بسمه تعالی

پاسخ سری دوم تمرینات درس یادگیری ماشین

امیرحسین رمضانی بناب (۹۹۲۱۰۲۹۴)

۱ سوال اول

۱.۱ الف

توزیع داده شده در صورت سوال را با دانستن دودویی بودن f به شکل زیر تغییر می‌دهیم:

$$\mathbb{P}(y|x) = \begin{cases} \lambda & y = f(x) \\ 1-\lambda & y \neq f(x) \end{cases} = \begin{cases} \lambda & y = f(x) \\ 1-\lambda & y = 1 - f(x) \end{cases}$$

می‌خواهیم احتمال اینکه $h(x)$ در تخمین y دچار خطا شود را تعیین کنیم. اگر $y = f(x)$ باشد در آن صورت طبق فرض داده شده در صورت سوال فرضیه‌ی h به احتمال μ دچار خطا می‌شود و اگر $y = 1 - f(x)$ باشد، فرضیه‌ی h در صورتی دچار خطا می شود که مقدار $f(x)$ را به درستی تخمین زده باشد (یا به عبارتی در تخمین مقدار $y = 1 - f(x)$ اشتباه کند) که احتمال آن برابر $1 - \mu$ است. به شکل صوری داریم :

$$\begin{aligned} \mathbb{P}(h(x) \neq y) &= \mathbb{P}(h(x) \neq y, y = f(x)) + \mathbb{P}(h(x) \neq y, y = 1 - f(x)) \\ &= \mathbb{P}(h(x) \neq y | y = f(x))\mathbb{P}(y = f(x)) + \mathbb{P}(h(x) \neq y | y = 1 - f(x))\mathbb{P}(y = f(x)) \\ &= \mu\lambda + (1-\mu)(1-\lambda) \\ &= \mu\lambda + 1 - \mu - \lambda + \mu\lambda \\ &= 2\mu\lambda - \mu - \lambda + 1 \end{aligned}$$

۲.۱ ب

برای اینکه μ را در محاسبه‌ی میزان خطای h بی اثر کنیم، در معادله‌ی بالا از آن فاکتور میگیریم:

$$\mathbb{P}(h(x) \neq y) = 2\mu\lambda - \mu - \lambda + 1 = \mu(2\lambda - 1) - \lambda + 1$$

کافی است ضریب μ را در معادله‌ی فوق صفر کنیم. داریم :

$$2\lambda - 1 = 0 \Rightarrow \lambda = 0.5$$

۲ سوال دوم

۱.۲ الف

۱.۱.۲ مساله‌ی سوپرمارکت

برای ماتریس ریسک سوپرمارکت داریم :

$$
\begin{aligned}
E_{in}(h) &= \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), f(x_n)) \\
&= \frac{1}{N}(\sum_{n=1}^{N} 0\ \times [[h(x_n) = +1]] \times [[f(x_n) = +1]] \\
&+ \quad \sum_{n=1}^{N} 1\ \times [[h(x_n) = +1]] \times [[f(x_n) = -1]] \\
&+ \quad \sum_{n=1}^{N} 10 \times [[h(x_n) = -1]] \times [[f(x_n) = +1]] \\
&+ \quad \sum_{n=1}^{N} 0\ \times [[h(x_n) = -1]] \times [[f(x_n) = -1]])
\end{aligned}
$$

که برابر است با :

$$
\begin{aligned}
E_{in}(h) &= \frac{1}{N}(\sum_{n=1}^{N} 1\ \times [[h(x_n) = +1]] \times [[f(x_n) = -1]] \\
&+ \quad \sum_{n=1}^{N} 10 \times [[h(x_n) = -1]] \times [[f(x_n) = +1]])
\end{aligned}
$$

که در صورت ساده سازی بر اساس $y = f(x)$ می‌توان به شکل زیر نیز نوشت :

$$
E_{in}(h) = \frac{1}{N}(10 \sum_{y_n=+1} [[h(x_n) = -1]] + \sum_{y_n=-1} [[h(x_n) = +1]])
$$

