* PAC for inconsistent-hypo sets.

① Supp. we use agnostic PAC to analyse algorithms in the "inconsistent-hyp set" case. Also, suppose that the algo. works by solving the following optimization pb:

ERM (empirical risk minimization).

$$h_S^{ERM} = \operatorname{argmin}_{h \in \mathcal{H}} \hat{R}_S(h).$$ for a given seq. S of samples.

Denote this algo. by A_{ERM}.

② Can we prove something like Thm 2.5 that says that: if $|\mathcal{H}|$ is at most exponential in n, A_{ERM} is an agnostic PAC algo.?

③ The answer is yes. Here is the thm.

Thm $\forall D \in Pr(\mathcal{X} \times \mathcal{Y})$ $\forall \varepsilon, \delta > 0$ $\forall m \in \mathbb{N}$.

if $m \geq \frac{2}{\varepsilon^2}\left(\log|\mathcal{H}| + \log\frac{2}{\delta}\right)$, then

$$\mathbb{P}_{S \sim D^m}\left[R(h_S^{ERM}) - \min_{h \in \mathcal{H}} R(h) \leq \varepsilon\right] \geq 1 - \delta.$$

We first note the following proposition from Ch4.

Prop 4.1 $$\mathbb{P}_{S \sim D^m}\left[R(h_S^{ERM}) - \min_{h \in \mathcal{H}} R(h) > \varepsilon\right] \leq \mathbb{P}_{S \sim D^m}\left[\max_{h \in \mathcal{H}} |R(h) - \hat{R}_S(h)| > \frac{\varepsilon}{2}\right].$$

for all $\varepsilon > 0$ and $m \in \mathbb{N}$.

Proof. Let $h_0 = \operatorname{argmin}_{h \in \mathcal{H}} R(h)$ (which is defined since $\mathcal{H}$ is finite).

Also, let $S = ((x_1, y_1), \cdots, (x_m, y_m))$ be the seq. of samples drawn independently from D. Then,

$$R(h_S^{ERM}) - \min_{h \in \mathcal{H}} R(h) = R(h_S^{ERM}) - R(h_0)$$

$$= R(h_S^{ERM}) - \widehat{R}_S(h_S^{ERM}) + \widehat{R}_S(h_S^{ERM}) - R(h_0)$$

$$\leq R(h_S^{ERM}) - \widehat{R}_S(h_S^{ERM}) + \widehat{R}_S(h_0) - R(h_0)$$

$$\leq 2 \max_{h \in \mathcal{H}} |R(h) - \widehat{R}_S(h)|$$

So, $\mathbb{P}[R(h_S^{ERM}) - \min_{h \in \mathcal{H}} R(h) > \varepsilon]$

$$\leq \mathbb{P}[\max_{h \in \mathcal{H}} |R(h) - \widehat{R}_S(h)| > \tfrac{\varepsilon}{2}].$$ D.

Before proving the theorem, we note that the following variant of Thm 2.13 also holds essentially by the same proof.

Thm 2.13 b. $\forall D \in \Pr(\mathcal{X} \times \mathcal{Y})$ $\forall \varepsilon, \delta > 0$ $\forall m \in \mathbb{N}$.

if $m \geq \frac{1}{2\varepsilon^2}\left(\log|\mathcal{H}| + \log\frac{2}{\delta}\right)$, then → new to this thm 2.13 b.

(missed in Thm 2.13 by mistake)

$$\mathbb{P}[\forall h \in \mathcal{H}.\ |R(h) - \widehat{R}_S(h)| \leq \varepsilon] \geq 1 - \delta.$$

Using what we have shown so far, we prove our thm.

Proof of the thm

We will show $\mathbb{P}_{S \sim D^m}[R(h_S^{ERM}) - \min_{h \in \mathcal{H}} R(h) > \varepsilon] \leq \delta.$

By Prop. 4.1,

$$\underset{S\sim D^m}{\mathbb{P}}\left[R(h_S^{ERM}) - \min_{h\in\mathcal{H}} R(h) > \varepsilon \right]$$

$$\leq \underset{S\sim D^m}{\mathbb{P}}\left[\max_{h\in\mathcal{H}} |R(h) - \widehat{R}_S(h)| > \frac{\varepsilon}{2} \right]$$

By assumption of the thm.,

$$m \geq \frac{1}{2\times\left(\frac{\varepsilon}{2}\right)^2}\left(\log|\mathcal{H}| + \log\frac{2}{\delta}\right).$$

Thus, we can use Thm 2.13b with $\frac{\varepsilon}{2}$, and derive:

$$" \leq \delta.$$

D.