamacher complexity
- Polynomial discrimination

Today

- Vapnik-Chervonenkis (VC) Theory.

VC Theory

We will develop another method to certify polynomial discrimination of $\{0,1\}$-valued functions $\mathcal{F}$.

Def: We say that

$$\mathcal{X} = \{x_1, \dots, x_n\}$$

is **shattered** by $\mathcal{F}$ if

$$\# \mathcal{F}(\mathcal{X}) = 2^{\# \mathcal{X}}$$

The **VC dimension** is

$$vc(\mathcal{F}) = \sup \{ n \in \mathbb{N} \mid \exists x \in \mathcal{X}^n \text{ shattered by } \mathcal{F} \}.$$

with

We will use the following notation, if S is a class of set and

$$\mathcal{F} = \{ \mathbb{1}_A \mid A \in S \},$$

then,

$$S(X) = \mathcal{F}(X) \quad \text{and} \quad vc(S) = vc(\mathcal{F}).$$

Example: Recall the last example from last class

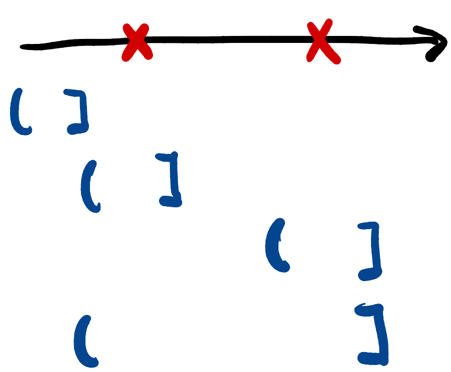
$$S_{one} = \{(-\infty, t] \mid t \in \mathbb{R}\}$$

then, we have that $vc(S_{one}) = 1$.

Similarly,

$$S_{two} = \{(a, b] \mid a, b \in \mathbb{R},\ a < b\}.$$

With $n=2$



$\rightarrow$ 00

$\rightarrow$ 10

$\rightarrow$ 01

$\rightarrow$ 11

But with $n=3$ we cannot form

101.

$\dashv$

Lemma (Sauer and Shelah): Suppose that $d = vc(S)$. Then for any $X = \{X_1, \ldots, X_n\}$,

$$\# S(X) \overset{(1)}{\leq} \underbrace{\sum_{i=0}^{d} \binom{n}{i}}_{\Phi_d(n)} \overset{(2)}{\leq} (n+1)^{vc(S)}.$$

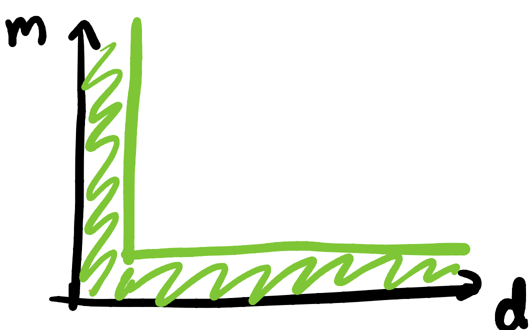
$\dashv$

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & 0 \leq k \leq n \\ 0 & \text{else.} \end{cases}$$

Thus, $\mathcal{F}$ has polynomial discrimination of order $vc(S)$ and by the results from last lecture we can bound

$$R_n(\mathcal{F}) \leq \sqrt{\frac{2\,vc(S)\log(n+1)}{n}}.$$

Proof: We prove (1) and leave (2) as an exercise. We will prove (1) using induction on $d+n$



Base case

Inductive step

Base case: Assume d arbitrary and $n=0$,

$$\#\mathcal{F}(\emptyset) = 0 \leq 1 = \sum_{i=0}^{d}\binom{0}{i} = \Phi_d(0)$$

$\hookrightarrow \binom{0}{0} = 1$

Assume n arbitrary $d=0$.

$$\#\mathcal{F}(X) = 1 \leq 1 = \binom{n}{0} = \Phi_0(n).$$

Inductive step: Fix d and m and suppose the inequality holds for any d', m' s.t. $d'+m' \leq d+m$. Let X be such that

$$X = \underset{|X|=m}{\operatorname{argmax}} \# \mathcal{F}(X).$$

Take any $a \in X$, and define

$$X^- = X \setminus \{a\}.$$

Label $H = \mathcal{F}(X) \subseteq \{0,1\}^{\# S}$, and define

$$H_0 = \{ h|_{X^-} \mid h \in H \text{ and } h(a) = 0 \},$$

$$H_1 = \{ h|_{X^-} \mid h \in H \text{ and } h(a) = 1 \}.$$

Define $H_\cap = H_0 \cap H_1$ these are binary strings that can be extended to X with both labels.

(Example

$$H = \left\{ \begin{array}{c} a \\ 010, \\ 000, \\ 011, \end{array} \right\} \Rightarrow \quad H_0 = \{01, 00\}, \quad H_1 = \{01\}$$

and $H_\cap = \{01\}$).

Therefore,

$$\# H = \#(H_0 \cup H_1) + \# H_\cap .$$

Convince yourself of this!

Inductive hypothesis

$$= \# \mathcal{F}(X^-) + \# H_\cap \qquad (\star)$$

$$\leq \Phi_d(n-1) + \# H_\cap .$$

Consider now

$$\tilde{\mathcal{F}} = \left\{ g : X^- \to \{0,1\} \;\middle|\; \begin{array}{l} \exists f_1, f_2 \in \mathcal{F} \text{ s.t. } f_1(a) \neq f_2(a) \\ \text{and } \forall x \in X^-,\ g(x) = f_1(x) = f(x). \end{array} \right\}$$

Then

$$H_\cap \subseteq \mathcal{F}_a(X^-).$$

We claim that $vc(\mathcal{F}_a) \leq d-1$. Suppose that it is not $\Rightarrow$ $\mathcal{F}_a$ can shatter some $Y \subseteq X^-$ and using the two extensions at a we would have

$$\# \mathcal{F}(Y \cup \{a\}) = 2^{d+1}$$

$\lightning$

Thus, we obtain

Inductive hypothesis

$$\# H_\cap \leq \# \mathcal{F}_a(X^-) \leq \Phi_{d-1}(n-1). \qquad (B)$$

Fact (Pascal's Identity):

$$\binom{n}{i} = \binom{n-1}{i} + \binom{n-1}{i-1} \qquad (☺)$$

Number of subsets of $[n]$ of size i

$= \quad \#\{S \subseteq [n-1] \mid \#S = i\}$

$\qquad + $

$\qquad \#\{S \cup \{n\} \mid S \subseteq [n-1],\ \#S = i-1\}.$

Then, (★), (♡) and (☺)

$$\#\mathcal{F}(X) \leq \sum_{i=0}^{d} \binom{n-1}{i} + \sum_{i=0}^{d-1} \binom{n-1}{i}$$

$$= \sum_{i=0}^{d} \binom{n-1}{i} + \binom{n-1}{i-1}$$

$$= \sum_{i=0}^{d} \binom{n}{i}.$$

$\square$