۲.۱.۲ مساله‌ی CIA

برای ماتریس ریسک CIA داریم :

$$
\begin{aligned}
E_{in}(h) &= \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), f(x_n)) \\
&= \frac{1}{N}(\sum_{n=1}^{N} 0 \quad\; \times [[h(x_n) = +1]] \times [[f(x_n) = +1]] \\
&+ \quad \sum_{n=1}^{N} 1000 \times [[h(x_n) = +1]] \times [[f(x_n) = -1]] \\
&+ \quad \sum_{n=1}^{N} 1 \quad\; \times [[h(x_n) = -1]] \times [[f(x_n) = +1]] \\
&+ \quad \sum_{n=1}^{N} 0 \quad\; \times [[h(x_n) = -1]] \times [[f(x_n) = -1]])
\end{aligned}
$$

که برابر است با :

$$
\begin{aligned}
E_{in}(h) &= \frac{1}{N}(\sum_{n=1}^{N} 1000 \times [[h(x_n) = +1]] \times [[f(x_n) = -1]] \\
&+ \quad \sum_{n=1}^{N} 1 \quad\; \times [[h(x_n) = -1]] \times [[f(x_n) = +1]])
\end{aligned}
$$

که در صورت ساده سازی بر اساس $y = f(x)$ می‌توان به شکل زیر نیز نوشت :

$$
E_{in}(h) = \frac{1}{N}(\sum_{y_n=+1} [[h(x_n) = -1]] + 1000 \sum_{y_n=-1} [[h(x_n) = +1]])
$$

۲.۲ ب

ابتدا با استفاده از تعریف $Expected\ Value$ عبارت E_{out} را باز می‌کنیم.

$$\mathbb{E}_{out}(h) = \mathbb{E}[(h(x)-y)^2] = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbb{P}(x,y)(h(x)-y)^2$$

حال از این عبارت نسبت به $h(x)$ مشتق می‌گیریم تا آن را کمینه کنیم.

$$\begin{aligned}\frac{\partial\,\mathbb{E}_{out}(h)}{\partial\,h} &= 2\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbb{P}(x,y)(h(x)-y)\\ &= 2\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathbb{P}(x)\mathbb{P}(y|x)(h(x)-y)\\ &= 2\sum_{x\in\mathcal{X}} \mathbb{P}(x)\big[\sum_{y\in\mathcal{Y}} \mathbb{P}(y|x)(h(x)-y)\big]\end{aligned}$$

برای اینکه این عبارت برابر صفر شود کافی است عبارت درون براکت برابر صفر شود :

$$\sum_{y\in\mathcal{Y}} \mathbb{P}(y|x)(h(x)-y) = 0 \Rightarrow \underbrace{\sum_{y\in\mathcal{Y}} \mathbb{P}(y|x)h(x)}_{h(x)} = \underbrace{\sum_{y\in\mathcal{Y}} \mathbb{P}(y|x)y}_{\mathbb{E}(y|x)}$$

پس پاسخ بهینه به شکل زیر خواهد بود :

$$h^*(x) = \mathbb{E}(y|x)$$

برای بخش دوم سوال داریم :

$$\begin{aligned}\epsilon(X) = y - h^*(x) \Rightarrow \mathbb{E}(\epsilon(x)) &= \mathbb{E}(y - h^*(x))\\ &= \mathbb{E}(y) - \mathbb{E}(h^*(x))\\ &= \mathbb{E}(y) - \mathbb{E}(\mathbb{E}(y|x))\\ (according\ to\ properties\ of\ conditional\ expectation)\quad &= \mathbb{E}(y) - \mathbb{E}(y)\\ &= 0\end{aligned}$$

۳ سوال سوم

۱.۳ آ

برای حالت $N = 4$ پازل به شکل زیر خواهد بود:

x_1	x_2	x_3	x_4
-1	-1	-1	-1
-1	-1	-1	$+1$
-1	-1	$+1$	-1
-1	-1	$+1$	$+1$
-1	$+1$	-1	-1
-1	$+1$	-1	$+1$
-1	$+1$	$+1$	-1
-1	$+1$	$+1$	$+1$
$+1$	-1	-1	-1
$+1$	-1	-1	$+1$
$+1$	-1	$+1$	-1
$+1$	-1	$+1$	$+1$
$+1$	$+1$	-1	-1
$+1$	$+1$	-1	$+1$
$+1$	$+1$	$+1$	-1
$+1$	$+1$	$+1$	$+1$

