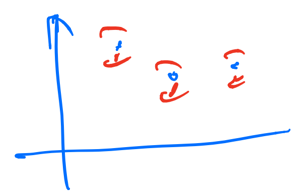
Problem 3.3. $V \geq 2$.

Math operator max min Sin

$\max$

Theorem 1. suppose $\mathcal{H}$. $VC(\mathcal{H}) = d$. Then, given a set of n i.i.d. training examples $(x_1, y_1) \ldots (x_n, y_n)$ from D, with prob. $1-\delta$,



$$\sup_{h \in \mathcal{H}} \left| err(h, S) - err(h, D) \right| \leq c_1 \cdot \sqrt{\frac{d \ln \frac{n}{d} + \ln \frac{1}{\delta}}{n}}$$

for some constant c_1.

(uniform convergence ;
uniform law of large numbers)

Thm 2 : suppose $S = (z_1 \cdots z_n)$ drawn iid from a distribution D. $\mathcal{F} \subseteq (Z \to \{0,1\})$ is a class of functions. Then, w.p. $1-\delta$,

$$\sup_{f \in \mathcal{F}} \left| \mathbb{E}_S f(z) - \mathbb{E}_D f(z) \right| \le \sqrt{\frac{32 \left(\ln\frac{4}{\delta} + \ln\left(S(\mathcal{F}, n) \right) \right)}{n}} .$$

$$\mathbb{E}_S f(z) = \frac{1}{n} \sum_{i=1}^{n} f(z_i)$$

$$S(\mathcal{F}, n) = \max_{z_1 \cdots z_n} \left| \{ (f(z_1), \cdots f(z_n)) : f \in \mathcal{F} \} \right|$$

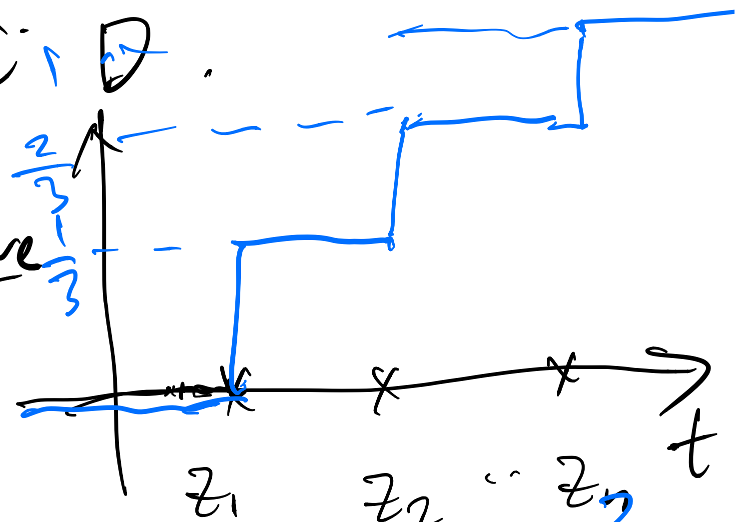
Example of Thm 2 :

Glivenko-Cantelli theorem:

D: distribution over $\mathbb{R}$

$Z_1 \ldots Z_n \overset{iid}{\sim} D$.

Empirical cumulative distribution fn (CDF):



