دانشکده‌ی مهندسی کامپیوتر

یادگیری ماشین — زمستان ۱۳۹۹

پاسخ تمرین سری دوم

مدرس: دکتر محمّدحسین رهبان

پاسخ سوال ۱ آ) فرض کنید نسخه‌ی حاوی نویز f را f' بنامیم، با توجه به اینکه f' و h به صورت مستقل دچار خطا می‌شوند؛ خواهیم داشت $P(A \cup B) = P(A) + P(B)$ ، بنابراین:

$$P[h(x) \neq f'(x)] = P\left[f'(x) \neq f(x) \cap h(x) = f(x)\right] + P\left[f'(x) = f(x) \cap h(x) \neq f(x)\right] = \mu \cdot \lambda + (1-\mu) \cdot (1-\lambda)$$

ب) اگر $\lambda = \frac{1}{2}$ خواهیم داشت $\lambda = 1 - \lambda$ بنابراین:

$$P\left[h(x) \neq f'(x)\right] = \mu \cdot \lambda + (1-\mu) \cdot \lambda = \lambda$$

در نتیجه، عملکرد h از μ مستقل می‌شود.

پاسخ سوال ۲ آ) برای مثال سوپرمارکت داریم:

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

و برای مثال CIA داریم:

$$\begin{aligned} E_{in}^{(s)}(h) &= \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), f(x_n)) \\ &= \frac{1}{N}[\sum_{y_n=1} e(h(x_n), 1) + \sum_{y_n=-1} e(h(x_n), -1)] \\ &= \frac{1}{N}[\sum_{y_n=1} [(h(x_n) \neq 1] + \sum_{y_n=-1} 1000 \cdot [h(x_n) \neq -1]] \end{aligned}$$

ب) ابتدا نشان می‌دهیم، $h^*(x) = \mathbb{E}_{y|x}[y|x]$ مقدار E_{out} را کمینه می‌کند. به ازای هر فرضیه h داریم:

$$\begin{aligned}
E_{out}(h) &= \mathbb{E}_{(x,y)}[(h(x)-y)^2] \\
&= \mathbb{E}_{(x,y)}[((h(x)-h^*(x))+(h^*(x)-y))^2] \\
&= \mathbb{E}_{(x,y)}[(h(x)-h^*(x))^2] + 2\underbrace{\mathbb{E}_{(x,y)}[(h(x)-h^*(x))(h^*(x)-y)]}_{(*)} + \mathbb{E}_{(x,y)}[(h^*(x)-y)^2]
\end{aligned}$$

و

$$\begin{aligned}
(*) &= \mathbb{E}_x[\mathbb{E}_{y|x}[(h(x)-h^*(x))(h^*(x)-y)|x]] \\
&= \mathbb{E}_x[(h(x)-h^*(x))\mathbb{E}_{y|x}[(h^*(x)-y)|x]] \\
&= \mathbb{E}_x[(h(x)-h^*(x))(\mathbb{E}_{y|x}[h^*(x)|x]-\mathbb{E}_{y|x}[y|x])] \\
&= \mathbb{E}_x[(h(x)-h^*(x))(h^*(x)-h^*(x))] = 0
\end{aligned}$$

بنابراین برای هر فرضیه‌ی h خواهیم داشت:

$$E_{out}(h) = \mathbb{E}_{(x,y)}[(h(x)-y)^2] = \mathbb{E}_{(x,y)}[(h(x)-h^*(x))^2] + \mathbb{E}_{(x,y)}[(h^*(x)-y)^2] \geq \mathbb{E}_{(x,y)}[(h^*(x)-y)^2]$$

در نتیجه $h^*(x)$ فرضیه‌ای است که E_{out} را کمینه می‌کند. حال واضح است که داریم:

$$y = h^*(x) + (y - h^*(x)) = h^*(x) + \epsilon(x)$$

بنابراین $\epsilon(x) = y - h^*(x)$ و امید ریاضی $\epsilon(x)$ به صورت زیر به دست می‌آید:

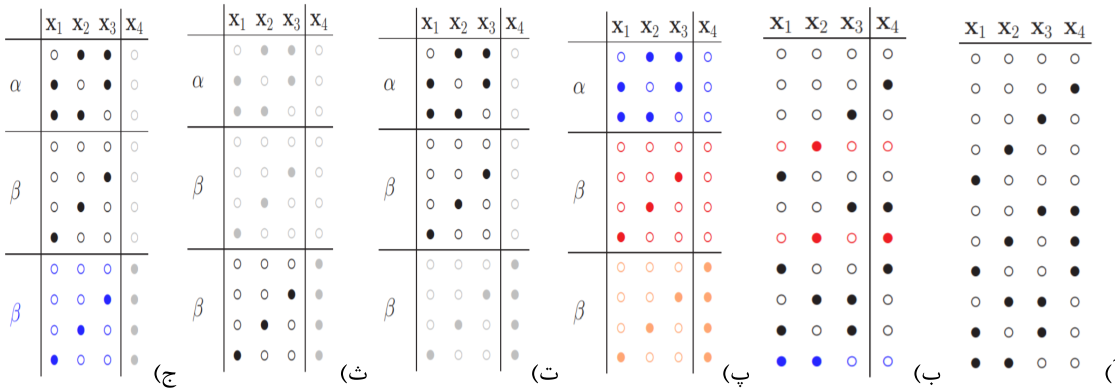
$$\mathbb{E}_{y|x}[\epsilon(x)|x] = \mathbb{E}_{y|x}[y|x] - \mathbb{E}_{y|x}[h^*(x)|x] = h^*(x) - h^*(x) = 0$$

