تمرین‌های سری ۶

فاطمه رحیمی

سوال ۶. ۲:

۱. ادعا می‌کنیم $VC\dim(H_k)=\min\{|X|-k,k\}$. برای نشان دادن این ادعا با فرض $m=\min\{|X|-k,k\}$ دو مورد را باید اثبات کنیم:

الف) هیچ $m+1$ نقطه‌ای توسط H_k خرد نمی‌شود.

ب) وجود دارد m نقطه که توسط H_k خرد می‌شود.

اثبات الف. فرض کنید $m=k$ و $m+1$ نقطه داریم، در این صورت چون می‌دانیم H_k دقیقا k نقطه را ۱ می‌کند، حالتی که در آن همه‌ی $k+1$ نقطه برابر ۱ باشند را نخواهیم داشت، پس $m+1$ نقطه خرد نمی‌شود. همچنین اگر فرض کنید $m=|X|-k$ در این صورت نیز چون H_k دقیقا $|X|-k$ نقطه را 0 می‌کند، حالتی که در آن همه‌ی $|X|-k+1$ نقطه برابر 0 باشند را نخواهیم داشت، پس باز هم $m+1$ خرد نمی‌شود.

اثبات ب. $C=\{x_1,x_2,...,x_m\}$ را نقاطی در نظر بگیرید که قرار است توسط H_k آن‌ها را خرد کنیم، همچنین $(y_1,y_2,...,y_m)$ را برچسب‌گذاری دلخواه این نقاط در نظر بگیرید، می‌دانیم در هر برچسب‌گذاری تعدادی 1 و 0 داریم، تعداد یک‌های هر برچسب‌گذاری را با s نشان می‌دهیم. حال با در نظر گرفتن $C^c=X/C$ و انتخاب زیر مجموعه‌ای $k-s$ عضوی از آن چون E می‌توانیم $h\in H_k$ را به این ترتیب تعریف کنیم:

$$\begin{cases} h(x_i)=y_i & x_i\in C \\ h(x)=1 & x\in E \\ h(x)=0 & o.w \end{cases}$$

روشن است که هر برچسب‌گذاری دلخواه را با این روش میتوان روی m نقطه داشت. پس ادعا ثابت می‌شود و داریم $VC\dim(H_k)=\min\{|X|-k,k\}$.

۲. ادعا می‌کنیم $VC\dim(H_{at-most-k})=k$. برای نشان دادن این ادعا دو مورد را باید اثبات کنیم:

الف) هیچ $k+1$ نقطه‌ای توسط $H_{at-most-k}$ خرد نمی‌شود.

ب) وجود دارد k نقطه که توسط $H_{at-most-k}$ خرد می‌شود.

اثبات الف. فرض کنید $k+1$ نقطه داریم، در این صورت چون می‌دانیم $H_{at-most-k}$ حداکثر k نقطه را ۱ می‌کند، حالتی که در آن همه‌ی $k+1$ نقطه برابر ۱ باشند را نخواهیم داشت، پس $k+1$ نقطه خرد نمی‌شود.

اثبات ب. $C=\{x_1,x_2,...,x_m\}$ را نقاطی در نظر بگیرید که قرار است توسط $H_{at-most-k}$ آن‌ها را خرد کنیم، همچنین $(y_1,y_2,...,y_m)$ را برچسب‌گذاری دلخواه این نقاط در نظر بگیرید، می‌توانیم $h\in H_k$ را به این ترتیب تعریف کنیم:

$$\begin{cases} h(x_i)=y_i & x_i\in C \\ h(x)=0 & o.w \end{cases}$$

روشن است که هر برچسب‌گذاری دلخواه را با این روش میتوان روی k نقطه داشت. پس ادعا ثابت می‌شود و داریم $VC\dim(H_{at-most-k})=k$.

سوال ۶. ۴:

با فرض $X=\mathbb{R}^d$ می‌خواهیم حالت‌های اکید و تساوی نامساوی‌ها را روی $X\times\{0,1\}$ نشان دهیم.

