$m_2 \leq m_1$

تمرین سری ۲

(2) $H_{singleton} = \{h_z : z \in X\} \cup \{h^-\}$

$h_z(x) = \begin{cases} 1 & x = z \\ 0 & x \neq z \end{cases}$ $\qquad h^- = \begin{cases} 0 & x \in X \\ 1 & x \notin X \end{cases}$

الف) ثابت شود ERM است $H_{singleton}$

الگوریتم: برای x_+ فرض h_{x_+} برنگردانده و برای $\{x_i\}$ ها می‌دهد

h_{x_-} برنگردانده واضح است که خطای تجربی صفر دارد، که ERM است

حال نشان دهید $H_{singleton}$ PAC learnable است

و کران بالا برای sample complexity بیابید:

طبق سوال $\bar{h}$ True lable ماست. اما یک x_+ وجود دارد

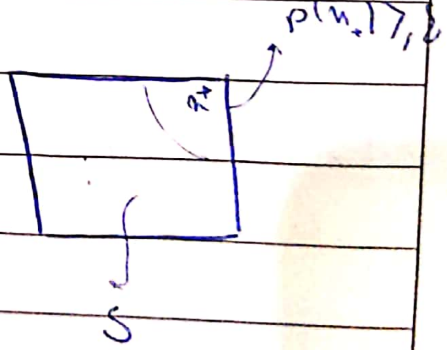
$p(x_+) \geq \epsilon$ حال برای [illegible] ها $p(x_+)$

$$(1-\epsilon)(1-\epsilon)\dots(1-\epsilon)^m \leq e^{-m\epsilon} \leq \delta$$

$$-m\epsilon \leq \log(\delta) \Rightarrow$$

پس $H_{singleton}$ دارای خاصیت PAC learnability، و کران بالای آن

$$\Rightarrow m_H(\epsilon,\delta) \leq \frac{\log(\frac{1}{\delta})}{\epsilon}$$



Date .... Subject ....

$X = \mathbb{R}^2 , \quad Y = \{0, 1\}$ 13

$H = \{h_r : r \in \mathbb{R}_+\} \qquad h_r(x) = \begin{cases} 1 & \|x\| \le r \\ 0 & o.w \end{cases}$

الگوریتم A را طوری تعریف می‌کنیم که به ازای نمونه دلخواه $S = \{(x_i, y_i)\}_{i=1}^{m}$

$h(S)$ برگرداند که کوچکترین دایره به عنوان نمونه‌های مثبت است (با توجه به ERM بودن)

که شعاع دایره $h(S)$ را با r نشان می‌دهیم و ما با توجه به فرض realizability

می‌دانیم که r^* وجود دارد که تمام لیبل‌ها را درست پیش بینی می‌کند

اگر $\varepsilon, \delta \in (0,1)$ فرض کنیم، $r' \le r^*$ باشد به طوری که

$P_x(\{x : r' \le \|x\| \le r^*\}) = \varepsilon$

و پیشامد خطا را به صورت زیر در نظر می‌گیریم

$E = \{x \in \mathbb{R}^2 : \bar{r} \le \|x\| \le r^*\}$

احتمال خطا $(1-\varepsilon)^m \le e^{-\varepsilon m} \le \delta \Rightarrow m_H(\varepsilon, \delta) \ge \frac{\log(\frac{1}{\delta})}{\varepsilon}$

$-\varepsilon m \le \log \delta \Rightarrow m \ge \frac{\log \frac{1}{\delta}}{\varepsilon}$

Scanned with CamScanner

پس H PCA learnable است و $r \leq r'$ زیرا اگر $r \neq r$ باشد

الگوریتم همان داده‌ها بر نمی‌گرداند

$\wedge$ ضرب
$\bar{n}$ متمم

چون ۳ حالت داریم یا x_i یا $\bar{x}_i$ یا هیچ کدام)

Date .... Subject ....

