## Relaxing realizability

$(x, y) \sim D$

True error $L_D(h) \triangleq \Pr_D[h(x) \neq y]$

Empirical error $L_S(h) \triangleq \frac{|\{i : h(x_i) \neq y_i\}|}{m}$

Goal: find $h$ minimizes $L_D(h)$

Bayes optimal predictor: $f_D(x) = \begin{cases} 1 & \text{if } P_D(y=1|x) \geq \frac{1}{2} \\ 0 & \end{cases}$

lowest true error of any predictor over $D$

but requires knowing $D$

## Agnostic PAC learnability

A class $\mathcal{H}$ is agnostic PAC learnable

if $\exists\, m_{\mathcal{H}}(\epsilon, \delta)$ & learning algorithm

$\forall$ distributions $D$, $\forall \epsilon > 0$, $\delta \in (0,1)$

s.t. Algo on $m_{\mathcal{H}}(\delta, \epsilon)$ ~~returns h~~

returns $h$ w.p. $1-\delta$

s.t. $L_D(h) \leq \min_{\hat{h} \in \mathcal{H}} L_D(\hat{h}) + \epsilon$

## Relaxing binary

Multiclass $Y = [k]$

$\ell_h(x, y) = \mathbb{I}[h(x) \neq y]$

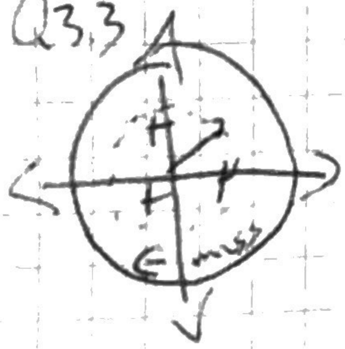
Regression loss (e.g.) $\ell_h(x, y) = (h(x) - y)^2$

Risk fn $L_D(h) \triangleq \underset{(x,y)\sim D}{\mathbb{E}} [\ell(h,(x,y))]$

Empirical RF $L_S(h) \triangleq \frac{1}{m} \sum_{(x_i,y_i)} \ell(h,(x_i,y_i))$

---

Q3.3



$\mathcal{X} = \mathbb{R}^2$, $\mathcal{Y} = \{0,1\}$ $\mathcal{H} = \{h_r : r \in \mathbb{R}_+\}$

$h_r(x) = \mathbb{1}_{\{\|x\| \le r\}}$

$\mathcal{H}$ is [realizably] PAC learnable w/

$$m_{\mathcal{H}}(\epsilon,\delta) \le \frac{\log(\frac{1}{\delta})}{\epsilon}$$

---

Pf: Alg: Return smallest circle including all +s.

Fixed $D$, $f_{r^*}$, $\epsilon$

$\mathcal{H}_{Bad} = \{ r : \tilde{r} \le r^*, D(\tilde{r} \le \|x\| \le r^*) = \epsilon \}$

$\Pr_S(H^{Algo}_{Bad}) = \Pr[\text{Algo finding } \tilde{r}] \le (1-\epsilon)^m$

$\le e^{-\epsilon m}$

when $m \ge \frac{\lg \frac{1}{\delta}}{\epsilon}$ $\longrightarrow$ $\le \delta$

---

Argument: you only return a smaller ○ than $r^*$, since realizable + algo def. So you have error from the space btwn $\tilde{r}$ + $r^*$

How does one get agnostic PAC learnability?

[reminder: realizable PAC for $|\mathcal{H}| = K$ comes from ERM]...

Possibly also true for agnostic case, dif sample complexity?

Chap 4: Uniform Convergence $\Rightarrow$ Agnostic PAC learnability

Def $S$ is an $\epsilon$-rep sample if $\forall h \in \mathcal{H}$

$$|L_S(h) - L_D(h)| \leq \epsilon$$

If we have, we are good. Pr ? ?

$\boxed{\text{Lemma}}$ Assume $S$ is $\frac{\epsilon}{2}$-rep sample for $\mathcal{H}$.

Then, $ERM_S \in \min_{h \in \mathcal{H}} L_S(h)$ has

$$L_D(ERM_S) \leq \min_{h \in \mathcal{H}} L_D(h) + \epsilon.$$

Pf [triangle $\neq$ ~~$|a|+|b| \geq |a+b|$~~ $a \leq b + |a-b|$

$$L_D(h_S) \stackrel{\Delta\neq}{\leq} |L_D(h_S) - L_S(h_S)| + L_S(h_S)$$

rep sample $\rightarrow \leq \epsilon/2 \quad \underbrace{+L_S(h_S) - L_S(h)}_{\leq 0 \text{ ERM}} + L_S(h)$

$$\leq \epsilon/2 + 0 + L_S(h)$$

$$\stackrel{\text{triangle}}{\leq} \epsilon/2 + |L_D(h) - L_S(h)| + L_D(h)$$

$$\leq \epsilon + L_D(h)$$

$\epsilon$-rep $\forall h \in \mathcal{H}$

It remains to show w.p. $1-\delta$, $S$ is $\epsilon/2$ repr.

