

Peer Eval for PS2 due Tue 10/27 in lecture

PS3 due Thu 10/29

Midterm Thu 11/5 in lecture

Last time:

- Sauer's lemma: for $\mathcal{F} \subseteq \{0,1\}^{\mathbb{Z}}$ w/ $d = \text{VC dim}(\mathcal{F}) < \infty$
and any $z_1, \dots, z_n \in \mathbb{Z}$,
$$|\mathcal{F}[z_{1:n}]| \leq \binom{n}{\leq d}$$

of behaviors on n pts is only $\text{poly}(n)$
rather than 2^n

Today: Prove uniform convergence for $\mathcal{F}$ w/ finite VC dim
(Vapnik & Chervonenkis, 1971)

Exercise: Suppose you want to show uniform convergence
of training error rates of $h \in \mathcal{H} \subseteq \{0,1\}^{\mathbb{Z}^k}$ (w/ iid training data)
to true error rates.

What $\mathcal{F} \subseteq \{0,1\}^{\mathbb{Z}}$ should we apply VC theorem to?

Uniform convergence of empirical averages to population averages for a function class $\mathcal{F} \subseteq \{0,1\}^{\mathcal{Z}}$

- P : distribution over $\mathcal{Z}$ (population)
- $\mathbf{z}_{1:n} = (z_1, \dots, z_n) \sim P^n$ size- n iid sample from P
- $\mathcal{F} \subseteq \{0,1\}^{\mathcal{Z}}$ function class
- $\frac{1}{n} \sum_{i=1}^n f(z_i)$ empirical avg of f
- $E[f(z)]$ where $z \sim P$ population avg of f

Thm If $d = \text{VCdim}(\mathcal{F}) < +\infty$, then

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \right] \leq \underbrace{\sqrt{\frac{2 \ln(2 \cdot \binom{2n}{\leq d})}{n}}}_{O\left(\sqrt{\frac{d \log n}{n}}\right)}$$

Thm (same setup) for any $\delta \in (0,1)$,
w.p. $\geq 1 - \delta$

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \leq 100 \sqrt{\frac{2 \ln(2 \cdot \binom{2n}{\leq d} / \delta)}{n}}$$

If $|\mathcal{F}| < \infty$, then we used union bound to show
w.p. $\geq 1 - \delta$

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \leq 100 \sqrt{\frac{2 \ln(2 \cdot |\mathcal{F}| / \delta)}{n}}$$

Thm If $d = VC \dim(F) < +\infty$, then

$$E \left[\sup_{f \in F} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \right] \leq \sqrt{\frac{2 \ln(2 \cdot \binom{2n}{\leq d})}{n}}$$

Important issues

① $\left| \left\{ \frac{1}{n} \sum_{i=1}^n f(z_i) \mid f \in F \right\} \right| \leq \binom{n}{\leq d}$ by Sauer's lemma

BUT

~~$\left| \{ E[f(z)] \mid f \in F \} \right|$ could be ∞~~

② $F[z_1, \dots, z_n] \subseteq \{0,1\}^n$ is a random subset of $\{0,1\}^n$

How do we understand its distribution???

Two tricks

① "Symmetrization": pass through symmetric r.v.'s only

② "conditioning" + "symmetrization"

Trick 1: Symmetrization

x, x' are iid real-valued r.v.'s

$$\frac{1}{2} E[(x - x')^2] = \frac{1}{2} E[(x - E(x) + E(x') - x')^2]$$

$$= \frac{1}{2} (E[(x - E(x))^2] + \text{crossterm} + E[(x' - E(x'))^2])$$

$$= \text{var}(x)$$

$$E[|x - E(x)|] \leq E[|x - x'|]$$

$$E[x] = E[x' | x]$$

$$E[|x - E(x)|] = E[|x - E[x' | x]|]$$

$$\leq E[|x - x'|] \quad \text{by Jensen's inequality}$$

Proposition Let $z_{1:n}$ & $z'_{1:n}$ be independent iid samples from $\mathcal{P}$ (both of size n)

"ghost sample"
empirical avg on ghost sample

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \right] \leq E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - \frac{1}{n} \sum_{i=1}^n f(z'_i) \right| \right]$$

