$\boxed{1.a}$

با توجه به fair بودن تاس برای وظیفه 1 احتمال رخداد هر عدد $\frac{1}{6}$ است

برای وظیفه 2 ابتدا تاس اول پرتاب شده و سپس هر عددی که آمد باید تاس دوم نیز همان عدد بیاید. یعنی

پرتاب اول اهمیتی در میزان سختی ندارد و پرتاب دوم با یک عدد ثابت باید برابر شود. پس احتمال این حالت نیز $\frac{1}{6}$ است

به عبارت دیگر توسط احتمال کل داریم $\longrightarrow$

$$\mathbb{P}(X_1 = X_2) = \sum_{i=1}^{6} \underbrace{\mathbb{P}(X_1 = X_2 \mid X_1 = i)}_{\mathbb{P}(X_2 = i)} \mathbb{P}(X_1 = i) = 6 \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{6}$$

$\boxed{1.b}$

$$\sum_{i=1}^{n} \left(\alpha_i - \frac{1}{n}\right)^2 = \sum_{i=1}^{n} \left(\alpha_i^2 - \frac{2\alpha_i}{n} + \frac{1}{n^2}\right) = \sum_{i=1}^{n} \alpha_i^2 \underbrace{- \frac{2}{n} \underbrace{\sum_{i=1}^{n} \alpha_i}_{1} + \frac{1}{n}}_{-\frac{1}{n}}$$

$$\longrightarrow 0 \leqslant \sum_{i=1}^{n} \left(\alpha_i - \frac{1}{n}\right)^2 = \sum_{i=1}^{n} \alpha_i^2 - \frac{1}{n} \longrightarrow \sum_{i=1}^{n} \alpha_i^2 \geqslant \frac{1}{n} \quad \circledast$$

$\boxed{1.c}$

$$\mathbb{P}(X_1 = X_2) \overset{\text{احتمال کل}}{=\!=} \sum_{i=1}^{n} \underbrace{\mathbb{P}(X_2 = i)}_{\alpha_i} \underbrace{\mathbb{P}(X_1 = i)}_{\alpha_i} = \sum_{i=1}^{n} \alpha_i^2 \overset{\circledast}{\geqslant} \frac{1}{n}$$

به ازای توزیع categorical با همه α_i های برابر $\longrightarrow$  این نابرابری به حداقل مقدار خود یعنی $\frac{1}{n}$ می‌رسد

2.a

$$\mathbb{E}[U] = \frac{1}{\binom{n}{2}} \sum_{j<k} \mathbb{E}[g(X_j, X_k)] \overset{X_i\text{s are iid}}{=\!=\!=} \frac{1}{\binom{n}{2}} \sum_{j<k} \mathbb{E}[g(X_1, X_2)]$$

$$= \frac{1}{\cancel{\binom{n}{2}}} \times \cancel{\binom{n}{2}}\, \mathbb{E}[g(X_1, X_2)]$$

$\longrightarrow$ bias $(U) = \mathbb{E}[U] - \mathbb{E}[g(X_1, X_2)] = 0 \longrightarrow$ unbiased

2.b

متغیر تصادفی $Y_{j,k} \triangleq g(X_j, X_k)$ را تعریف می‌کنیم و چون X_i ها iid هستند پس $Y_{j,k}$ ها نیز iid هستند

و $0 < Y_{j,k} \leq b$: پس می‌توان برای آن‌ها hoeffding نوشت :

$$\mathbb{P}\left(\left| \frac{1}{\binom{n}{2}} \sum_{j<k} Y_{j,k} - \mathbb{E}[Y_{j,k}] \right| \geq t \right) \leq \underbrace{2e^{\frac{-2\binom{n}{2} t^2}{b^2}}}_{\leq e^{\frac{-nt^2}{2b^2}}}$$

3

در گراف G احتمال وجود یا عدم وجود هر یال را با متغیرهای تصادفی $X_1 \dots X_n \overset{iid}{\sim} Bern(p)$ مدل می‌کنیم

و واضح است که $C(G)$ تابعی از X_i هاست.