$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} I(Z_i \le t)$$

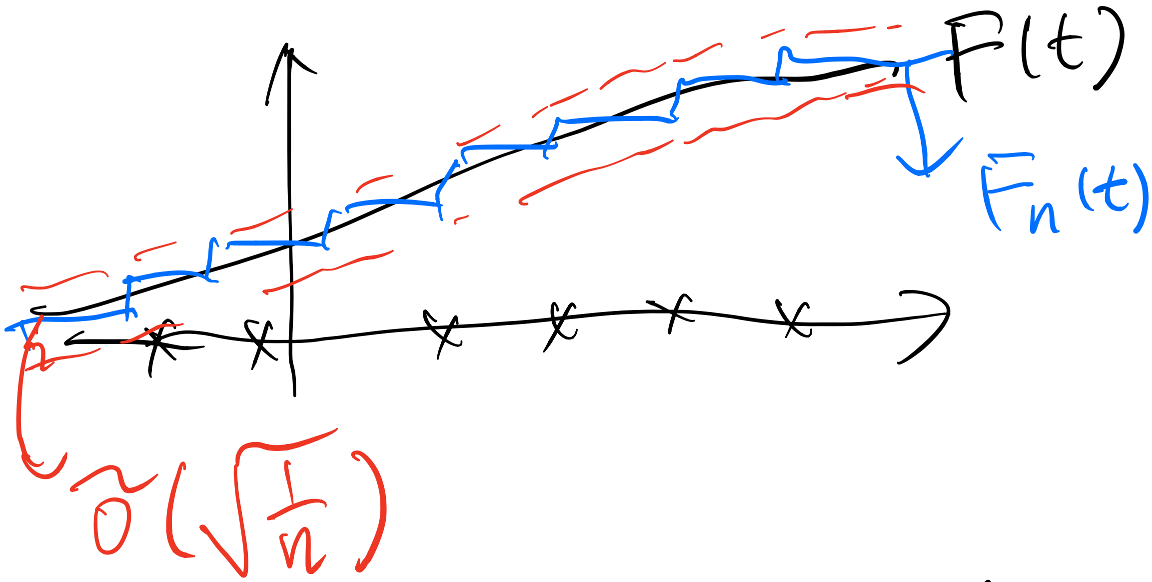
$n=3$

Population CDF:

$$F(t) = P_{Z \sim D}(Z \le t)$$

Thm 2 $\Rightarrow$ $\sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$

$$\le O\left(\sqrt{\frac{\ln\frac{1}{\delta}+\ln 4}{n}}\right)$$



$$\mathcal{F} = \{ \mathbb{I}(x \le t) : t \in \mathbb{R} \}$$

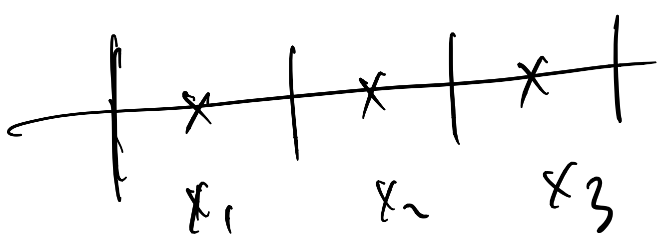
Apply thm 2 : w.p. $1-\delta$,

$$\sup_{f \in \mathcal{F}} \left| \mathbb{E}_S f(z) - \mathbb{E}_D f(z) \right|$$

$$\le O\left(\sqrt{\frac{\ln\frac{1}{\delta} + \ln S(\mathcal{F}, n)}{n}} \right)$$

$$\text{LHS} = \sup_{t\in\mathbb{R}} \Big| \underbrace{\mathbb{E}_S\, h_t(z)}_{\substack{\| \\ \frac{1}{n}\sum_{i=1}^{n} I(z_i \le t) \\ \| \\ F_n(t)}} - \underbrace{\mathbb{E}_D\, h_t(z)}_{\substack{\| \\ F(t)}} \Big|$$

$$\text{RHS}: \quad S(\mathcal{F}, n) \overset{?}{=} n+1.$$



$$\Rightarrow \sup_{t\in\mathbb{R}} \big| F_n(t) - F(t) \big| \le \tilde{O}\left(\sqrt{\frac{1}{n}}\right) \quad \text{w.h.p.}$$

Thm 2 $\Rightarrow$ Thm 1 ?

$\forall i,\ z_i = (x_i, y_i)$

0-1 loss function class induced by $\mathcal{H}$

$$\mathcal{F} = \{ l_h : h \in \mathcal{H} \} \quad \Big| \quad l_h(x,y) \doteq I(h(x) \neq y)$$

$$l_h \in \mathcal{F} \longleftrightarrow h \in \mathcal{H}$$

$$\mathbb{E}_S\, l_h(z) = err(h, S)$$

$$\mathbb{E}_D\, l_h(z) = err(h, D).$$

Thm 2 $\stackrel{?}{\Rightarrow}$ w.p. $1-\delta$:

