

## PS 2 due Thursday

Last time: - Finished probability bounds for large deviations

- PAC learning algorithms based on minimizing training error rate (over finite  $\mathcal{H}$ .)

- CH

- ERM

- Occam's razor - please read proof in lecture notes

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

Today: - Miscellaneous comments

- Uniform convergence

• VC dimension

• Sauer's lemma (if time)

## Miscellaneous comments

### ① Proper vs. improper learners

If a learning alg. for concept class  $\mathcal{C}$  actually returns a concept  $\hat{c} \in \mathcal{C}$ , then it is proper.

$$\mathcal{C} = 3\text{-term DNF}$$

$$\mathcal{C} \subseteq \mathcal{H} = 3\text{-CNF}$$

### ② ERM & agnostic learning

Thm  $S \sim P^T$  for some dist.  $P$  over  $\mathcal{X} \times \mathcal{Y}$ ,  
&  $\hat{h} = \arg\min_{h \in \mathcal{H}} \text{err}(h; S)$   
then w.p.  $\geq 1 - \delta$ ,

$$\text{err}(\hat{h}; P) \leq \min_{h \in \mathcal{H}} \text{err}(h; P) + \sqrt{\frac{2 \ln(2|\mathcal{H}|/\delta)}{T}}$$

- This makes no assumption on  $P$
- Guarantees "agnostic learning"
  - statistical analogue of "regret" from online learning.

## Uniform Convergence

$S \sim P^T$

To analyze ERM for hyp. class  $\mathcal{H}$ , we proved:

For any fixed  $h \in \mathcal{H}$ ,

w.p.  $\geq 1 - \delta$ ,

$$|\text{err}(h; S) - \text{err}(h; P)| \leq O\left(\sqrt{\frac{\ln \frac{1}{\delta}}{T}}\right)$$

Then by union bound,

w.p.  $\geq 1 - \cancel{|\mathcal{H}|} \delta$

$$\forall h \in \mathcal{H}, |\text{err}(h; S) - \text{err}(h; P)| \leq O\left(\sqrt{\frac{\ln(|\mathcal{H}|/\delta)}{T}}\right)$$

---

$$\mathcal{X} = [0, 1], \mathcal{Y} = \{0, 1\}, \mathcal{H} = \{h_V \mid V \subseteq [0, 1], \underbrace{|V|}_{\text{finite set}} < \infty\}$$

$$h_V(x) = \mathbb{1}\{x \in V\}$$

$P_X$  = uniform on  $[0, 1]$ ,  $Y = 0$  always.

$$\text{err}(h_V; P) = \Pr(h_V(Y) = 1) = \int_0^1 \mathbb{1}\{x \in V\} dx = 0$$

$\nearrow \{(x_i, y_i), \dots, (x_T, y_T)\}$

For any  $S \subset \mathcal{X} \times \mathcal{Y}$  finite

$$\exists h_V \in \mathcal{H} \text{ w/ } \text{err}(h_V; S) = 1$$

$$V = \{x_1, \dots, x_T\}$$

Q: Which  $\mathcal{H}$  have a uniform convergence guarantee for every  $P$ ?

A: given by Vapnik - Chervonenkis theory.

① Which function classes have finite approximation — empirical covers

② how to use empirical covers to establish uniform convergence.

— Uses ideas from Empirical Process Theory.

## VC-dimension

$\mathcal{F} \subseteq \{0,1\}^{\mathbb{Z}}$  function class of  $\{0,1\}$ -valued fns on domain  $\mathbb{Z}$   
 $T = \{z_1, \dots, z_m\}$  m pts from  $\mathbb{Z}$

$$\mathcal{F}[T] = \{(f(z_1), \dots, f(z_m)) \mid f \in \mathcal{F}\} \subseteq \{0,1\}^m$$

- Projection of  $\mathcal{F}$  onto  $T$

- Behaviors of  $\mathcal{F}$  on  $T$

Sometimes  $\mathcal{F}[T]$  is large,  
 " " " " small

$\mathcal{F}$  shatters  $T$  if  $\mathcal{F}[T] = \{0,1\}^m$

-  $\forall (b_1, \dots, b_m) \in \{0,1\}^m, \exists f \in \mathcal{F}$  s.t.  $f(z_i) = b_i \forall i \in [m]$ .

$\text{VCdim}(\mathcal{F}) =$  largest  $d$  s.t.  $\exists T \subseteq \mathbb{Z}, |T|=d$

&  $T$  shattered by  $\mathcal{F}$ .

( $\infty$  if arbitrarily large  $T$  are shattered by  $\mathcal{F}$ .)

Lemma If  $|\mathcal{F}| < \infty$ ,  $\text{VCdim}(\mathcal{F}) \leq \log_2 |\mathcal{F}|$ .

