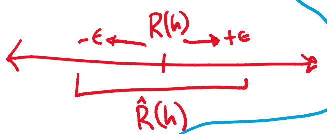
Section B

Monday, March 13, 2023 8:43 AM

Def: PAC Criterion

$$\Pr\left(\forall h \in \mathcal{H},\ |R(h) - \hat{R}(h)| < \epsilon\right) \geq (1-\delta)$$

ϵ ← small δ ← small



Q: What is 'random' here?

A: $\hat{R}(h)$ is based on a random sample of training data from $p^*(x)$

Def: sample complexity is the minimum number of training examples N s.t. the PAC criterion is satisfied for some ϵ and δ

Def: a hypothesis $h \in \mathcal{H}$ is consistent with the training data if $\hat{R}(h) = 0$

Thm 1: Realizable, Finite $|\mathcal{H}|$

$N \geq \frac{1}{\epsilon}\left[\ln(|\mathcal{H}|) + \ln(1/\delta)\right]$ labeled examples are sufficient to ensure that w/prob. $(1-\delta)$ all $h \in \mathcal{H}$ with $\hat{R}(h) = 0$ have $R(h) < \epsilon$

Proof of Thm. 1

1. Assume k bad hypotheses $h_1, h_2, \ldots, h_k$ with $R(h_i) > \epsilon$
2. Pick bad h_i: prob. of h_i consistent w/first training example $\leq (1-\epsilon)$ — $x^{(1)} \sim p^*(\cdot)$
   prob. of h_i consistent w/first N training examples $\leq (1-\epsilon)^N = \underbrace{(1-\epsilon)\cdots(1-\epsilon)}_{N \text{ times}}$
   → $\hat{R}(h) = 0$
3. Prob. that at least one bad h_i is consistent w/first N train examples $\leq k(1-\epsilon)^N$
   $\exists h \in \mathcal{H}$ s.t. $\hat{R}(h) = 0$ and $R(h) > \epsilon$
   by Union Bound: $P(A \cup B) \leq P(A) + P(B)$
   $\leq |\mathcal{H}|(1-\epsilon)^N$
   Q: which is larger? A: $k \leq |\mathcal{H}|$
4. Fact: $(1-x) \leq \exp(-x)$
   $\Rightarrow |\mathcal{H}|(1-\epsilon)^N \leq |\mathcal{H}| \exp(-\epsilon)^N = |\mathcal{H}| \exp(-\epsilon N)$
5. Choose δ s.t. $|\mathcal{H}| \exp(-\epsilon N) \leq \delta$
6. Solve for N: $\Rightarrow |\mathcal{H}|(1/\delta) \leq 1/\exp(-\epsilon N)$
   $\Rightarrow \ln(|\mathcal{H}|) + \ln(1/\delta) \leq \ln(1) - \underbrace{\ln(\exp(-\epsilon N))}_{\epsilon N}$
   Statement 1: $\Rightarrow \frac{1}{\epsilon}\left[\ln(|\mathcal{H}|) + \ln(1/\delta)\right] \leq N$

Assume Stat. 1 then:

$\epsilon[\ldots] \leq N$

Assume Stat. 1, then:

w/prob δ

$\exists h \in \mathcal{H}$ s.t. $R(h) > \epsilon$ and $\hat{R}(h) = 0$] "bad"

w/prob. $(1-\delta)$

all $h \in \mathcal{H}$ with $\underbrace{\boxed{R(h) > \epsilon}}_{A}$ have $\underbrace{\boxed{\hat{R}(h) > 0}}_{B}$] "not bad"

$\Rightarrow$ all $h \in \mathcal{H}$ with $\underbrace{\boxed{\hat{R}(h) = 0}}_{\neg B}$ have $\underbrace{\boxed{R(h) \leq \epsilon}}_{\neg A}$

if the learner returns $h \in \mathcal{H}$ w/zero training error then w/prob. $(1-\delta)$ we know that h has $\leq \epsilon$ true error

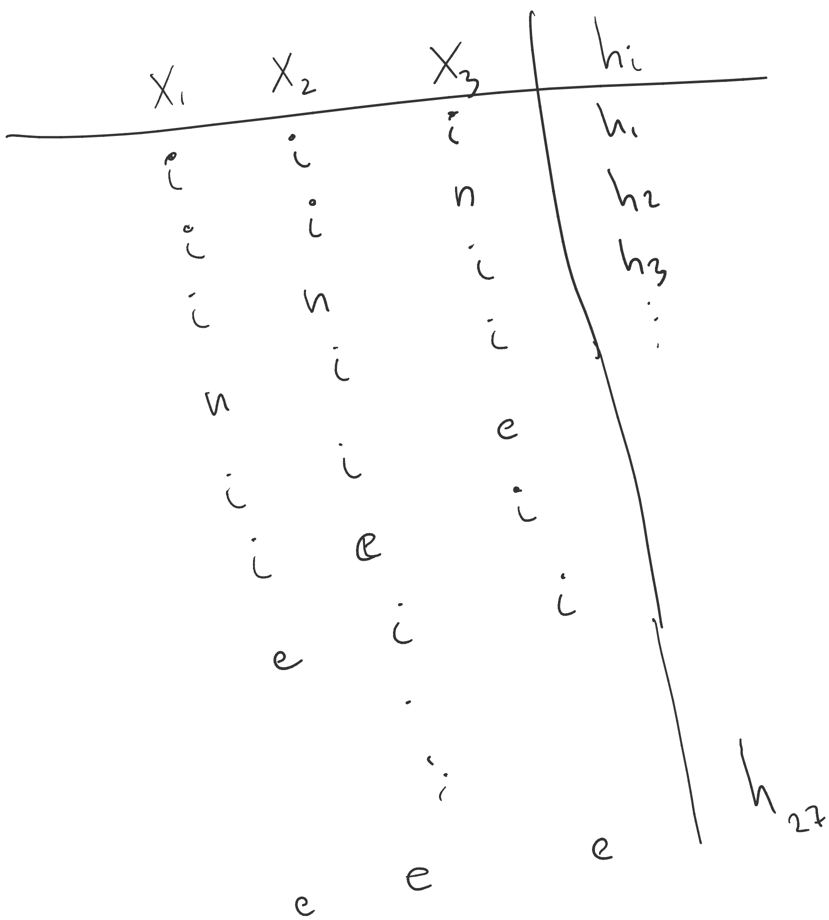
Contrapositive

$A \Rightarrow B$

equiv.

$\neg B \Rightarrow \neg A$

OH



$l \in \mathbb{R}^k$

$\vec{b} = \beta \vec{z}$

$b \in \mathbb{R}^k$

$z \in \mathbb{R}^D$

$\beta \in \mathbb{R}^{k \times D}$

$\frac{\partial l}{\partial b} \in \mathbb{R}^k$

$\frac{\partial b}{\partial \beta} \in \cancel{(k \times D) \times k}$

$\frac{\partial b_k}{\partial \beta} = \left[\left[\left[\partial b_k / \partial \beta_{11} \quad \cdots \quad \partial \beta_{kD} \right] \right] \right]$

$\frac{\partial b_k}{\partial \beta_{jd}}$

1D vector k

2D matrix $\cancel{k \times D}$ = 3D tensor $k \times D \times k$ $\times$

$\frac{\partial l}{\partial \beta_{jd}} =$

$$\begin{array}{cccccc} & 1 & 2 & 3 & 4 & 5 \\ \vec{y} = \{ & 0 & 0 & 1 & 0 & 0 \} \end{array}$$

$\sum y_l = 1$

$0.7 - \frac{1}{?} = 9$

$0.7 - 0 = 10$