همانطور که مشخص است ۱۱ تفکیک مختلف وجود دارد و بقیه‌ی نقاط در ستون های مشخص شده باعث تشکیل همه سه تایی های ممکن می‌شوند. حال به سراغ دسته بندی ارائه شده در کلاس می‌رویم.

name	*count*	x_1	x_2	x_3	x_4
S_1	$\alpha = 3$	-1	$+1$	$+1$	-1
		$+1$	-1	$+1$	-1
		$+1$	$+1$	-1	-1
S_2^+	$\beta = 4$	-1	-1	-1	$+1$
		-1	-1	$+1$	$+1$
		-1	$+1$	-1	$+1$
		$+1$	-1	-1	$+1$
S_2^-	$\beta = 4$	-1	-1	-1	-1
		-1	-1	$+1$	-1
		-1	$+1$	-1	-1
		$+1$	-1	-1	-1

حال داریم :

$$B(N,k) = B(4,3) \overset{definition}{=} \alpha + 2\beta = 11$$
$$B(N-1,k) = B(3,3) \overset{obvious}{=} 7 \geq \alpha + \beta$$
$$B(N-1,k-1) = B(3,2) \overset{puzzle}{=} 4 \geq \beta$$

پس :

$$\alpha + \beta \leq B(4-1, 3-1)$$
$$\beta \leq B(4-1, 3)$$

اگر دو طرف نامساوی فوق را جمع کنیم داریم :

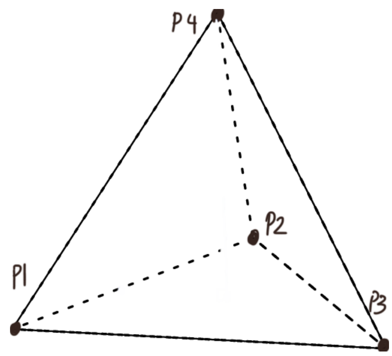
$$\alpha + 2\beta \leq B(4-1, 3-1) + B(4-1, 3)$$

که چون $\alpha + 2\beta = B(4, 3)$:

$$B(4, 3) \leq B(4-1, 3-1) + B(4-1, 3)$$

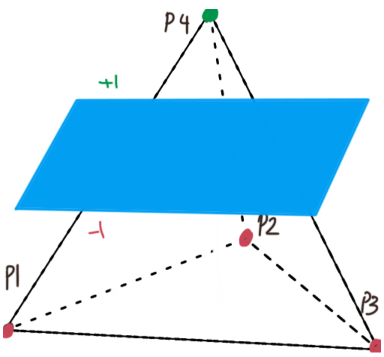
۲.۳ ب

در فضای پرسپترون سه بعدی با کمتر از ۳ نقطه به سادگی می‌توان تحقیق کرد که می‌توانیم به بیشترین تعداد تفکیک ممکن (2^N) برسیم. برای ۴ نقطه بهترین حالت این است که تا حد ممکن این نقاط را دور از هم نگه داریم و به این صورت نقطه‌ی چهارم را در نقطه‌ای از فضا قرار می‌دهیم که روی صفحه گذرنده از ۳ نقطه‌ی دیگر نباشد. و سه نقطه‌ی دیگر را دور از هم قرار می‌دهیم تا تشکیل یک مثلث متساوی الاضلاع دهند که در کل، چهار نقطه تشکیل یک چهاروجهی منتظم می‌دهند.

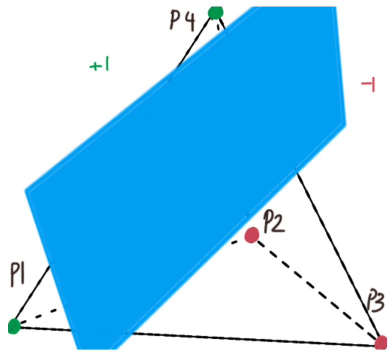


حالت های غیرمشابه را بررسی میکنیم:

- حالتی که همه‌ی نقاط برچسب -1 داشته باشند : در این حالت به وضوح کافی است صفحه‌ای را در فضا رسم کنیم که همه‌ی ۴ نقطه در سمتی از آن باشند که برچسب -1 می‌خورند
- حالتی که نقطه‌ی P_4 برچسب $+1$ داشته باشد و بقیه برچسب -1 : در این حالت کافی است صفحه‌ای را چنان رسم کنیم که P_4 در سمتی از آن قرار داشته باشد که برچسب -1 می‌خورد و بقیه در سمت دیگر.