- فرض کنید $d\geq 2$ و کلاس H را تمام گوی‌های هم مرکز در نظر بگیرید. در واقع هر $h_r\in H$ به این ترتیب است که به تمام xهایی که $\|x\|_2\leq r$ مقدار ۱ و در غیر اینصورت مقدار 0 می‌دهد. حال روشن است که $VC\dim(H)=1$ چرا که یک نقطه را می‌تواند خرد کند، به این ترتیب که این نقطه داخل یا خارج گوی باشد. ولی هیچ دو نقطه‌ای را نمی‌تواند خرد کند به این ترتیب که اگر $\|x_i\|_2\leq\|x_j\|_2$ هیچوقت برچسب‌گذاری $y_j=1$ و $y_i=0$ را نمی‌توانیم با اعضای H تولید کنیم. حال کافی است A را زیر مجموعه دو عضوی دلخواهی از دامنه در نظر بگیریم. با توجه به تعریف H روشن است که $|H_A|=2$، همچنین زیر مجموعه‌هایی از A که خرد می‌شوند تک عضوی‌ها و تهی هستند، بنابراین داریم:

$$\underbrace{|H_A|}_{2}<\underbrace{|\{B\subseteq A: H \text{ shatter } B\}|}_{3}=\underbrace{\sum_{i=1}^{d}\binom{|A|}{i}}_{3}$$

- این بار کلاس H را تمام مستطیل‌هایی که اضلاع موازی با محورهای مختصات در $\mathbb{R}^2$ دارند در نظر بگیرید. می‌دانیم $VC\dim(H)=4$. حال با در نظر گرفتن $A=\{(0,1),(0,2),(0,3)\}$ تمام برچسب‌گذاری‌ها به جز $(1,0,1)$ را خواهیم داشت، بنابراین:

$$|H_A|=2^3-1=7$$

$$|\{B\subseteq A: H \text{ shatter } B\}|=2^3-1=7$$

$$\sum_{i=0}^{d}\binom{|A|}{i}=\binom{3}{0}+\binom{3}{1}+\binom{3}{2}+\binom{3}{3}=1+3+3+1=8$$

$$\underbrace{|H_A|}_{7}=\underbrace{|\{B\subseteq A: H \text{ shatter } B\}|}_{7}\leq\underbrace{\sum_{i=1}^{d}\binom{|A|}{i}}_{8}$$

سوال ۹. ۶:

ادعا می‌کنیم $VC\dim(H_{a,b,s})=3$. برای نشان دادن این ادعا دو مورد را باید اثبات کنیم:

الف) هیچ 4 نقطه‌ای توسط $H_{a,b,s}$ خرد نمی‌شود.

ب) وجود دارد 3 نقطه که توسط $H_{a,b,s}$ خرد می‌شود.

- اثبات الف. چهار نقطه‌ی دلخواه را به صورت $\{x_1,x_2,x_3,x_4\}$ در نظر می‌گیریم. بدون از دست رفتن کلیت می‌توانیم فرض کنیم $x_1<x_2<x_3<x_4$ ، حال اگر بخواهیم برچسب‌گذاری روی این مجموعه در نظر بگیریم که برچسب نقاط یکی در میان با هم متفاوت باشد با هیچکدام از اعضای $H_{a,b,s}$ این کار ممکن نیست. پس $VC\dim(H_{a,b,s})\leq 3$.
- اثبات ب. سه نقطه دلخواه را به صورت $\{1,2,3\}$ در نظر بگیرید. این سه نقطه ۸ برچسب‌گذاری متفاوت دارند، کافی است مقادیر a ، b و s را به صورت‌های زیر قرار دهیم تا همه‌ی حالت‌ها تولید شود:

1	2	3	a	b	s
+	+	+	0.5	3.5	1
−	+	+	1.5	3.5	1
+	−	+	1.5	2.5	−1
+	+	−	0.5	2.5	1
−	−	+	2.5	3.5	1
−	+	−	1.5	2.5	1
+	−	−	0.5	1.5	1
−	−	−	0.5	3.5	−1

بنابراین $VC\dim(H_{a,b,s})\geq 3$ و با توجه به اثبات الف می‌توان نتیجه گرفت $VC\dim(H_{a,b,s})=3$.

