Aryan Sabounchi

1. تعریف Pac learnability، $\forall\, m \geq m_H(\epsilon_1, \delta)$ و هر توزیع D روی X، به روی S train set داریم:

$$P\{L_D(A(S_m)) < \epsilon_1\} \geq 1-\delta$$

اگر $L_D(A(S_m)) < \epsilon_1$ باشد، بر اساس این که $\epsilon_1 \leq \epsilon_2$ (فرض مسأله)، $L_D((S_m)) < \epsilon_2$ و به ازای هر $m \geq m_H(\epsilon_2, \delta)$

$$P\{L_D(A(S_m)) < \epsilon_2\} \geq 1-\delta$$

چون $m = m_H(\epsilon_1, \delta)$ در قسمت اول صدق می‌کند، پس در قسمت دوم هم صدق می‌کند یعنی $m_H(\epsilon_1, \delta) \geq m_H(\epsilon_2, \delta)$. برای $0 < \delta_1 < \delta_2 < 1$ نیز دقیقاً به همین ترتیب است.

2.

a) الگوریتم داده شده برای H singleton به این صورت است که برای هر توزیع D روی X اگر مجموعه آموزشی S با m عضو داشته باشیم $S=\{(x_1,f(x_1)),\dots,(x_m,f(x_m))\}$ در صورتی که یک (x_i, y_i) وجود داشته باشد که $y_i=1$، الگوریتم h_{x_i} را برمی‌گرداند. در غیر این صورت h^- را برمی‌گرداند. با توجه به realizability condition که تنها یک نقطه در X ممکن است وجود داشته باشد که برچسب واقعی آن 1 باشد

واضح است که ~~[illegible]~~ الگوریتم بر روی train set همیشه درست عمل می کند.

یعنی $(L_S(h_S)=0)$. پس الگوریتم ERM است.

b) باید ثابت کنیم به ازای ε و $\delta \in (0,1)$ و

$$P(\{S_x \mid L_{(D,f)}(h_S)>\varepsilon\}) < \delta$$

که D توزیع دلخواه روی X و f تابع true labling است. با توجه به الگوریتم

اگر $f=h_-$ پس صورت خطای واقعی همیشه صفر است و شرط برقرار است

هم چنین اگر $f=h_{x_0}$ و $x_0 \in S$ ، در این حالت باز هم خطای حقیقی

صفر است و شرط برقرار است. در نتیجه باید حالتی را بررسی کنیم که $f=h_{x_0}$ اما

$x_0 \notin S$ و الگوریتم h_- را برمی گرداند یعنی $h_S=h_-$ پس

$$P(\{S_x \mid L_{(D,f)}(h_S)>\varepsilon\}) \leq P(\{x_0 \notin S_x\})$$

چون اعضای S به صورت i.i.d انتخاب می شوند احتمال نبودن x_0 در S_x

برابر است با $1-P(\{x_0\})$

$$\varepsilon < L_{(D,f)}(h_-)$$

$$L_{(D,f)}(h_-) = P(\{h_-(x) \neq f(x)\}) = P(\{x_0\})$$

$$\Longrightarrow \varepsilon < P(\{x_0\}) \longrightarrow 1-P(\{x_0\}) < 1-\varepsilon$$

$$\longrightarrow (1-P(\{x_0\}))^m < (1-\varepsilon)^m \longrightarrow$$ چون m عضو داریم

بنابراین

$$P(\{S_x \mid L_{(D,f)}(h_-) > \varepsilon\}) \leq P(\{x_0 \notin S_x\}) < (1-\varepsilon)^{m-} < e^{-\varepsilon m}$$

$$\Longrightarrow m \geq \frac{\ln(1/\delta)}{\varepsilon} \quad (\mathrm{I})$$

پس به ازای (I) ، خطای حقیقی با احتمال $1-\delta$ از ε کوچکتر است.

---

3. طبق realizability assumption ، یک h^* وجود دارد که دایره‌ای را به L برمی‌گرداند که نقاط داخل دایره برچسب 1 و نقاط خارج آن برچسب 0 دارد.

شعاع دایره را r^* می‌نامیم. الگوریتم A را به گونه‌ای در نظر می‌گیریم که کوچک‌ترین دایره‌ای را با تمام نقاط مثبت مجموعه S در بر می‌گیرد، به ما برمی‌گرداند. A ، ERM است.

