Halving Algorithm

Input: $\mathcal{H}$

Init: $V = \mathcal{H}$, $Z = \emptyset$

while $|V| > 1$

$b(x) = \text{MajorityVote}(h(x) : h \in V) \quad \forall x$

Let S be a minimal specifying set for $b \in \{0,1\}^n$

and request $y = h^*(x) \ \forall x \in S$. $Z = Z \cup \{(x_i, y_i)\}_{i \in S}$

Update $V = \{h \in V : h(x) = y \ \forall (x,y) \in Z\}$

Note: $|S| < \text{EXT-TD}(\mathcal{H})$ at all rounds

Case 1: $\exists h \in \mathcal{H}$: $y = h(x) \ \forall x \in S$, but by defn there is at most one of these, and it is equal to h^*. Thus, algorithm terminates w/ $V = \{h^*\}$

Case 2: $\exists x \in S$: $y \neq b(x)$. But by defn of Majority vote, at least <u>half</u> of hypotheses in V agreed that $h^*(x)$ was $b(x)$. Thus, when proven wrong, $|V| \to |V|/2$

$\Longrightarrow$ At each round, either $|V|$ is halved or terminates

at most $\text{EXT-TD}(\mathcal{H})$ queries per round

$\Rightarrow$ # labels $\leq \lceil \log_2(\mathcal{H}) \rceil \cdot \text{EXT-TD}(\mathcal{H})$ ∎

Another simpler algorithm...

<u>Generalized</u> <u>Binary</u> <u>Search</u> (GBS)

Input: $\mathcal{H}$, $p \in \Delta_{\mathcal{H}}$: $\sum_{h \in \mathcal{H}} p(h) = 1$

Init: $V = \mathcal{H}$ $\quad (h : \mathcal{X} \to \{0,1\})$

while $|V| > 1$

$$x' \leftarrow \operatorname*{argmin}_{x \in \mathcal{X}} \left| \frac{1}{2} - \sum_{h \in V} p(h) h(x) \right|$$

Request $y' = h^*(x')$ update $V = \{h \in \mathcal{H} : h(x') = y'\}$

<u>Theorem</u>) Fix any $\mathcal{H}$ and $p \in \Delta_{\mathcal{H}}$. Let OPT $=$ $\min_{\mathcal{A}} \mathbb{E}_{h^* \sim p}\left[\varsigma(\mathcal{H}, \mathcal{H}, h^*, \mathcal{A}) \right]$. Then GBS satisfies $\mathbb{E}_{h^* \sim p}\left[\varsigma(\mathcal{H}, \mathcal{H}, h^*, \mathcal{A}_{GBS}) \right] \le \mathrm{OPT} \cdot 4 \log\left(\frac{1}{\min_h p(h)} \right)$

If $p(h)$ is uniform $\frac{1}{|\mathcal{H}|}$ then this is a $\log(|\mathcal{H}|)$ approx ratio.

Separable, streaming setting

$\mathcal{X}$ can be uncountable but I have access to an oracle $\mathcal{D}_X$ s.t. I can request $x \sim \mathcal{D}_X$. Assume $\exists h^* \in \mathcal{H}$ s.t. $Y = h^*(X)$ for any $X \sim \mathcal{D}_X$ and requested label Y.

$$\text{Risk}(h) = \mathbb{E}_{(X,Y)\sim\mathcal{D}}\left[\mathbb{1}\{h(X) \neq Y\}\right] \qquad (Y = h^*(X),\ X \sim \mathcal{D}_X)$$

How many unlabeled examples and queried labels must I make to identify an $\hat{h}$: $R(\hat{h}) \leq \varepsilon$? (Note $R(h^*) = 0$)

Passive learning Review (Realizable / Separable $\equiv h^*(X) = Y$)

Suppose $(x_i, y_i) \overset{iid}{\sim} \mathcal{D}$ for $i = 1, 2, \dots, n$.

and compute $\hat{h}_n = \underset{h \in \mathcal{H}}{\arg\min} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{h(x_i) \neq y_i\}}_{\hat{R}_n(h)}$

what is $R(\hat{h}_n)$? Rather, how large must n be to have $R(\hat{h}_n) \leq \varepsilon$ w.p. $\geq 1-\delta$?

$\boxed{\text{Thm}}$ Assume $\mathcal{H}$ is finite. $R(\hat{h}_n) \leq \varepsilon$ w.p. $\geq 1-\delta$ whenever $n \geq \varepsilon^{-1} \log(|\mathcal{H}|/\delta)$.

Proof $\hat{R}_n(h^*) = \hat{R}_n(\hat{h}_n) = 0$.

$$\mathbb{P}(R(\hat{h}_n) > \varepsilon) \leq \mathbb{P}\left(\bigcup_{h \in \mathcal{H}} \{R(h) > \varepsilon,\ \hat{R}_n(h) = 0\}\right)$$

$$\le \sum_{h \in \mathcal{H}} \mathbb{P}\left(R(h) > \varepsilon,\ \hat{R}_n(h) = 0 \right)$$