تغییر هر X_i حداکثر یک واحد می‌تواند $C(G)$ را تغییر دهد $\longrightarrow$ ✓ bounded difference property with $c_i = 1$

$$\text{Mc Diarmid} \longrightarrow \mathbb{P}\left(\left| C(G) - \mathbb{E}[C(G)] \right| \geq \varepsilon \right) \leq 2e^{\frac{-2\varepsilon^2}{\sum c_i^2}} = 2e^{\frac{-2\varepsilon^2}{n}}$$

$$\xrightarrow{\varepsilon = n\delta} \mathbb{P}\left(\frac{1}{n} \left| C(G) - \mathbb{E}[C(G)] \right| \geq \delta \right) \leq 2e^{-2n\delta^2}$$

$\boxed{4}$

⊛ هر $f \in \mathcal{F}_S$ به این صورت عمل می‌کند که یک X_i می‌گیرد و اگر آن ورودی در مجموعه S وجود داشت

1 خروجی می‌دهد و در غیر این صورت 0.

بنابراین اندازه فضای فرضیه $|\mathcal{F}_S|$ بی‌نهایت است.

طبق تعریف uniform convergence برای حالت infinite داریم:

$X_1 \cdots X_n \overset{iid}{\sim} P \quad \forall f_S \in \mathcal{F}_S$

$\forall \delta, \varepsilon > 0 , \quad \exists n_0 = n_0(\delta, \varepsilon)$ such that $\forall n \geq n_0$

$$\mathbb{P}\left(\left| \mathbb{E}_p[f_S(X)] - \frac{1}{n}\sum_{i=1}^{n} f_S(X_i) \right| \leq \varepsilon \right) \geq 1-\delta$$

یعنی به ازای δ , ε فیکس شده باید n_0 ای پیدا شود که از آن جا به بعد کران را داشته باشیم

و P پیوستگی ندارد

حال با توجه به اینکه اعضای S محدود است $\mathbb{E}_p[f_S(X)] = 0$ زیرا مقدار f_S به ازای همه حالات 0 می‌شود

ضمناً می‌توان $X_1 \cdots X_n$ و مجموعه S را جوری انتخاب کرد که تمام X_i ها در S موجود باشند $\longrightarrow 1 = \frac{1}{n}\sum_{i=1}^{n} f_S(X_i)$

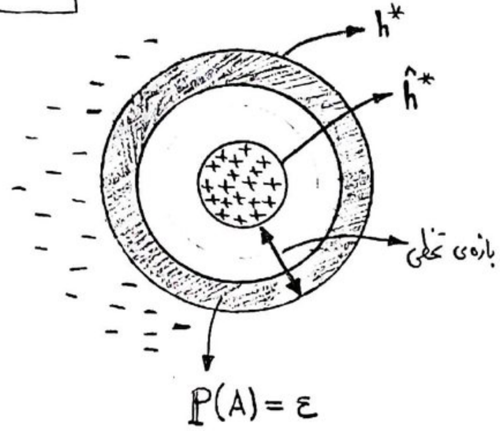
$$\longrightarrow \mathbb{P}(1 \leq \varepsilon) \geq 1-\delta$$

بنابراین مشاهده می‌شود که به ازای هر ε ای باند را نداریم. مثلاً برای $\varepsilon = 0.1$ به تناقض می‌رسیم

پس خانواده توابع $\mathcal{F}_S$ و توزیع P همگرایی یکنواخت ندارد

5

الگوریتم پیشنهادی برای تعیین $\hat{h}^*$ ← کوچکترین دایره به مرکز مبدأ مختصات که نمونه های train set را دربر بگیرد



⊛ طبق صورت سوال حالت PAC است و $R(h^*)=0$

از h^* به اندازه ای به سمت داخل می آییم که احتمال آن بازه برابر ε شود
حال اگر تخطی $\hat{h}^*$ بیشتر از ε باشد به این معناست که هیچ نمونه ای داخل بازه تخطی نخواهد بود (زیرا اگر باشد با فرض الگوریتم پیشنهادی مبنی بر در برگرفتن تمام نمونه های trainset به تناقض می خوریم)

$$\underbrace{\mathbb{P}\left(R(\hat{h}^*) \geq \varepsilon\right)}_{\text{احتمال اینکه تخطی } \hat{h}^* \text{ بیشتر از } \varepsilon \text{ باشد}} = \mathbb{P}\left(\text{هیچ داده ای داخل بازه‌ی تخطی نباشد}\right) \leq \mathbb{P}\left(\text{هیچ داده ای داخل بازه هاشور خورده نباشد}\right) = (1-\delta)^n \leq e^{-n\delta} \leq \delta$$

$$\longrightarrow n \geq \frac{-\lg\delta}{\varepsilon} \longrightarrow n \geq \frac{1}{\varepsilon}\lg\frac{1}{\delta}$$