- حالتی که نقطه‌های P_4 و P_1 برچسب $+1$ داشته باشد:



بقیه‌ی حالات نیاز به بررسی نداشته و قابل تبدیل به یکی از این ۳ حالت می‌باشند. پس برعکس فضای دوبعدی، در فضای سه بعدی با ۴ نقطه می‌توانیم به همه پترن‌های ممکن برسیم.
اما وقتی تعداد نقاط به ۵ افزایش پیدا می‌کند دیگر امکان تولید همه‌ی پترن‌های ممکن وجود ندارد. به عنوان مثال در همان شکل بالا اگر نقطه‌ی P_5 را به نحوی اضافه کنیم تا تعداد دایکاتمی‌ها بیشینه بماند، باز هم امکان تولید برخی پترن‌ها وجود ندارد. پس نقطه‌ی شکست پرسپترون سه بعدی ۵ است. در نتیجه داریم :

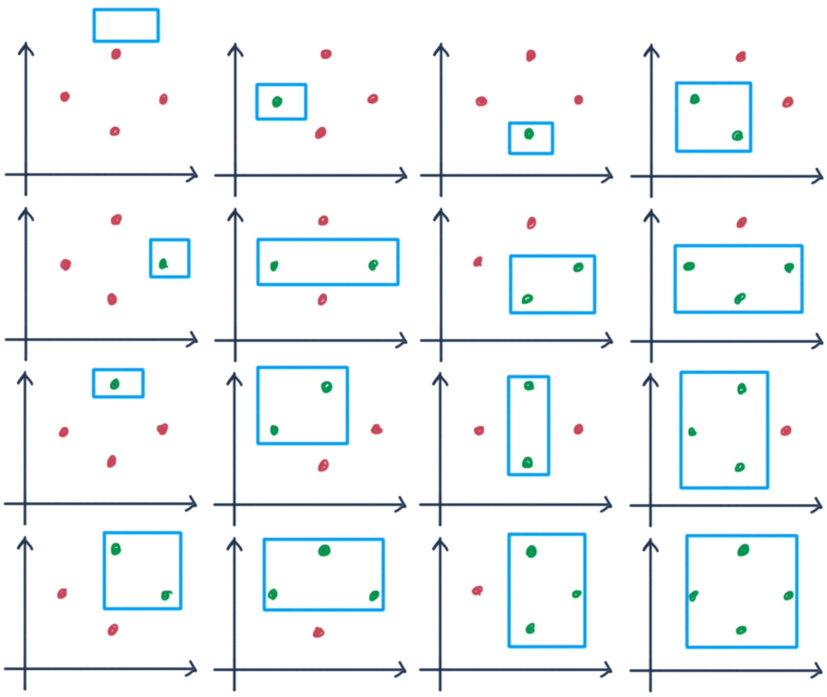
$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i} = \sum_{i=0}^{4} \binom{N}{i}$$

۴ سوال چهارم

برای این سوال نقاط سبزرنگ نقش برچسب $+1$ و نقاط قرمزرنگ نقش برچسب -1 را دارند.

۱.۴ آ

برای یک و دو و سه نقطه مشخص است که به بیشتری تعداد دایکاتمی می‌رسیم.برای چهار نقطه به شکل زیر می‌توان به همه‌ی ۱۶ دایکاتمی متفاوت رسید.