$$\le \sum_{h \in \mathcal{H}} (1-\varepsilon)^n$$

$$\le |\mathcal{H}| (1-\varepsilon)^n \qquad \text{(union bound)}$$

$$\le |\mathcal{H}| \exp(-n\varepsilon) \qquad 1-x \le e^{-x}$$

$$\le \delta$$

when $n \ge \varepsilon^{-1} \log(|\mathcal{H}|/\delta)$. $\boxtimes$

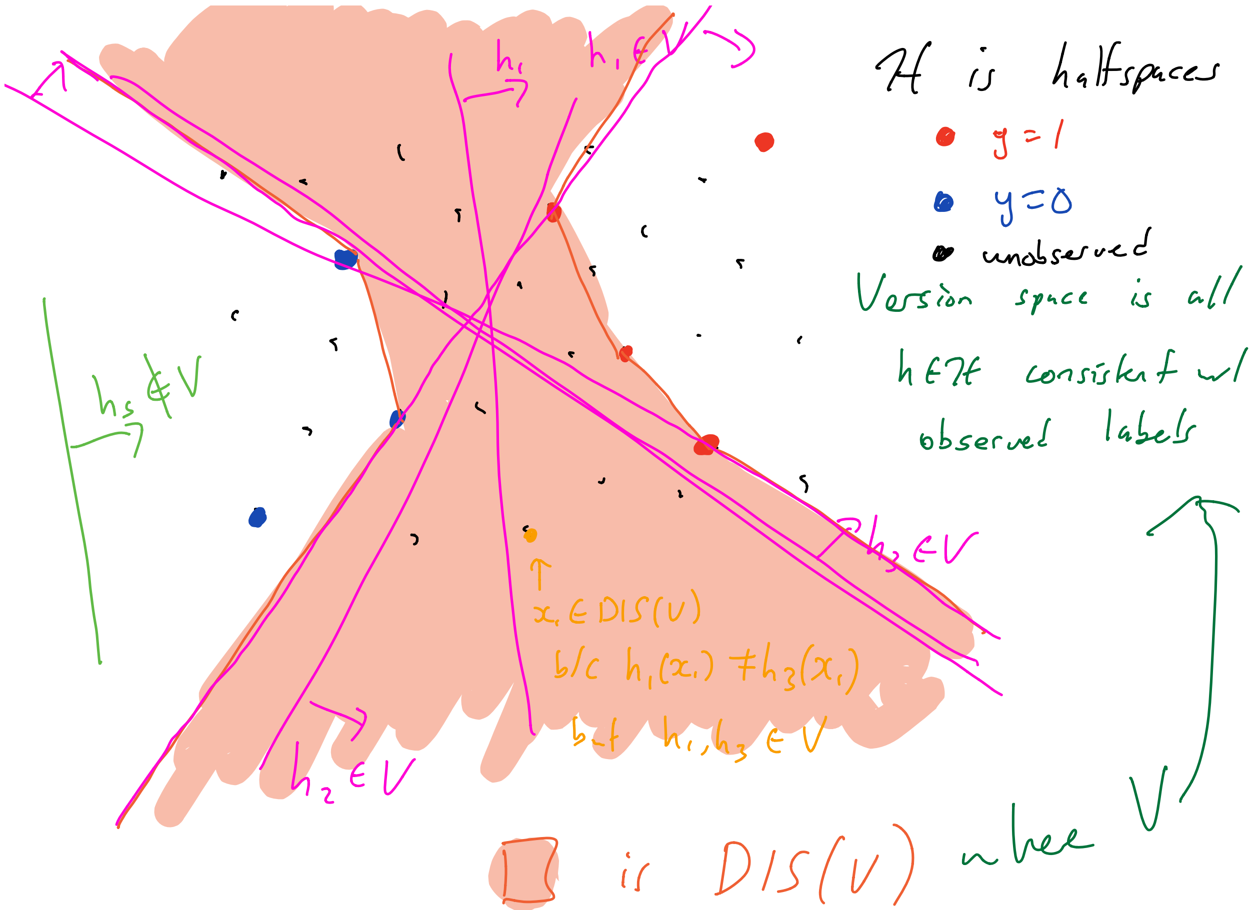
(Empirical)
Aside: Bernstein says

$$|\hat{R}_n(h) - R(h)| \le \sqrt{\frac{2\hat{R}_n(h) \log(|\mathcal{H}|/\delta)}{n}} + \frac{\log(|\mathcal{H}|/\delta)}{3n}$$

Active Learning / Disagreement-based learning.

Def] For any $V \subset \mathcal{H}$ define <u>region of disagreement</u> of V as

$$DIS(V) = \{x : \exists h, h' \in V : h(x) \neq h'(x)\}$$



CAL (Cohn, Atlas, Ladner 1993)

Init $Z_0 = \emptyset$, $V_0 = \mathcal{H}$

for $t = 1, 2, \ldots$

Nature reveals $x_t \sim \mathcal{D}_x$

if $x_t \in DIS(V_{t-1})$ then

query $y_t = h^*(x_t)$, $Z_t = Z_{t-1} \cup \{(x_t, y_t)\}$

else

$Z_t = Z_{t-1}$

$$V_t = \{h \in \mathcal{H} : h(x_s) = y_s \quad \forall (x_s, y_s) \in Z_t\}$$

Hanneke in mid '00s analyzed it.

Dasgupta, Hsu, Monteleoni in '09 (?)

presented computational eff. version

Efficient CAL

Init $Z_0 = \emptyset$

for $t = 1, 2, \ldots$

Nature reveals $x_t \sim \mathcal{D}_x$

if for $\hat{y} \in \{0,1\}$ $\exists h_{\hat{y}} \in \mathcal{H}$ with

$$h_{\hat{y}}(x_s) = y_s \quad \forall (x_s, y_s) \in Z_{t-1} \cup (x_t, \hat{y})$$

then request y_t, $Z_t = Z_{t-1} \cup (x_t, y_t)$

else

$Z_t = Z_{t-1}$.

If $\mathcal{H}$ is halfspaces linear program.