بنابراین به تعریف PAC رسیدیم و $n \geq poly\left(\frac{1}{\delta},\frac{1}{\varepsilon}\right)$
پس این کلاس از مسائل PAC Learnable هستند

⊛ برای تخطی $\hat{h}^*$ از h^* به سمت بیرون دایره را در نظر نگرفتیم زیرا این حالت با فرض الگوریتم پیشنهادی مبنی بر کوچکترین دایره ای که تمام نمونه های trainset را در بر می گیرد تناقض خواهد داشت و رخ نمی دهد

$\boxed{6.a}$

$$\mathbb{E}\left[\hat{F}_n(t)\right] = \frac{1}{n}\sum_{i=1}^{n} \overbrace{\mathbb{E}\left[\mathbb{1}(X_i \leqslant t)\right]}^{F(t)} = F(t)$$

$$\operatorname{Var}\left[\hat{F}_n(t)\right] \overset{\text{iid}}{=} \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\operatorname{Var}\left[\mathbb{1}(X_i \leqslant t)\right]}_{\mathbb{E}\left[\underbrace{\mathbb{1}(X_i \leqslant t)^2}_{\mathbb{1}(X_i \leqslant t)}\right] - \mathbb{E}\left[\mathbb{1}(X_i \leqslant t)\right]^2} = \frac{F(t) - F(t)^2}{n}$$

بنابراین مشاهده می‌شود که تخمینگر $\hat{F}_n(t)$ بایاس صفر دارد و $\lim_{n\to\infty} \operatorname{Var}\left[\hat{F}_n(t)\right] = 0$.

طبق قضیه ای که از درس آمار داشتیم : بایاس و واریانس به سمت صفر میل کند $\longleftrightarrow$ consistent ✓

بنابراین $\lim_{n\to\infty} \hat{F}_n(t) \overset{a.s.}{=\!=} F(t)$

حال اما با Borel-cantelli نیز اثبات می‌کنیم :

دنباله event های A_n را به صورت روبه رو تعریف می‌کنیم : $A_n \triangleq \left\{ \left| \hat{F}_n(t) - F(t) \right| \geqslant \varepsilon \right\}$

چون $\hat{F}_n(t)$ میانگین تجربی n تا bernoulli R.V. است می‌توان هافدینگ نوشت :

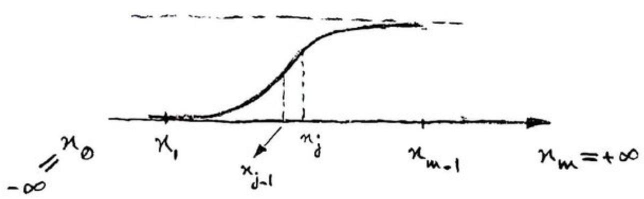
$$\mathbb{P}\left(\left| \hat{F}_n(t) - \overbrace{\mathbb{E}\left[\hat{F}_n(t)\right]}^{F(t)} \right| \geqslant \varepsilon \right) \leqslant 2e^{-2n\varepsilon^2}$$

$$\longrightarrow \sum_{n=1}^{\infty} \mathbb{P}(A_n) \leqslant \sum_{n=1}^{\infty} 2e^{-2n\varepsilon^2} < \infty \longrightarrow$$ می‌توان از Borel-cantelli استفاده کرد

$$\longrightarrow \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0 \longrightarrow$$ مجموعه outcome هایی که ∞ بار تکرار شده اند با احتمال صفر رخ می‌دهد $\longrightarrow$ یعنی n هایی که به ازای آنها تفاوت $\hat{F}_n$ و F زیاد است، ∞ بار رخ نمی‌دهد

$$\longrightarrow \mathbb{P}\left(\lim_{n\to\infty} \left| \hat{F}_n(t) - F(t) \right| = 0\right) = 1 \equiv \lim_{n\to\infty} \hat{F}_n(t) \overset{a.s.}{=\!=} F(t)$$

