Assignment 5

Fatemeh Rahimi

April 7, 2020

Question 3.2

part a

We know that the true hypothesis f labels negatively all examples in the domain, perhaps except one, so we suppose our algorithm as following: if there is a positive instance x in S, algorithm returns h_x, otherwise it returns h^- ,it is clear that this algorithm is an ERM.

part b

First suppose two cases:
if true labeling function is all-negative: our algorithm returns a perfect hypothesis.
if there exists a unique positive instance x^+, and it appears in the S, our algorithm returns a perfect hypothesis again. now we have two cases, if $D|\{x^+\}| < \varepsilon$ the returned hypothesis has a generalization error of at most ε. if $D|\{x^+\}| > \varepsilon$ then $D|\{x_i \neq x^+ : x_i \in S\}| \leq 1 - \varepsilon$. so $D^m|F| \leq (1-\varepsilon)^m \leq e^{\varepsilon m}$, F is event that x^+ doesn't appear in S. Hence, $H_{Singleton}$ is PAC learnable, and its sample complexity is bounded by:

$$m_h(\varepsilon, \delta) \leq \left[\frac{log(\frac{|H_{Singleton}|}{\delta})}{\varepsilon} \right]$$

Question 3.3

Suppose the ERM algorithm A which given a training sequence S = returns the hypothesis h corresponding to the tightest circle which contains all the positive instances. with realizability assumption, let h^* be true hypothesis. Denote its radius by r^*. now let $\varepsilon, \delta \in (0, 1)$ and $\bar{r} \leq r^*$ that $P[x : \bar{r} \leq \|x\| \leq r^*] = \varepsilon$ and set $E = \{x : \bar{r} \leq \|x\| \leq r^*\}$.now note that:

$$\mathbf{P}[L_D(h_s) \geq \varepsilon] = \mathbf{P}[x_i \in S \text{ s.t } x \notin E] = \prod_i (1 - \mathbf{P}[x_i \in E])^m = (1 - \varepsilon)^m \leq e^{m\varepsilon}$$

now if we set δ such that $e^{m\varepsilon} \leq \delta$ we have $m \geq \frac{log \frac{1}{\delta}}{\varepsilon}$, so H is PAC learnable with $m_h(\varepsilon, \delta) \leq \left[\frac{log(\frac{1}{\delta})}{\varepsilon} \right]$.

Question 3.4

We know H is finite hypothesis class, so its sample complexity, m is less than $\frac{log(\frac{|H|}{\delta})}{\varepsilon}$, and we konw each hypothesis in H is a conjunction, that contains x^i or $\hat{x}^i$ or none of them. so $|H| = 3^d + 1$ (we plus 1 because empty hypothesis).therefore we have $m \leq \frac{log(\frac{3^d+1}{\delta})}{\varepsilon}$.