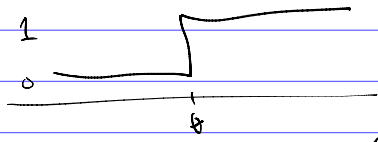
Pf:  $T = \{z_1, \dots, z_m\}$  is shattered by  $\mathcal{F}$ .

$$|\mathcal{F}| \geq |\mathcal{F}[T]| = 2^m$$

$$\log_2 |\mathcal{F}| \geq m. \quad \square$$

$$\text{LTF}_{\mathbb{R}^n} = \{h_{w,\theta} \mid w \in \mathbb{R}^n, \theta \in \mathbb{R}\}$$

$$h_{w,\theta}: \mathbb{R}^n \rightarrow \{0,1\}, \quad h_{w,\theta}(x) = \begin{cases} 1 & \text{if } \langle w, x \rangle > \theta \\ 0 & \text{o.w.} \end{cases} = \mathbb{1}\{\langle w, x \rangle > \theta\}$$



$$\text{Claim: } \log_2 |\text{LTF}_n| = \Theta(n^2) = \text{VCdim}(\text{LTF}_n)$$

$$\text{Claim: } \text{VCdim}(\text{LTF}_{\mathbb{R}^n}) = n+1$$

$$\text{VCdim}(\text{LTF}_n) = n+1$$

$$\textcircled{1} \text{ VCdim}(\text{LTF}_{\mathbb{R}^n}) \geq n+1$$

$$T = \{0, e_1, \dots, e_n\}$$

$$b_0 \quad b_1$$

$$b_n$$

$$e \in \{0,1\}^?$$

$$\tilde{b}_0 \quad \tilde{b}_1$$

$$\tilde{b}_n$$

$$\tilde{b}_i = 2b_i - 1 \in \{-1,1\}$$

$$\langle w, 0 \rangle > \theta \iff \tilde{b}_0 = 1$$

$$w = \langle w, e_1 \rangle e_1 + \dots + \langle w, e_n \rangle e_n$$

$$-\theta > 0$$

$$\text{set } \theta = -\tilde{b}_0$$

$$\langle w, e_i \rangle > \theta \iff \tilde{b}_i = 1$$

$$\text{"Set"} \quad \langle w, e_i \rangle = 2\tilde{b}_i$$

$$2\tilde{b}_i + \tilde{b}_0 > 0 \iff \tilde{b}_i = 1 \quad \checkmark$$

$$\textcircled{2} \text{ VCdim}(\text{LTF}_{\mathbb{R}^n}) \leq n+2.$$

Take any  $T \subset \mathbb{R}^n$ .  $|T| = n+2$

and show it's not shattered by  $\text{LTF}_{\mathbb{R}^n}$ .

Fact: If  $T$  is shattered by  $\text{LTF}_{\mathbb{R}^n}$ ,

then  $\tilde{T} = \{(x, -1) \mid x \in T\}$  is shattered by  $\text{HLTF}_{\mathbb{R}^{n+1}}$

Take any  $T \subset \mathbb{R}^{n+1}$ .  $|T| = n+2$

and show it's not shattered by  $\text{HLTF}_{\mathbb{R}^{n+1}}$

Take  $x_1, \dots, x_{n+2} \in \mathbb{R}^{n+1}$

They are linearly dependent, so  $\exists c_1, \dots, c_{n+2}$  not all 0

s.t.

$$\sum_{i=1}^{n+2} c_i x_i = 0$$

Assume WLOG  $c_i > 0$

Consider the labeling  $(b_1, \dots, b_{n+2}) \in \{0,1\}^{n+2}$

$$b_i = \mathbb{1}\{c_i > 0\}$$

for sake of contradiction

Suppose there's HLTF given by  $w \in \mathbb{R}^{n+1}$  s.t.

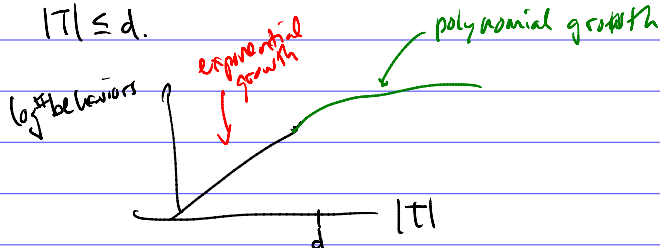
$$\langle w, x_i \rangle > 0 \iff b_i = 1 \iff c_i > 0.$$

$$0 = \left\langle w, \sum_{i=1}^{n+2} c_i x_i \right\rangle = \sum_{i=1}^{n+2} c_i \langle w, x_i \rangle > 0 \quad \perp$$

If  $\mathcal{F}$  has  $\text{VCdim}(\mathcal{F}) = d$

then  $\exists$  sets  $T$  s.t.  $\mathcal{F}[T] =$  all behaviors on  $T$

for  $|T| \leq d$ .



For  $|T| \geq d$ ,  $\# \text{behaviors} \leq O(n^d) \leftarrow$  Sauer's lemma.