$\boxed{6.b}$



⊛ x_i ها به طوری انتخاب می‌شن که روی محور y بازه‌ها uniform باشن

$$F(x_{j-1}) \leqslant F(x) \leqslant F(x_j)$$

$\forall x \in \mathbb{R}$

def. $j \in \{1, \dots, m\}$ such that $x \in [x_{j-1}, x_j)$ $\quad m \in \mathbb{N}$

$$\begin{cases} \hat{F}_n(x) - F(x) \leqslant \hat{F}_n(x_j) - F(x_{j-1}) = \hat{F}_n(x_j) - F(x_j) + \frac{1}{m} \\ \hat{F}_n(x) - F(x) \geqslant \hat{F}_n(x_{j-1}) - F(x_j) = \hat{F}_n(x_{j-1}) - F(x_{j-1}) - \frac{1}{m} \end{cases}$$

$$\to \left| \hat{F}_n(x) - F(x) \right| \leqslant \max_{j=1\dots m} \left| \hat{F}_n(x_j) - F(x_j) \right| + \frac{1}{m}$$

$$\to \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right| \leqslant \underbrace{\max_{j=1\dots m}}_{\text{deterministic}} \left| \hat{F}_n(x_j) - F(x_j) \right| + \frac{1}{m}$$

⊛ m رو هم مثلاً $\sqrt{n}$ می‌گیریم و کلش بر اساس تعداد سمپل ها میشه

⊛ این جا چون بازه بندی x_i ها ربطی به x نداره ، میشه جاشون رو عوض کرد پس رسیدن به $\sup_{x \in \mathbb{R}}$ مثل $R(\hat{h}^*)$ نیست که داستان دار بشه

$\forall \varepsilon > 0$ choose $m \in \mathbb{N}$ such that $\frac{1}{m} < \varepsilon$

$$\forall j \quad \mathbb{P}\left(|\hat{F} - F| > \sqrt{\frac{\log \frac{2}{\delta}}{2n}} \right) < \delta$$

$$\xrightarrow[\text{bound}]{\text{union}} \mathbb{P}\left(\max_{j=1 \dots m} |\hat{F}_n(x_j) - F(x_j)| > \sqrt{\frac{\log \frac{2}{\delta}}{2n}} \right) < m\delta$$

rename

$\delta \to \frac{\delta}{m}$

$$\sup_x |\hat{F}_n(x) - F(x)| \overset{1-\delta}{\leqslant} \frac{1}{m} + \sqrt{\frac{\log \frac{2m}{\delta}}{2n}}$$

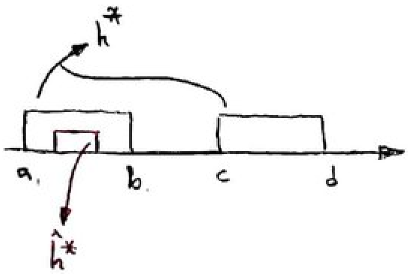
اگر m را توان n مثلاً $\sqrt{n}$ بگذاریم

$$\longrightarrow O\left(\frac{1}{\sqrt{n}}\right) + O\left(\sqrt{\frac{\log \frac{n}{\delta}}{n}}\right) = O\left(\sqrt{\frac{\log \frac{n}{\delta}}{n}}\right)$$

7

الگوریتم پیشنهادی: ابتدا سمپل ها را بر اساس مقادیرشان سورت می‌کنیم و سپس از ابتدا یک دور تمام مقادیر را traverse می‌کنیم و هرجا سمپل + دیدیم مقدار آن را به عنوان شروع بازه در نظر گرفته و تا زمانی که سمپل های + پشت سرهم مشاهده می‌کنیم ادامه می‌دهیم. هرجا این توالی قطع شد مقدار آخرین سمپل + را به عنوان پایان بازه ثبت می‌کنیم (به این صورت p بازه tight به دست می‌آید)

اجتماع این بازه ها خروجی الگوریتم ماست



برای به دست آوردن sample complexity ابتدا حالت $p=2$ را در نظر می‌گیریم.

تخطی $\hat{h}^*$ نسبت به h^* به سه حالت می‌تواند باشد

R1 ⟵ ریسک حالتی که $\hat{h}^*$ بازه کوچکتری داخل $[a,b]$ را انتخاب کرده باشد

R2 ⟵ 〃 〃 〃 〃 $[c,d]$ 〃 〃

R3 ⟵ ریسک حالتی که به دلیل ندیدن سمپل training تمام بازه $[b,c]$ نیز جزء بازه $\hat{h}^*$ بیاید