$$\sup_{l_h \in \mathcal{F}} \left| \mathbb{E}_S\, l_h(z) - \mathbb{E}_D\, l_h(z) \right|$$

$$\leq \sqrt{\frac{32\left(\ln\frac{4}{\delta}+\ln(S(\mathcal{F},n))\right)}{n}}$$

$$S(\mathcal{H},n) \leq \binom{n}{\leq d} \leq \left(\frac{en}{d}\right)^d$$

$$LHS = \sup_{h\in\mathcal{H}} \left| err(h,S) - err(h,D) \right|$$

RHS: bound $S(\mathcal{F},n)$ in terms of $VC(\mathcal{H}) = d$?

$$S(\mathcal{F},n) = S(\mathcal{H},n)$$

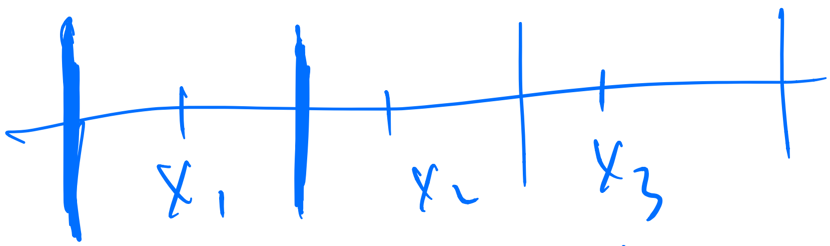
$$S(\mathcal{F},n) = \max_{z_1\ldots z_n} \left| \{ (l_h(z_1)\ldots l_h(z_n)) : h\in\mathcal{H} \} \right|$$

correct / incorrect

$$= \max_{(x_1,y_1)\ldots(x_n,y_n)} \left| \{ (\underline{I(h(x_1)\neq y_1)} \ldots I(h(x_n)\neq y_n)) : h\in\mathcal{H} \} \right|$$

+ | − || 0 ✓

4 $\quad \left|\left\{\left(h(x_1), h(x_2) \dots h(x_n)\right) : h \in \mathcal{H}\right\}\right|$



$+ \qquad - \qquad +$

$1 \quad 0 \quad 1 \longleftrightarrow + \quad + \quad +$

$0 \quad 0 \quad 1 \longleftrightarrow - \quad + \quad +$

$$= \max_{x_1 \dots x_n} \left|\left\{\left(h(x_1) \dots h(x_n)\right) : h \in \mathcal{H}\right\}\right|$$

$$= S(\mathcal{H}, n).$$

Apply Sauer's Lemma on the RHS concludes the pf of thm. 1.

<u>pf of Thm 2</u>:

We only prove that w.p. $1-\delta/2$,

$$(*) \quad \sup_{f \in \mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z) \le \sqrt{\frac{32(\ln 4/\delta + \ln S(\mathcal{F}, n))}{n}}$$

Can show the lower concentration $\mathbb{E}_S f(z)$ · $\mathbb{E}_D f(z)$ by similar techniques.

taking union bound gives thm 2.

<u>Lemma 1</u>: w.p. $1-\delta/2$:

$$\sup_{f \in \mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z) \le \mathbb{E}\left[\sup_{f \in \mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)\right] + \sqrt{\frac{\ln 4/\delta}{2n}}$$

Lemma 2: $\mathbb{E}\left[\sup_{f\in\mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)\right]$

$$\le 2\cdot\frac{1}{n}\cdot \mathop{\mathbb{E}}_{S\sim D^n}\ \mathop{\mathbb{E}}_{\sigma\sim U(\pm1)^n} \sup_{f\in\mathcal{F}} \sum_{i=1}^n f(z_i)\sigma_i$$

$$= 2\cdot \mathrm{Rad}_n(\mathcal{F})$$

$$\mathrm{Rad}_S(\mathcal{F}) = \frac{1}{n}\mathop{\mathbb{E}}_{\sigma\sim U(\pm1)^n} \sup_{f\in\mathcal{F}} \sum_{i=1}^n f(z_i)\sigma_i :$$

Rademacher distribution

empirical Rademacher complexity of $\mathcal{F}$.

$$\mathrm{Rad}_n(\mathcal{F}) = \mathop{\mathbb{E}}_{S\sim D^n} \mathrm{Rad}_S(\mathcal{F}).$$

Ex: $\mathcal{F}=\{f\}$. $\mathrm{Rad}_S(\mathcal{F}) = 0$.