برای پنج نقطه، برای اینکه بتوانیم به همه‌ی دایکاتمی های ممکن برسیم آن ها را حول یک پنج صلعی منتظم قرار می‌دهیم. ولی حتی در این حالت نیز به هیچ وجه نمی‌توانیم چنین برچسب‌دهی را بسازیم:

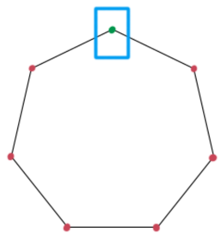
پس برای این فضای فرضیه نقطه‌ی شکست ۵ است. در این حالت :

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i} = \sum_{i=0}^{4} \binom{N}{i}$$

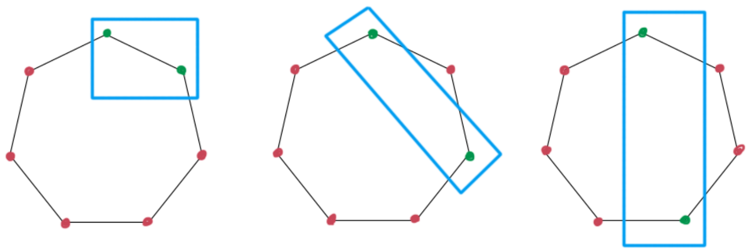
۲.۴ ب

ابتدا نشان می‌دهیم که برای ۷ نقطه، می‌توانیم به بیشترین تعداد دایکاتمی برسیم. بدین منظور روی تعداد نمونه‌ها با برچسب مثبت حالت بندی می‌کنیم تا از بررسی حالات تکراری پرهیز کنیم. همچنین برای اینکه بیشترین تعداد دایکاتمی را به دست آوریم، ۷ نقطه را رئوس یک هفت ضلعی منتظم در نظر می‌گیریم:

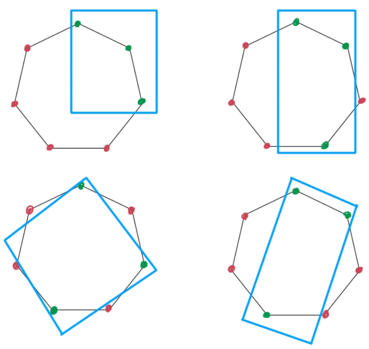
- یک مثبت : بدیهی است که پترن‌هایی که یک برچسب مثبت دارند به سادگی قابلیت ایجاد توسط فضای فرضیه را دارند :



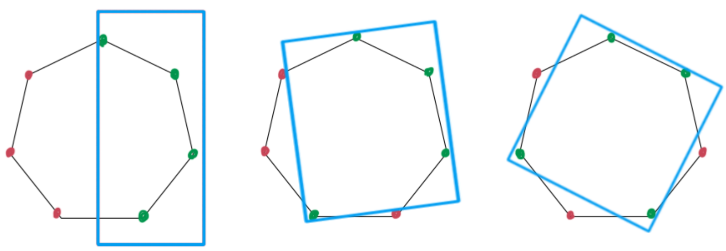
- دو مثبت : پترن‌هایی که دو برچسب مثبت دارند قابلیت ایجاد توسط فضای فرضیه را دارند :



- سه مثبت : پترن‌هایی که سه برچسب مثبت دارند قابلیت ایجاد توسط فضای فرضیه را دارند :

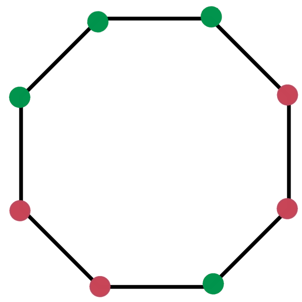


- چهار مثبت : پترن‌هایی که چهار برچسب مثبت دارند قابلیت ایجاد توسط فضای فرضیه را دارند :



غیر از حالات ارائه شده، بقیه‌ی حالات تکراری بوده و به همین روش قابلیت این را دارند که از فرضیه به دست بیایند. پس توانستیم ۷ نقطع را توسط این فضای فرضیه $Shatter$ کنیم.

حال برای ۸ نقطه، آن ها را روی رئوس یک ۸ ضلعی منتظم قرار می‌دهیم که بتوانیم بیشترین دایکاتمی را به دست آوریم. همانطور که از تصویر مشخص است برچسب دهی زیر قابل تولید توسط این فضای فرضیه نیست.