چون هرکدام از این حالات به صورت مستقل می‌تواند رخ دهد

$$\mathbb{P}(R(\hat{h}^*) \geq \varepsilon) \leq \mathbb{P}(R1 \geq \tfrac{\varepsilon}{3} \cup \underbrace{R2 \geq \tfrac{\varepsilon}{3}}_{\text{مشابه}} \cup R3 \geq \tfrac{\varepsilon}{3})$$

$$\overset{\text{union bound}}{\leq} 2\underbrace{\mathbb{P}(R1 \geq \tfrac{\varepsilon}{3})} + \underbrace{\mathbb{P}(R3 \geq \tfrac{\varepsilon}{3})}$$

$$\leq \mathbb{P}\left(\text{تخطی } \hat{h}^* \text{ از سمت چپ بیشتر از } \tfrac{\varepsilon}{6} \text{ باشد} \cup \text{تخطی } \hat{h}^* \text{ از سمت راست بیشتر از } \tfrac{\varepsilon}{6} \text{ باشد}\right) \qquad \leq \underbrace{\mathbb{P}\left(\text{هیچ نمونه ای در بازه به جرم } \tfrac{\varepsilon}{3} \text{ قرار نگیرد}\right)}_{\underbrace{(1-\frac{\varepsilon}{3})^m}_{\leq e^{-\frac{m\varepsilon}{3}}}}$$

$$\overset{\text{union bound}}{\leq} 2\underbrace{\mathbb{P}\left(\text{هیچ نمونه ای در بازه به جرم } \tfrac{\varepsilon}{6} \text{ قرار نگیرد}\right)}_{\underbrace{(1-\frac{\varepsilon}{6})^m}_{\leq e^{-\frac{m\varepsilon}{6}}}}$$

$$\longrightarrow \mathbb{P}(R(\hat{h}^*) \geq \varepsilon) \leq 4e^{-\frac{m\varepsilon}{6}} + e^{-\frac{m\varepsilon}{3}} \leq 5e^{-\frac{m\varepsilon}{6}} \leq \delta$$

sample complexity
$$m \geq \frac{6}{\varepsilon}\lg\frac{5}{\delta}$$

⊛ اگر دو بازه $[a,b]$ و $[c,d]$ disjoint نبودند داشتیم

$$\mathbb{P}(R(\hat{h}^*) \geq \varepsilon) \leq \mathbb{P}\left(\text{تخطی } \hat{h}^* \text{ از سمت چپ بیشتر از } \tfrac{\varepsilon}{2} \text{ باشد} \cup \text{تخطی } \hat{h}^* \text{ از سمت راست بیشتر از } \tfrac{\varepsilon}{2} \text{ باشد}\right)$$

$$\overset{\text{union bound}}{\leq} 2(1-\tfrac{\varepsilon}{2})^m \leq 2e^{-\frac{m\varepsilon}{2}} \longrightarrow$$

sample complexity
$$m \geq \frac{2}{\varepsilon}\lg\frac{2}{\delta}$$

7 ادامه

(*) چون sample complexity حالت disjoint بیشتر است آن را درنظر می‌گیریم.

در حالت کلی که p بازه داشته باشیم $p-1$ تا ریسک از جنس $R3$ داریم و $2p$ تا ریسک از جنس $R1$.

با تقسیم کردن ε به تعداد این بازه‌ها خواهیم داشت:

$$\mathbb{P}\left(R(\hat{h}^*) \geq \varepsilon\right) \leq (2p)\, e^{-\frac{m\varepsilon}{4p-2}} + (p-1)\, e^{-\frac{n\varepsilon}{2p-1}} \leq (3p-1)\, e^{-\frac{m\varepsilon}{4p-2}} \leq \delta$$

$\longrightarrow$ sample complexity : $m \geq \frac{4p-2}{\varepsilon} \lg \frac{3p-1}{\delta}$

پیچیدگی زمانی الگوریتم نیز به خاطر مرتب سازی اولیه $O(m \lg m)$ است

8.a

با توجه به اینکه $|x_i| \leqslant M$ تابع f خاصیت bounded difference property را دارد و هر x_i حداکثر $\frac{2M}{m}$ تغییر ایجاد می‌کند

$$\text{Mc Diarmid} \longrightarrow \mathbb{P}\left(\left|f(x) - \mathbb{E}[f(x)]\right| \geqslant \varepsilon\right) \leqslant 2e^{\frac{-2\varepsilon^2}{\sum c_i^2}}$$

$$\xrightarrow{c_i = \frac{2M}{m}} \mathbb{P}\left(\left|f(x) - \mathbb{E}[f(x)]\right| \geqslant \varepsilon\right) \leqslant 2e^{\frac{-2\varepsilon^2}{m \times \frac{4M^2}{m^2}}} = 2e^{\frac{-m\varepsilon^2}{2M^2}}$$

8.b

با توجه به اینکه h خاصیت bounded difference property را دارد می‌توان Mc Diarmid نوشت.