سوال ۱۰. ۶:

- با توجه به صورت سوال می‌دانیم $VC\dim(H)\geq d$، پس وجود دارد مجموعه C با اندازه‌ی d به طوری که توسط H خرد شود. بدون از دست رفتن کلیت می‌توانیم فرض کنیم $X=C$، برای این حالت توزیعی روی X را در نظر بگیرید که در خارج از C مقداری نمی‌گیرد. حال با توجه به این که

H شامل تمام توابع از C به $\{0,1\}$ است و با توجه به تمرین فصل ۵، برای همه‌ی الگوریتم‌ها توزیعی چون D وجود دارد که $\min_{h\in H} L_D(h)=0$ ولی $E[L_D(A(S))]\geq \frac{k-1}{2k}$ ، که در آن :

$$|X|=|C|=d=km \Rightarrow k=\frac{d}{m}$$

$$\Rightarrow E[L_D(A(S))]\geq \frac{\frac{d}{m}-1}{\frac{2d}{m}}=\frac{d-m}{2d}$$

- فرض کنید $VC\dim(H)=\infty$ می‌خواهیم نشان دهیم در این صورت H نمی‌تواند PAC Learnable باشد. چون $VC\dim(H)=\infty$ پس می‌توانیم اندازه‌ی مجموعه‌ای که توسط آن خرد می‌شود را به اندازه دلخواه بزرگ در نظر بگیریم، اگر اندازه این مجموعه را به صورت $d=2m$ در نظر بگیریم، با توجه به قسمت قبلی می‌توانیم بگوئیم توزیعی چون D وجود دارد که $\min_{h\in H} L_D(h)=0$ ولی $E[L_D(A(S))]\geq \frac{1}{4}$. بنابراین با توجه قضیه ۶,۴ با احتمال حداقل $\frac{1}{7}$ داریم:

$$L_D(A(S))-\min_{h\in H} L_D(h)=L_D(A(S))\geq \frac{1}{8}$$

حال با در نظر گرفتن $\varepsilon=\frac{1}{9}$ و $\delta=\frac{1}{8}$ روشن است که H، PAC Learnable نمی‌باشد.

سوال ۶.۱۱:

- می‌دانیم اگر $VC\dim(\bigcup_{i=1}^{r} H_i)=k$، آن گاه $\tau_{\bigcup_{i=1}^{r} H_i}(k)=2^k$، با توجه به تعریف تابع رشد داریم:

$$\tau_{\bigcup_{i=1}^{r} H_i}(k)\leq \sum_{i=1}^{r}\tau_{H_i}(k)$$

اکنون با استفاده از لم Sauer روی هر کدام از جملات τ_{H_i} داریم:

$$\tau_{\bigcup_{i=1}^{r} H_i}(k)\leq rm^d$$

اکنون با لگاریتم گرفتن از طرفین داریم:

$$k < d\log m + \log r$$

اکنون به سادگی با استفاده از لم A.2 در راهنمایی مسئله داریم:

$$k < 4d\log(2d) + 2\log r$$

- بدون از دست رفتن کلیت فرض کنید $VC\dim(H_1) = VC\dim(H_2) = d$ و k را به طوری در نظر بگیرید که $k < 2d+1$، حال داریم:

$$\begin{aligned}
&\tau_{H_1 \bigcup H_2} \le \tau_{H_1} + \tau_{H_2} \\
&\le \sum_{i=1}^{d}\binom{k}{i} + \sum_{i=1}^{d}\binom{k}{i} \\
&= \sum_{i=1}^{d}\binom{k}{i} + \sum_{i=k-d}^{k}\binom{k}{i} \\
&\le \sum_{i=1}^{d}\binom{k}{i} + \sum_{i=d+2}^{k}\binom{k}{i} \\
&< \sum_{i=1}^{d}\binom{k}{i} + \sum_{i=d+1}^{k}\binom{k}{i} = \sum_{i=1}^{k}\binom{k}{i} = 2^k
\end{aligned}$$

بنابراین $VC\dim(H_1 \bigcup H_2) \le k \le 2d+1$.