$\mathcal{F}=\{$all fns from X to $\{\pm1\}\}$.

$$\mathrm{Rad}_S(\mathcal{F}) = 1$$

?

$\mathcal{F} = \{$ all fn from X to $\{0,1\}\}$.

$$\mathrm{Rad}_S(\mathcal{F}) = \frac{1}{2}$$

Lemma 3 : controlling Rademacher complexity w/ growth fn.

for any set S of size n,

$$\mathrm{Rad}_S(\mathcal{F}) \leq \sqrt{\frac{2\ln(S(\mathcal{F},n))}{n}}$$

$$S(\mathcal{F},n) \leq n^d \qquad \sqrt{\frac{d\ln n}{n}}$$

Lemmas 1–3 $\Rightarrow$ Thm 2 .

$$\sup_{f\in\mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)$$

$$\leq \sqrt{\frac{\ln 4/\delta}{2n}} + 2\cdot\sqrt{\frac{2\ln S(\mathcal{F},n)}{n}}$$

$$\left(\forall A, B.\ \sqrt{A} + \sqrt{B} \le \sqrt{2(A+B)} \right)$$

$$\le \sqrt{2\left(\frac{\ln 4/\delta}{2n} + \frac{8 \ln S(\mathcal{F}, n)}{n} \right)}$$

Lemma 1. w.p. $1-\delta/2$:

$$\underbrace{\boxed{\sup_{f\in\mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z)}}_{g(z_1 \ldots z_n) = \frac{1}{n}\text{-sensitive}} \le \mathbb{E}\left[\sup_{f\in\mathcal{F}} \mathbb{E}_S f(z) - \mathbb{E}_D f(z) \right] + \sqrt{\frac{\ln 4/\delta}{2n}}.$$

Pf: McDiarmid's Inequality:
(bounded difference ineq.)

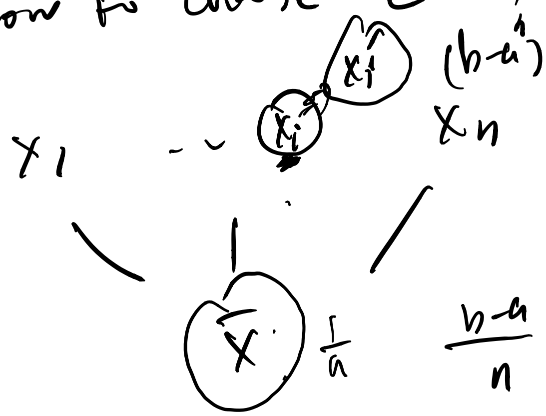
sensitivity: $f: V^n \to \mathbb{R}$ is c-sensitive, if for every $i \in \{1 \ldots n\}$, $x_1 \ldots x_n$, x_i'

$$\left| f(x_1 \ldots x_n) - f(x_1 \ldots x_i' \ldots x_n) \right| \le c.$$

Ex: $f(x_1 \dots x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$

$V = [a, b]$.

how to choose c ? $c = \frac{b-a}{n}$.



McDiarmid's Lemma:

f is c-sensitive. $x_1 \dots x_n$ are iid. D supported on V. Then,

$$P\left(\left| f(x_1 \dots x_n) - \mathbb{E} f(x_1 \dots x_n) \right| \geq \varepsilon \right)$$

$$\leq 2 \cdot e^{-\frac{2\varepsilon^2}{nc^2}}$$

equivalently. w.p. $1-\delta'$.

$$\left| f(X_1 \dots X_n) - \mathbb{E} f(X_1 \dots X_n) \right|$$

$$\leq C \cdot \frac{1}{n} \sqrt{\frac{n}{2} \cdot \ln \frac{2}{\delta'}}$$

McDiarmid $\Rightarrow$ Hoeffding's Ineq ? $\checkmark$