فرض می‌کنیم $g(m) \in O(m^k)$ باشد و می‌خواهیم k را به نحوی پیدا کنیم که کران Mc Diarmid با افزایش m به صفر همگرا شود

$$\xrightarrow{c_i = g(m) \in O(m^k)} \mathbb{P}\left(\left|h(x) - \mathbb{E}[h(x)]\right| \geqslant \varepsilon\right) \leqslant 2e^{\frac{-2\varepsilon^2}{\sum m^{2k}}} = 2e^{\frac{-2\varepsilon^2}{m \times m^{2k}}}$$

$$= 2e^{-2\varepsilon^2 m^{-2k-1}}$$

برای اینکه این کران با افزایش m به 0 میل کند باید داشته باشیم $-2k-1 > 0 \longrightarrow k < \frac{-1}{2}$

یعنی سرعت نزول $g(m)$ باید بیشتر از $O\left(\frac{1}{\sqrt{m}}\right)$ باشد مثلاً $O\left(m^{-\left(\frac{1}{2}+\delta\right)}\right)$ اوکی است

$\forall \delta > 0$

8.c

بله تابع f پایدار است زیرا $g(m) = \frac{2M}{m}$ که نزولی است

چون $g(m) \in O\left(\frac{1}{m}\right)$ طبق قسمت قبل می‌دانیم که کران همگرا به صفر می‌شود

$$\xrightarrow[c_1 = \frac{2M}{m}]{\text{Mc Diarmid}} \mathbb{P}\left(\left|f(x) - \mathbb{E}[f(x)]\right| \geqslant \varepsilon\right) \leqslant 2e^{\frac{-2\varepsilon^2}{m \times \frac{4M^2}{m^2}}} = 2e^{\frac{-2m\varepsilon^2}{4M^2}}$$

خیر f' با افزایش m، $g(m)$ اش لزوماً نزولی نیست و بنابراین پایدار نیست

در این جا حداکثر $2M$ تغییر می‌تواند ایجاد شود

$$\xrightarrow[c_i = 2M]{\text{Mc Diarmid}} \mathbb{P}\left(\left|f'(x) - \mathbb{E}[f'(x)]\right| \geqslant \varepsilon\right) \leqslant 2e^{\frac{-2\varepsilon^2}{m \times 4M^2}}$$

همان طور که مشاهده می‌شود با افزایش m

کران بهتر نمی‌شود و چیز به درد بخوری برای f' نیست

$\boxed{9}$

اگر یک مسئله در معنای کلاسیک (تک اوراکل) به طور کارآمد قابل یادگیری PAC باشد (لاس = zero-one)

* چون به طور کارآمد PAC learnable هستیم پس قطعاً یک $\hat{h}^*$ ای توسط این الگوریتم به دست ما می‌رسد و تخطی آن‌را $\frac{\varepsilon}{2}$ در نظر می‌گیریم

$$R(\hat{h}^*) \overset{1-\delta}{\leqslant} \frac{\varepsilon}{2} \qquad n \geq \mathrm{poly}\left(\frac{1}{\delta}, \frac{1}{\varepsilon}\right)$$

$$\rightarrow \mathbb{E}_x\left[\ell(\hat{h}^*(x), h^*(x))\right] \overset{1-\delta}{\leqslant} \varepsilon/2$$

$$\xrightarrow[\text{loss}]{\text{zero-one}} \mathbb{P}_x\left(\hat{h}^*(x) \neq h^*(x)\right) \overset{1-\delta}{\leqslant} \varepsilon \longrightarrow \mathbb{P}_x\left(\hat{h}^*(x) = 1,\ h^*(x) = 0\right) + \mathbb{P}_x\left(\hat{h}^*(x) = 0,\ h^*(x) = 1\right) \overset{1-\delta}{\leqslant} \varepsilon/2$$

$$\longrightarrow \mathbb{P}_x\left(\hat{h}^*(x) = 1 \mid h^*(x) = 0\right) \mathbb{P}_x\left(h^*(x) = 0\right) + \mathbb{P}_x\left(\hat{h}^*(x) = 0 \mid h^*(x) = 1\right) \mathbb{P}_x\left(h^*(x) = 1\right) \overset{1-\delta}{\leqslant} \frac{\varepsilon}{2}$$