(4) تعداد کل فرضیه‌ها برابر 3^d+1 است (Al-negative) و داریم

$$m(\varepsilon,\delta) < \frac{\log(|H|/\delta)}{\varepsilon} < \frac{\log(3^d+1/\delta)}{\varepsilon} < \frac{\log(3^d/\delta)}{\varepsilon}$$

$$< \frac{d\log(3)+\log\frac{1}{\delta}}{\varepsilon}$$

الگوریتم h_0 را به صورت زیر تعریف می‌کنیم.

$$h_0 = x_1 \wedge \bar{x}_1 \wedge x_2 \wedge \bar{x}_2 \ldots x_d \wedge \bar{x}_d$$

h_0 همیشه منفی است بنابراین اگر y در مجموعه آموزشی منفی باشد این تابع مقدار درست به ما برمی‌گرداند و برای مقادیر مثبت را هم باید h_0 نوشته دهد تا ERM باشد

5

$$L_{(\bar{D}_m,f)}(h) > \epsilon$$

$$\frac{P_{x\sim D_1}[h(x)=f(x)]+\dots+P_{x\sim D_m}[h(x)=f(x)]}{m} < 1-\epsilon$$

$$P_{S\sim \prod_{i=1}^{m} D_i}[L_S(h)=0] = \prod_{i=1}^{m} P_{x\sim D_i}[h(x)=f(x)]$$

$$= \left(\left(\prod_{i=1}^{m} P_{x\sim D_i}[h(x)=f(x)]\right)^{\frac{1}{m}}\right)^{m} \leq \left(\frac{\sum P_{x\sim D_i}[h(x)=f(x)]}{m}\right)^{m}$$

$$< (1-\epsilon)^m \leq e^{-\epsilon m}$$

$$\frac{x_1+x_2+\dots+x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$ (و عبارت تساوی هرگاه $x_1=x_2=\dots=x_n$)

برای h ها : $P(L_{(\bar{D}_m,f)}(h) > \epsilon) \leq |H| \cancel{\exp(-\epsilon m)}$

زیرا باز $h\in H$ تعداد را اینجا داریم که در شرط $L_{\bar{D}_m,f}(h) > \epsilon$ صدق کند

پس حداکثر از $|H|$ کوچکتر است کوچکتر بودن از $e^{-\epsilon m}$ را در بالا ثابت کردیم پس

$$P(L_{\bar{D}_m,f}(h) > \epsilon) \leq |H|\, e^{-\epsilon m}$$

6) میدانیم H Agnostic PAC learnable است

پس $\epsilon, \delta \in (0,1)$ و توزیع X بر روی $X \times Y$ داریم

در اینجا $Y = \{0,1\}$ منظور تشخیص می‌دانیم

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

رابطه بالا با فرض Agnostic PAC L برقرار است و

با فرض realizability ، $\exists h^* \in H \quad L_D(h^*) = 0$ پس:

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

که یعنی H PAC learnable است

pessimistic

Date .... Subject $\alpha_x = p(Y=1 | X=x)$

7

$P[f_D(X) \neq y | X=x] =$

$1_{[\alpha_x \geq 0.5]} P[Y=0 | X=x] + 1_{[\alpha_x < 0.5]} P(Y=1 | X=x)$

$= 1_{[\alpha_x \geq 0.5]} (1-\alpha_x) + 1_{[\alpha_x < 0.5]} \alpha_x$

$= \min\{\alpha_x, 1-\alpha_x\}$

$L_D(f_D) = E_{(x,y) \sim D}(1_{f_D(x) \neq y})$

$= E_{x \sim D_x}[\underbrace{E_{y \sim D_{y|x}}[1_{[f_D(x) \neq y]} | X=x]}_{\alpha_x}]$

$\leq E_{x \sim D_x}[E_{y \sim D_{Y|x}}[1_{g(x) \neq y} | X=x]]$

$= L_D(g)$

$L_D(f_D) \leq L_D(g)$

پس optimal f_D است

Scanned with CamScanner

Date .... Subject ....

زمل

5) 2. در مدلی که از همه متغیرها استفاده شده امکان وقوع overfitting بیشتر

است و همچنین با توجه به این که بعد H افزایش را یابد طبیعتا تعداد مجموعه

آموزش بیشتری مورد نیاز است پس بهتر است که الگوریتم با دو متغیر

BP و BMI انتخاب شود.