Pf $\Delta f := \frac{1}{n} \sum_{i=1}^n f(z_i) - \frac{1}{n} \sum_{i=1}^n f(z'_i)$

$$E[f(z)] = E \left[\frac{1}{n} \sum_{i=1}^n f(z_i) \right] = E \left[\frac{1}{n} \sum_{i=1}^n f(z'_i) \mid z_{1:n} \right]$$

$$\left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| = \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E \left[\frac{1}{n} \sum_{i=1}^n f(z'_i) \mid z_{1:n} \right] \right|$$

$$= \left| E \left[\frac{1}{n} \sum_{i=1}^n f(z_i) - \frac{1}{n} \sum_{i=1}^n f(z'_i) \mid z_{1:n} \right] \right|$$

$$= \left| E[\Delta f \mid z_{1:n}] \right|$$

$$\leq E[|\Delta f| \mid z_{1:n}] \quad (\text{Jensen})$$

$$\leq E \left[\sup_{f \in \mathcal{F}} |\Delta f'| \mid z_{1:n} \right]$$

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \right] \leq E \left[E \left[\sup_{f \in \mathcal{F}} |\Delta f'| \mid z_{1:n} \right] \right]$$

Proposition let $z_{1:n}$ & $z'_{1:n}$ be independent iid samples from P
(both of size n)

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \right] \leq E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - \frac{1}{n} \sum_{i=1}^n f(z'_i) \right| \right]$$

size $- 2n$ iid sample from P

$$(z_1^a, z_1^b, z_2^a, z_2^b, \dots, z_n^a, z_n^b) \sim P^{2n}$$

Choice for assigning z_i^a & z_i^b to real & ghost
given by $\sigma_i \in \{-1, 1\}$

If $\sigma_i = 1$, then $z_i := z_i^a$

$$z'_i = z_i^b$$

If $\sigma_i = -1$, then $z_i := z_i^b$

$$z'_i = z_i^a$$

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n f(z_i) - \frac{1}{n} \sum_{i=1}^n f(z'_i) &= \frac{1}{n} \sum_{i=1}^n (f(z_i) - f(z'_i)) \\ &= \frac{1}{n} \sum_{i=1}^n \sigma_i (f(z_i^a) - f(z_i^b)) \end{aligned}$$

for every $\sigma_{1:n} \in \{-1, 1\}^n$

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \right] \leq E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i (f(z_i^a) - f(z_i^b)) \right| \right]$$

Think of $\sigma_{1:n} \sim \text{Unif}(\{-1, 1\}^n)$ and regarded as indep. of $z_{1:n}^a, z_{1:n}^b$

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \right] \leq E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i (f(z_i^a) - f(z_i^b)) \right| \right]$$

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(z_i) - E[f(z)] \right| \right] \leq E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i (f(z_i^a) - f(z_i^b)) \right| \right]$$

Conditioning

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i (f(z_i^a) - f(z_i^b)) \right| \mid z_{1:n}^a, z_{1:n}^b \right]$$

$$\mathcal{F}[z_{1:n}^a, z_{1:n}^b] \subseteq \{0, 1\}^{z^n} \quad \|v\| \leq |\mathcal{F}[z_{1:n}^a, z_{1:n}^b]| \leq \binom{n}{\leq d}$$

↓

$$V := \left\{ (f(z_{1:n}^a) - f(z_{1:n}^b), \dots, f(z_{1:n}^a) - f(z_{1:n}^b)) \mid f \in \mathcal{F} \right\} \subseteq \{-1, 0, 1\}^n \quad \leftarrow$$

So

$$E \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i (f(z_i^a) - f(z_i^b)) \right| \mid z_{1:n}^a, z_{1:n}^b \right] \\ = E \left[\sup_{v \in V} \left| \frac{1}{n} \langle \sigma, v \rangle \right| \mid z_{1:n}^a, z_{1:n}^b \right]$$

for a fixed choice $v \in V$

$$\langle \sigma, v \rangle = \sum_{i=1}^n \sigma_i v_i \sim \|v\|^2 \text{-subgaussian}$$

$$K_{\langle \sigma, v \rangle}(\lambda) \leq \frac{\|v\|^2}{2} \cdot \lambda^2$$

$$\Pr(|\langle \sigma, v \rangle| > t) \leq 2 \exp\left(-\frac{t^2}{2\|v\|_2^2}\right)$$

$$\Pr(\exists v \in V \text{ s.t. } |\langle \sigma, v \rangle| > t) \leq 2|V| \exp\left(-\frac{t^2}{2\|v\|_2^2}\right)$$

So every $v \in V$ has $\|v\|_2^2 \leq n$
w.p. $\geq 1 - \delta$

$$\max_{v \in V} \left| \frac{1}{n} \langle \sigma, v \rangle \right| \leq \sqrt{\frac{2 \ln(2|V|/\delta)}{n}} \quad \left| \begin{array}{l} \text{oop not} \\ \text{what we} \\ \text{wanted} \end{array} \right.$$

Massart's finite class lemma: for any $V \subseteq \mathbb{R}^n$ w/ $\|v\|_2 \leq \sqrt{n} \quad \forall v \in V$

$$E \left[\max_{v \in V} \left| \frac{1}{n} \langle \sigma, v \rangle \right| \right] \leq \sqrt{\frac{2 \ln(2|V|)}{n}}$$