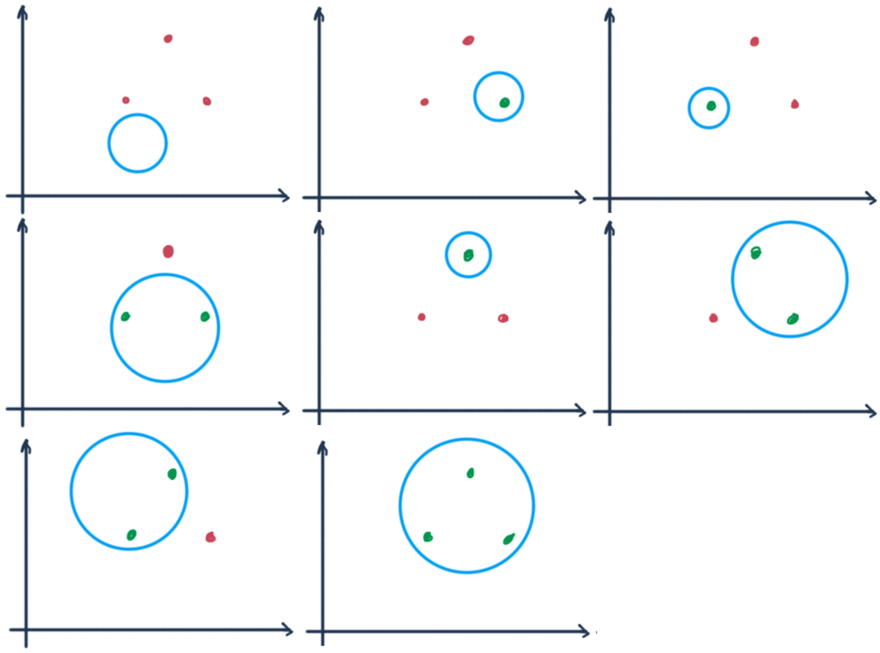


پس برای این فضای فرضیه نقطه‌ی شکست ۸ است. در این حالت :

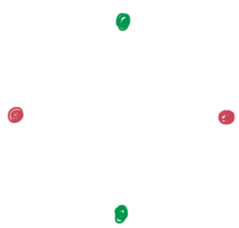
$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i} = \sum_{i=0}^{7} \binom{N}{i}$$

۳.۴ پ

برای یک و دو نقطه مشخص است که می‌توانیم به بیشترین تعداد دایکاتمی های ممکن برسیم. برای سه نقطه به شکل زیر می‌توان به همه‌ی ۸ دایکاتمی ممکن رسید



برای چهار نقطه، برای اینکه بتوانیم به همه‌ی دایکاتمی های ممکن برسیم آن ها را حول یک چهار صلعی منتظم قرار می‌دهیم. ولی حتی در این حالت نیز به هیچ وجه نمی‌توانیم چنین برچسب‌دهی را بسازیم:



پس برای این فضای فرضیه نقطه‌ی شکست ۴ است. در این حالت :

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i} = \sum_{i=0}^{3} \binom{N}{i}$$

۵ سوال پنجم

نامساوی هافدینگ با شرط $\epsilon = 0.05$ و $2Me^{-2\epsilon^2 N} \leq 0.03$ به شکل زیر در می‌آید :

$$P(|E_{in}(g) - E_{out}(g)| > 0.05|) \leq 2Me^{-2(0.05)^2 N} \leq 0.03$$

پس داریم :

$$\begin{aligned}
&2Me^{-2(0.05)^2 N} \leq 0.03 \\
\Rightarrow &Me^{-2(0.05)^2 N} \leq 0.015 \\
\Rightarrow &\log_e(M) + (-2(0.05)^2 N) \leq \log_e(0.015) \\
\Rightarrow &\log_e(M) + (-0.005N) \leq \log_e(0.015) \\
\Rightarrow &\log_e(M) - \log_e(0.015) \leq 0.005N \\
\Rightarrow &200(\log_e(M) - \log_e(0.015)) \leq N
\end{aligned} \tag{۱}$$

حال به ازای M های مختلف تعداد نمونه‌های مورد نیاز را محاسبه می‌کنیم:

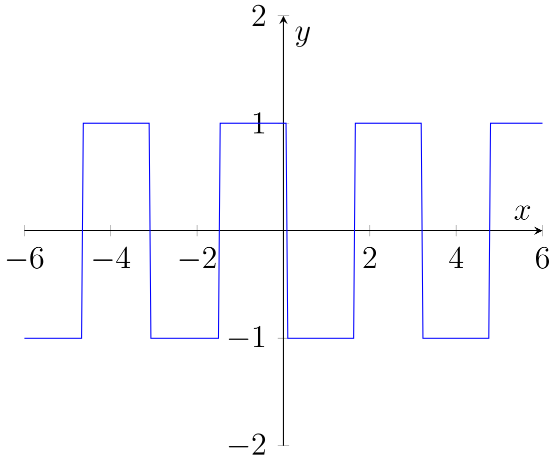
$$M = 1 \Rightarrow 200(\log_e(1) - \log_e(0.015)) \leq N \Rightarrow N \geq 839.94 \Rightarrow N \geq 840$$
$$M = 10 \Rightarrow 200(\log_e(10) - \log_e(0.015)) \leq N \Rightarrow N \geq 1300.45 \Rightarrow N \geq 1301$$
$$M = 1000 \Rightarrow 200(\log_e(1000) - \log_e(0.015)) \leq N \Rightarrow N \geq 2221.49 \Rightarrow N \geq 2222$$

نتیجه‌ای که به دست می‌آید این است که هر چه تعداد عناصر درون فضای فرضیه رشد می‌کند، همانطور که از ۱ نیز برمی‌آید تعداد نمونه‌های مورد نیاز به شکل لگاریتمی رشد می‌کند و باعث می‌شود برای فضاهای فرضیه‌ی بسیار بزرگ به همان نسبت نمونه نیاز نداشته باشیم.

۶ سوال ششم

۱.۶ آ

تصویر زیر، یکی از عناصر این فضای فرضیه را نمایش می‌دهد که در آن $a = 2$ و $b = 3$ است.



متاسفانه صورت سوال اشکال دارد. اگر در همه‌ی عناصر فضای فرضیه $b = 0$ باشد در آن صورت اثبات می‌شود که نمی‌توان این نقاط را $Shatter$ کرد. ولی در غیر اینصورت اگر $b \neq 0$ مثال نقض خواهیم زد که می‌توان نقاط اشاره شده را $Shatter$ کرد.
برای حالتی که $b = 0$ فرض کنید یک a وجود دارد که $sin(ax)$ چنین برچسب گذاری تولید می‌کند:

$$x : -, 2x : -, 3x : +, 4x : -$$

پس

$$sin(ax) < 0, sin(2ax) < 0, sin(3ax) \geq 0, sin(4ax) < 0$$

از $sin(4ax) < 0$ نتیجه می‌گیریم که

$$sin(4ax) = 2sin(2ax)cos(2ax) < 0$$

چون $sin(2ax) < 0$ پس $cos(2ax) > 0$. در نتیجه

$$cos(2ax) = 1 - 2sin^2(ax) > 0 \Rightarrow sin^2(ax) < \frac{1}{2} \qquad (۱)$$

از طرف دیگر

$$sin(3ax) = 3sin(ax) - 4sin^3(ax) \geq 0 \xrightarrow{sin(3ax) \geq 0} 3 - 4sin^2(ax) \leq 0 \Rightarrow sin^2(ax) \geq \frac{3}{4} \qquad (۲)$$

اگر ۱ و ۲ را در کنار هم قرار دهیم داریم :

$$sin^2(ax) < \frac{1}{2} \ , \ sin^2(ax) \geq \frac{3}{4}$$

که این یک تناقض است. پس به ازای هر $x \in \mathbb{R}$ نمی‌توان نقاط اشاره شده را $Shatter$ کرد.
برای حالتی که b بتواند غیر صفر باشد، فرض کنید $x = 6$. در اینصورت همه‌ی حالات برچسب‌گذاری این ۴ نقطه توسط فرضیه‌هایی که پارامترهای آن در تصویر آمده است قابل $Shatter$ کردن است.

```
-1,-1,-1,-1   a=1,b=24
-1,-1,-1,+1   a=1,b=23
-1,-1,+1,-1   a=6,b=38
-1,-1,+1,+1   a=1,b=29
-1,+1,-1,-1   a=6,b=30
-1,+1,-1,+1   a=8,b=31
-1,+1,+1,-1   a=5,b=24
-1,+1,+1,+1   a=1,b=35
+1,-1,-1,-1   a=1,b=38
+1,-1,-1,+1   a=4,b=20
+1,-1,+1,-1   a=8,b=34
+1,-1,+1,+1   a=6,b=27
+1,+1,-1,-1   a=1,b=32
+1,+1,-1,+1   a=6,b=35
+1,+1,+1,-1   a=1,b=26
+1,+1,+1,+1   a=1,b=20
```