شعاع این دایره را r_s می‌نامیم پس $r_s \leq r^*$

$$L_D(h_s) = P_{(x,y)\sim D}(r_s < \|x\| \leq r^*)$$

$$P(\{x : r_s < \|x\| \leq r^*\}) = \varepsilon \quad , \quad E = \{x \in \mathbb{R}^2 , r_s \leq \|x\| < r^*\}$$

احتمال بزرگتر بودن خطای حقیقی از ε برابر است با احتمال این که هیچکدام از training set عضو E نباشد یعنی $(1-\varepsilon)^m$ پس

$$P(\{S_x \mid L_D(h)) > \varepsilon\}) \leq P(\{S_x \mid S_x \cap E = \emptyset\}$$

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

$$\rightarrow m \geq \frac{\ln(1/\delta)}{\varepsilon}$$

و ثابت شد با احتمال $1-\delta$ ، خطای حقیقی از ε کوچکتر است.

---

4. در نخست ثابت می کنیم فرضیات H متناهی است.

هر h می تواند با 3 حالت (x_i ، $\bar{x}_i$ ، هیچ کدام) در Conjunction مشخص شود.

اندازه ی H برابر است با

$$|H| = 3^d + 1$$

چون فرضیات H متناهی است و به ازای $m \geq \frac{\log(|H|/\delta)}{\varepsilon}$ ، برای هر توزیع دلخواه D و تابع true lable ، H PAC leanible است.

$$m \geq \frac{\log(\frac{3^d+1}{\delta})}{\varepsilon} \geq \frac{\log(3^d/\delta)}{\varepsilon}$$

Sample complexity : $m_H(\varepsilon, \delta) \leq \frac{\log(3^d/\delta)}{\varepsilon} = \frac{d(\log 3 + \log(1/\delta)}{\varepsilon}$

6.

اگر مجموعه فرضیات H، Agnostic PAC learnable باشد یعنی

$\exists m_H : (0,1)^2 \to \mathbb{N}$

و یک الگوریتم A داشته باشیم که برای هر $\delta \in (0,1)$ و ϵ و هر توزیع دلخواه D روی $X \times Y$ به ازای $m \geq m_H(\epsilon, \delta)$ نمونه از توزیع D با احتمال $1-\delta$ تابع h را برمی‌گرداند طوری که

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

اکنون اگر فرض کنیم شرط realizability برقرار باشد و تابع f true labling وجود داشته باشد، بنابراین D یک توزیع دلخواه روی $X \times Y$ است طوری که $P_{Y|X}(y|x)$ با f معلوم می‌شود. پس D برابر می‌شود با توزیع روی X. همچنین چون شرط realizability برقرار است $\min_{h' \in H} L_{(D,f)}(h') = 0$ پس با احتمال $1-\delta$

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

7.

$$f_D(x) = \begin{cases} 1 & P[y=1 \mid x] \geq 1/2 \\ 0 & \text{o.w} \end{cases}$$

$$P[f_D(x) \neq y \mid x = x] = 1_{[\alpha_x \geq 1/2]} \cdot P[y=0 \mid x=x] + 1\, P[y=1 \mid x=x]$$

$$= 1_{[\alpha_x \geq 1/2]} \cdot (1-\alpha_x) + 1_{[\alpha_x < 1/2]}\, \alpha_x = \min\{\alpha_x, 1-\alpha_x\}$$

$$\Longrightarrow L_D(f_D) = \min\{\alpha_x, 1-\alpha_x\}$$

y را یک تابع classifier دلخواه فرض می‌کنیم

$$L_D(y) = E_{(x,y)\sim D}\left[1_{y(x)=y}\right] = E_{x\sim D_x}\left[E_{y\sim D_{y|x}}\left[1_{y(x)\neq y}\right]\right]$$

$$= E_{x\sim D_x}\left[P(y(x)\neq y|x)\right]$$

$$P(y(x)\neq y|x) = P(y(x)=0|x)\cdot P(y=1|x) +$$

$$+ P(y(x)=1|x)\, P(y=0|x)$$

$$\geq P(y(x)=0|x)\cdot \min\{\alpha_x, 1-\alpha_x\}$$

$$+ P(y(x)=1|x) = \min\{\alpha_x, 1-\alpha_x\}$$

$$\longrightarrow E_{x\sim D_x}\left[P(y(x)\neq y|x)\right] \geq \min\{\alpha_x, 1-\alpha_x\}$$

$$\longrightarrow L_D(y) \geq L_D(f_D) \quad \blacksquare$$