ML.exer.ch2

samira moradzade

5 azar 1400

1 Exercise 3.2:

let X be a discrete domain, and let

$$H_{singleton} = \{h_z : z \in X\} \cup \left\{h^-\right\}$$

where for each $z \in X$, h_z is the function defined by $h_z(x) = 1$ if $x = z$ and $h_z(x) = 0$ if $x \neq z$. h^- is simply the all-negative hypothesis, namely, $\forall x \in X$, $h^-(x) = 0$. The realizability assumption here implies that the true hypothesis f labels negatively all examples in the domain, perhaps except one.

1.1 a:

we want to describe that the algorithm implements is an ERM algorithm. so we have to show that the empirical risk is minimum. Consider the training set as $S = X \times \{0, 1\}$ by the facts given if $(x_+, 1) \in S$ the algorithm returns h_{x_+} else it returns h^- and because there can be only one point takes positive label (means label 1) so $L_s(h_s) = 0$ because the error is the minimum so its ERM.

1.2 b:

in this part we want to show that $S = X \times \{0, 1\}$ is PAC learnable and provide an upper bound on the sample complexity. let $\epsilon \in (0, 1)$ and the distribution of D over X if the h^- be the true hypothesis then the algorithm returns a perfect h. We know that there is a uniqe point x_+. If $(x_+, 1) \in S$ the algorithm returns perfect hypothesis if we consider $D[x_+] \leq \epsilon$:

$$L_D\left\{h_s\right\} = D\left(\{x : h_s(x) \neq f(x)\}\right) = D(\{x_+\}) \leqslant \epsilon$$

it means the max value that generalization error of h_s takes is ϵ. Now we need to bound the probability of bad cases ($D(\{x_+\}) > \epsilon$, but $(x_+, 1) \notin S$)

$$D^m\left(\{s : L_D(hs) > \epsilon\}\right) = D^m\left(\{s : x_+ \notin S\}\right) = (1 - \epsilon)^m \leq e^{-\epsilon m}$$

so $H_{singleton}$ is PAC learnable with sample complexity:

$$m_{H_{singleton}} \leq \left\lceil \frac{log\left(\frac{1}{\delta}\right)}{\epsilon} \right\rceil$$

2 Exercise 3.3:

proving that $H = \{h_r : r \in \mathbb{R}_+\}$ where $h_r = 1_{[\|x\| \leq r]}$ is PAC learnable.

2.1 solution:

let $\delta, \epsilon \in (0,1)$ and the distribution D over X and consider $S = ((x_i, y_i))_{i=1}^m$ and r^* as the radius of the tightest circle which contains all the positive constants so it returns the hypothesis h^* Besides assume realizability and let h_0 be a circle with zero generation error by the radius r_0 so $r^* \leq r_0$ Now define the set $R := \{x : r^* \leq \|x\| \leq r_0\} = \epsilon$ if a point in R be in S for the error hr we have:

$$\begin{aligned} L_D(h_r) = D(\{x : h_r(x) \neq h_{r^*}(x)\}) = D(\{x : r \leq \|x\| \leq r_0\}) \\ \leq D(\{x : r^* \leq \|x\| \leq r_0\}) \leq D(R) = \epsilon \end{aligned}$$

and now for the worst set we have the bounding below:

$$\begin{aligned} D^m(\{S|_x : L_D(h_s) > \epsilon\}) = D^m(\{S|_x : S|_x \cap R = \emptyset\}) = \\ \bigcap_{i=0}^{m} D(\{x : x \in S|_x, x \notin R\}) = \prod_{i=0}^{m}(i - \epsilon) = (1-\epsilon)^m \leq e^{-\epsilon m} \end{aligned}$$

Therefore H is PAC learable and the sample complexity can find by considering the above equation be lower than δ :

$$m_{(\epsilon,\delta)} \leq \left\lceil \frac{log(\frac{1}{\delta})}{\epsilon} \right\rceil$$

3 Exercise 3.6:

Let H be a hypothesis class of binary classifiers. Show that if H is agnostic PAC learnable, then H is PAC laernable as well. Furthermore, if A is a successful agnostic PAC learner for H, then A successful PAC learner for H. **solution:**
H is agnostic PAC learnable so there is $m_H : (0,1)^2 \to \mathbb{N}$ and the algorithm named A (the learner for H) for every $\epsilon, \delta \in (0,1)$ and the distribution D over $(x \times y)$. if A running on $m \geq m_H(\epsilon, \delta)$ returns hypothesis h s.t. with probability $1 - \delta$

$$L_D(h) \leq \min_{h \in H} L_D(h^{'}) + \epsilon$$

if the realizibality assumption holds then $\min_{h \in H} L_D(h^{'}) = 0$ and if f labeled m iid constances set S. then we have

$$L_D(h) \leq 0 + \epsilon = \epsilon$$

Therefore A is successful PAC learnable for H.