این نتایج از طریق یک کد در زبان پایتون به دست آمده است و به سادگی قابل تحقیق است که این فرضیه‌ها با پارامترهای داده شده می‌توانند همه‌ی پترن ها را ایجاد کنند. با چشم پوشی از این فرضیه‌ها ی فضای فرضیه که $b \neq 0$ اثبات ارائه شد.

۲.۶ ب

برای هر $m > 0$ مجموعه‌ی نقاط $(x_1, x_2, ..., x_m)$ با مجموعه برچسب‌های $(y_1, y_2, ..., y_m) \in \{-1, +1\}^m$ را در نظر بگیرید که $x_j = 2^{-m}$. قرار می‌دهیم:

$$a = \pi(1 + \sum_{i=1}^{m} 2^i(\frac{1-y_i}{2})) \quad , \quad b = 0$$

ثابت می‌کنیم این پارامترها روی داده های اشاره شده منجر به تولید بدون خطای برچسب‌های همه‌ی نقاط می‌شوند و در نتیجه برای هر m همه‌ی پترن‌های ممکن برای m نقطه‌ی ذکر شده تولید می‌شود و $vc - dimension$ برابر ∞ خواهد بود. برای هر j داریم :

$$\begin{aligned} ax_j = a2^{-j} &= \pi(2^{-j} + \sum_{i=1}^{m} 2^{i-j}(\frac{1-y_i}{2})) \\ &= \pi(2^{-j} + \sum_{i=1}^{j-1} 2^{i-j}(\frac{1-y_i}{2}) + (\frac{1-y_j}{2}) + \sum_{i=1}^{m-j} 2^{i}(\frac{1-y_i}{2})) \end{aligned}$$

از آن‌جایی که جمله‌ی آخر عبارت بالا به اندازه مضرب صحیحی از 2π در نتیجه نهایی مشارکت می‌کند پس می‌توانیم آن را حذف کنیم. زیرا :

$$sin(\theta) = sin(\theta + 2k\pi)$$

پس عبارت باقی مانده به شکل زیر خواهد بود:

$$\pi(2^{-j} + \sum_{i=1}^{j-1} 2^{i-j}(\frac{1-y_i}{2}) + (\frac{1-y_j}{2})) = \pi(2^{-j} + \sum_{i=1}^{j-1} 2^{-i}(\frac{1-y_i}{2}) + (\frac{1-y_j}{2}))$$

حال با توجه به اینکه $\frac{1-y_i}{2} \in \{0, 1\}$ ، یک کران پایین و بالا برای عبارت بالا به دست می‌آوریم.

$$\pi(2^{-j} + \sum_{i=1}^{j-1} 2^{-i}(\frac{1-y_i}{2}) + (\frac{1-y_j}{2})) \leq \pi(\sum_{i=1}^{j} 2^{-i} + (\frac{1-y_j}{2})) < \pi(1 + \frac{1-y_j}{2})$$

$$\pi(2^{-j} + \sum_{i=1}^{j-1} 2^{-i}(\frac{1-y_i}{2}) + (\frac{1-y_j}{2})) > \pi(\frac{1-y_j}{2})$$

پس اگر $y_j = +1$ کران بالا برابر π و کران پایین برابر 0 خواهد بود. یعنی :

$$0 < ax_j < \pi \Rightarrow sign(sin(ax_j)) = +1 = y_j$$

و به طریق مشابه اگر $y_j = -1$ کران بالا برابر 0 و کران پایین برابر $-\pi$ خواهد بود. یعنی :

$$-\pi < ax_j < 0 \Rightarrow sign(sin(ax_j)) = -1 = y_j$$

در نتیجه در هر دو حالت پترن مورد نظر توسط فرضیه با پارامتر a تولید شد. پس اثبات کامل است.