Probabilistic Graphical Models.

Assignment 2. - 20171047

3.5) 2)

a) Algorithm:

We have sample S from X.

if S has a positive instance x^+
then return true hypothesis $h_z(x) = \begin{cases} 1 & x = z \\ 0 & \text{else} \end{cases}$

else
return all negative hypothesis h^- i.e $h^-(x) = 0$ $\forall x$

- This is an ERM as it minimizes $L_S(h)$
$= |\{ i \in [m] : h(x_i) \neq y_i \}|$ trivially.

$[m] = \{1, 2, 3, \ldots, m\}$

b) $\epsilon \in (0,1)$, D: probability distribution of samples over X

If h^- was indeed the True hypothesis, then the above ERM gives zero generalization error → we have a perfect hypothesis and we are done.

Now, if we have a single (+)ve instance i.e x^+.
if $x^+ \in$ training sequence S, the above ERM still gives a perfect solution.

if $D[\{x^+\}] \leq \epsilon$, generalization error or true error is upper bounded by ϵ w.p 1.

If $D[\{x^+\}] > \epsilon$ & x^+ is not in S → $D[\{x^+\}] > \epsilon$

Let E be the event x^+ not in S s.t

Now, $\mathbb{P}_{S|x \sim D^m}[E] \leq (1-\epsilon)^m \leq e^{-m\epsilon}$

∴ $\mathcal{H}_{singleton}$ defined in the problem is PAC learnable

And sample complexity $m_{\mathcal{H}}(\epsilon, \delta) \leq \frac{\log(|\mathcal{H}|/\delta)}{\epsilon}$

$= \log(1/\delta)/\epsilon = -\frac{\log \delta}{\epsilon}$ & $m_{\mathcal{H}}(\epsilon, \delta) \in \mathbb{I}$.

4.5) 1) A is the learning algorithm.

$\therefore A(S)$ is the generated hypothesis.

D : probability distribution over X.

Given : Range of loss functions L is $[0,1]$

Assume $\forall\ \epsilon, \delta \in (0,1)$ & $\forall\ D'$ over $X \times \{0,1\}$,

$\exists\ m(\epsilon,\delta) \in \mathbb{N}$ (upper bound) s.t $\forall\ m \geq m(\epsilon,\delta)$

$$\mathbb{P}_{S\sim D^m}[L_D(A(S)) > \epsilon] < \delta$$

→ Assuming binary labels $\{0,1\}$ for instances in X.

To show that $\exists\ m_0 \in \mathbb{N}$ s.t $\forall\ m \geq m_0$, $\mathbb{E}_{S\sim D^m}[L_D(A(S))]$ $\leq \lambda$ for some $\lambda > 0$ ⟶

• Let $\epsilon = \dfrac{\min(1,\lambda)}{2}$. $m_0 = m_{\mathcal{H}}(\epsilon,\epsilon)$

Now, $\forall\ m \geq m_0$ since $L \leq 1$;

$$\mathbb{E}_{S\sim D^m}[L_D(A(S))] \leq \mathbb{P}_{S\sim D^m}[L_D(A(S)) > \lambda/2]\cdot 1 + \mathbb{P}_{S\sim D^m}[L_D(A(S)) \leq \lambda/2]\cdot\lambda/2$$

$$\leq \mathbb{P}_{S\sim D^m}[L_D(A(S)) > \epsilon] + \lambda/2$$

$$\leq \epsilon + \lambda/2 \leq \lambda/2 + \lambda/2 = \lambda.$$

Also, if $\lim_{m\to\infty} \mathbb{E}_{S\sim D^m}[L_D(A(S))] = 0$

for $\epsilon,\delta \in (0,1)$; $\exists\ m_0 \in \mathbb{N}$ s.t $\forall\ m \geq m_0$,

$$\mathbb{E}_{S\sim D^m}[L_D(A(S))] \leq \epsilon\cdot\delta.$$

Markov's Inequality gives us :

$$\mathbb{P}_{S\sim D^m}[L_D(A(S)) > \epsilon] \leq \frac{\mathbb{E}_{S\sim D^m}[L_D(A(S))]}{\epsilon} \leq \frac{\cancel{\epsilon}\delta}{\cancel{\epsilon}} = \delta$$

6.8) 1) $\mathcal{H}' \subseteq \mathcal{H}$; $\therefore$ for all $C = \{c_1, c_2, \dots, c_m\}$. m being the number of samples. $C \subseteq X$; we obtain

$\mathcal{H}'_C \subseteq \mathcal{H}_C$.

So, if C is shattered by some h in $\mathcal{H}'$; $h \in \mathcal{H}$ as well as $\mathcal{H}' \subseteq \mathcal{H}$ & in that $\mathcal{H}$ shatters C.

$\therefore$ By definition : $VC\dim(\mathcal{H}') \leq VC\dim(\mathcal{H})$