$\boxed{\text{Def}}$ ~~the~~ $\mathcal{H}$ has the uniform convergence prop if $\exists\ m^{uc}_{\mathcal{H}}(\epsilon,\delta)$ s.t. $\forall \epsilon,\delta, D$ $m \geq m^{uc}_{\mathcal{H}}(\epsilon,\delta)$ the $S$ is $\epsilon$-representative w.p. $1-\delta$.

---

Cor 1 $\mathcal{H}$ is agnostically PAC-learn w. sample complex $m_{\mathcal{H}}(\epsilon,\delta) \leq m^{uc}_{\mathcal{H}}(\frac{\epsilon}{2},\delta)$ ERM is a PAC learner.

---

[Now in remains to show UC / $\epsilon/2$ rep sample w.p. $1-\delta$]

Claim: Any finite $|\mathcal{H}| \leq \infty$ is UC, w.

$$m_{\mathcal{H}}(\epsilon,\delta) \leq \left\lceil \frac{\lg \frac{2|\mathcal{H}|}{\delta}}{\epsilon^2} \right\rceil$$

Cor 2 ERM is an agnostic ~~ally~~ PAC learner w. sample complexity

$$m_{\mathcal{H}}(\epsilon,\delta) \underset{\text{Cor 1}}{\leq} m^{uc}_{\mathcal{H}}\left(\frac{\epsilon}{2},\delta\right) \underset{\text{Claim}}{\leq} \left\lceil \frac{2\lg \frac{2\mathcal{H}}{\delta}}{\epsilon^2} \right\rceil$$

Proof of Claim

Q: what sample size guars $|L_D(h) - L_S(h)| \leq \epsilon$ w.p $1-\delta$

$$\text{N.t.s. } \mathbb{P}_S(\{S : \forall h \in \mathcal{H}, |L_S(h) - L_D(h)| \leq \epsilon\}) \geq 1-\delta$$

$$\Leftarrow \mathbb{P}_S(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_D(h)| > \epsilon\}) < \delta$$

Union Bd $\leq \sum_{h \in \mathcal{H}} \mathbb{P}_S(\{S : |L_S(h) - L_D(h)| > \epsilon\})$ *

$\boxed{\text{Assume } 0 \leq \ell(h,z) \leq 1}$

$$L_S(h) = \frac{1}{m} \sum_{i=1}^{m} \ell(h, z_i)$$ ← avg of iid RV $z_i = (x_i, y_i)$

$$L_D(h) = \mathbb{E}_{z \sim D}[\ell(h,z)] = \mathbb{E}_S[L_S(h)]$$ ← expectation of each term

So we need to argue the measure (dist) of $L_S(h)$ concentrates around its mean $L_D(h)$

Many tail inequalities do this in var. contexts!

Let's use Hoeffding's Inequality:

If $\theta_1, \ldots \theta_m$ is a set of iid RVs; $\mu \triangleq \mathbb{E}[\theta_i]$

- $\Pr[a \leq \theta_i \leq b] = 1 \quad \forall i$

then $\forall \epsilon > 0$

$$\Pr\left(\left|\frac{1}{m}\sum \theta_i - \mu\right| > \epsilon\right) \leq 2e^{\frac{-2m\epsilon^2}{(b-a)^2}}$$

Then, using $\theta_i = \ell(h, z_i)$ $a = 0 \leq \ell(h, z_i) \leq 1 \leq b$

$\mu = L_D(h)$

$$\mathbb{P}[|L_D(h) - L_S(h)| > \epsilon] \leq \mathbb{P}[|\tfrac{1}{m}\sum \theta_i - \mu| > \epsilon)$$

Hoeff $\rightarrow$ $\leq 2e^{-2m\epsilon^2}$

then Union Bd $\star$ $\leq 2|\mathcal{H}|e^{-2m\epsilon^2}$ $\leq \delta$

& $m(\epsilon, \delta) \geq \dfrac{\lg\left(\frac{2|\mathcal{H}|}{\delta}\right)}{2\epsilon^2}$

---

Summary

$$\boxed{\text{If } m \geq \left\lceil \dfrac{\log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{2\epsilon^2} \right\rceil \text{ the } \mathcal{H} \text{ has u.c.}}$$

Lemma: $(\epsilon, \delta)$ UC $\Rightarrow$ $m_{\mathcal{H}}^{UC}(\frac{\epsilon}{2}, \delta) \geq m_{\mathcal{H}}(\epsilon, \delta)$

$\downarrow$

$(\epsilon, \delta)$ PAC learnable w. ERM

If $m \geq \left\lceil \dfrac{2 \log\left(\frac{2|\mathcal{H}|}{\delta}\right)}{2\epsilon^2} \right\rceil \geq m_{\mathcal{H}}^{UC}(\frac{\epsilon}{2}, \delta)$

then $\mathcal{H}$ is $(\epsilon,\delta)$ PAC learnable $\geq m_{\mathcal{H}}(\epsilon, \delta)$

w. ERM