حال اگر فرض کنیم در مدل دو اوراکل از هرکدام توزیع‌ها با احتمال برابر سمپل می‌گیریم خواهیم داشت ⟵

$$\mathbb{P}_x\left(h^*(x) = 0\right) = \mathbb{P}_x\left(h^*(x) = 1\right) = \frac{1}{2}$$

$$\longrightarrow \mathbb{P}_x\left(\hat{h}^*(x) = 1 \mid h^*(x) = 0\right) + \mathbb{P}_x\left(\hat{h}^*(x) = 0 \mid h^*(x) = 1\right) \overset{1-\delta}{\leqslant} \varepsilon$$

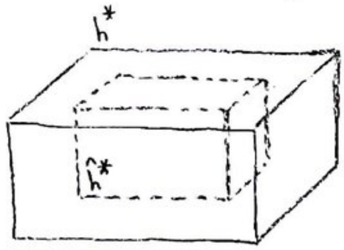
$$\longrightarrow \begin{array}{l} \mathbb{P}_x\left(\hat{h}^*(x) = 1 \mid h^*(x) = 0\right) \overset{1-\delta}{\leqslant} \varepsilon \\ , \\ \mathbb{P}_x\left(\hat{h}^*(x) = 0 \mid h^*(x) = 1\right) \overset{1-\delta}{\leqslant} \varepsilon \end{array} \longrightarrow \begin{array}{l} \mathbb{P}_{x \sim D^-}\left(\hat{h}^*(x) = 1\right) \overset{1-\delta}{\leqslant} \varepsilon \\ \\ \mathbb{P}_{x \sim D^+}\left(\hat{h}^*(x) = 0\right) \overset{1-\delta}{\leqslant} \varepsilon \end{array}$$

بنابراین از فرض efficiently PAC learnable بودن در حالت کلاسیک به تعریف efficiently PAC learnable بودن در حالت دو اوراکل رسیدیم ✓

10

الگوریتم پیشنهادی $\longleftarrow$ کوچکترین مستطیل n بعدی که داده های trainset را در بر بگیرد ، در نظر می گیریم

اثبات PAC learnability $\longleftarrow$ مسئله مستطیل های 2 بعدی که سر کلاس حل شد را تعمیم می دهیم :



می دانیم که هر مستطیل axis aligned که n بعدی باشند $2n$ تا وجه دارند
فرض می کنیم حجم ناحیه داخل h^* بیشتر از ε باشد چون اگه کمتر باشد already
باند مدنظر را داریم . حال از هر وجه h^* به اندازه ای رو به داخل می آییم که
جرم احتمال آن ناحیه $\frac{\varepsilon}{2n}$ شود ، احتمال تخطی کل $\hat{h}^*$ نسبت به h^* کمتر مساوی
احتمال اجتماع تخطی هر وجه به اندازه $\frac{\varepsilon}{2n}$ است . پس هر وجه را به صورت مستقل
و جدا در نظر می گیریم و می توان نوشت :

تخطی هر وجه به اندازه حداقل $\frac{\varepsilon}{2n}$

$$\mathbb{P}\left(R(\hat{h}^*) \geqslant \varepsilon\right) \leqslant \mathbb{P}\left(\bigcup_{i=1}^{2n} E_i\right) \overset{\text{union bound}}{\leqslant} \sum_{i=1}^{2n} \underbrace{\mathbb{P}(E_i)}_{\leqslant \underbrace{\mathbb{P}\left(\text{هیچ داده ای در بازه به جرم } \frac{\varepsilon}{2n} \text{ نباشد}\right)}_{\left(1-\frac{\varepsilon}{2n}\right)^m}}$$

$$\rightarrow \mathbb{P}\left(R(\hat{h}^*) \geqslant \varepsilon\right) \leqslant 2n\left(1-\frac{\varepsilon}{2n}\right)^m \leqslant 2n\, e^{-\frac{m\varepsilon}{2n}} \leqslant \delta$$

$$\rightarrow \text{for } m \geqslant \frac{2n}{\varepsilon} \lg \frac{2n}{\delta} \qquad R(\hat{h}^*) \overset{1-\delta}{\geqslant} \varepsilon$$

در برخی سوال ها با علی بابابیگ همفکری داشتم