پاسخ سوال ۳ آ)تعداد تفکیک‌ها روی چهار نقطه به طوری که هیچ زیرمجموعه سه‌تایی از آن‌ها $shattered$ نشود، برای مثال پازل در شکل ۱-(آ) داده شده است:

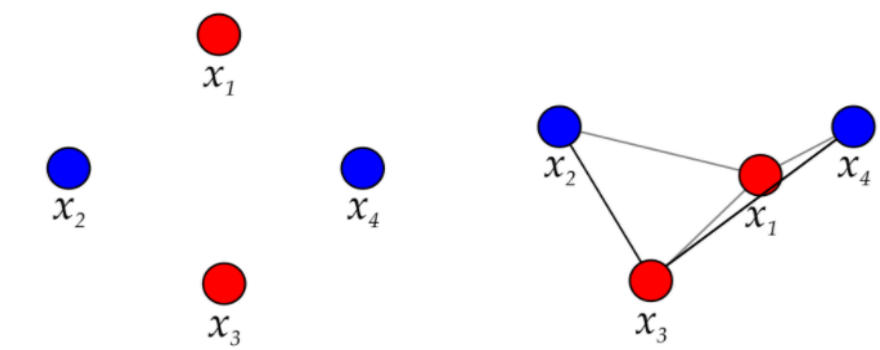
حال اگر نقطه آخر را جدا کنیم به شکل ۱-(ب) میرسیم. اکنون میتوانیم دسته‌بندی α و β را به گونه‌ای انجام دهیم که α شامل الگوهایی باشد که x_1 تا x_3 در آن‌ها یکبار ظاهر شده و β شامل الگوهایی باشد که x_1 تا x_3 در آن‌ها دوبار ظاهر شده. بنابراین به جدول شکل ۱-(پ) خواهیم رسید که $B(3,4) = \alpha + 2\beta$. می‌دانیم $\alpha + \beta \leq B(3,3)$ که این موضوع از شکل ۱-(ت) هم مشخص است. همچنین داریم $\beta \geq B(3,2)$ شکل ۱-(ث). بنابراین $B(4,3) = \alpha + \beta + \beta \leq B(3,3) + B(3,2)$ شکل ۱-(ج). و رابطه داده شده به رابطه موجود در صورت سوال قابل تعمیم است:

$$B(N,k) \leq B(N-1,k) + B(N-1,k-1)$$

ب) نشان می‌دهیم که هیچ پنج نقطه‌ای با مدل پرسپترون در فضای سه بعدی قابل $shattered$ کردن نیست و بنابراین نقطه شکست برابر پنج است.



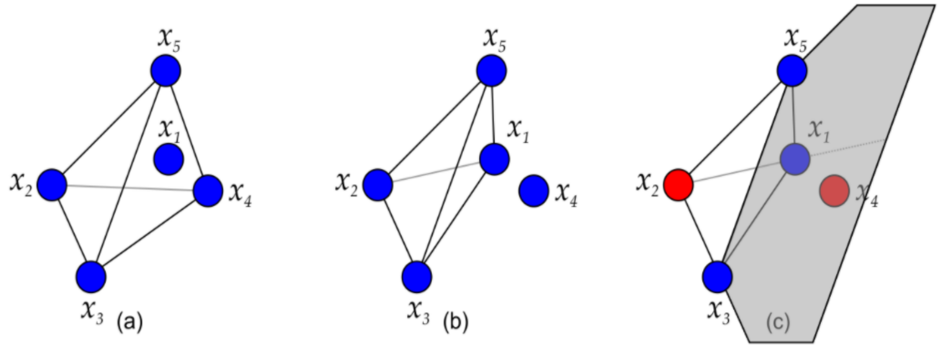
شکل ۱: دسته‌بندی مثال پازل



شکل ۲: تفکیک چهار نقطه در فضای دو بعدی(سمت چپ) و سه بعدی(سمت راست)

ابتدا پنج نقطه در فضای سه بعدی در نظر می‌گیریم. می‌دانیم که هیچ چهار نقطه ای از این پنج نقطه نباید در یک صفحه قرار بگیرند(چرا که در غیر این صورت واضحا قابل $shattered$ کردن نیستند (شکل ۲)). اولین نکته ای که باید توجه داشته باشیم این است که همیشه می‌توانیم از پنج نقطه، یک زیرمجموعه‌ی چهار نقطه‌ای را انتخاب کرده و یک چهار ضلعی بسازیم به طوری که نقطه باقیمانده خارج از آن قرار بگیرد. این موضوع در شکل ۳ نشان داده شده است: برای ساخت این چهار ضلعی، می‌توانیم با چهار نقطه شروع کنیم(شکل ۳-(a)) و اگر نقطه باقیمانده درون چهارضلعی قرار بگیرد، جوری آن را تغییر می‌دهیم تا نقطه باقیمانده قسمتی از یک چهارضلعی جدید شود، که پس از این کار لزوما یکی از دیگر از نقاط از چهارضلعی حذف می‌شود(شکل ۳-(b)). اکنون تباید توجه کنیم که نقطه‌ای که در خارج از چهارضلعی باقی مانده است ، همیشه در داخل "سایه" ایجاد شده توسط یکی از طرفین آن چهارضلعی قرار دارد. چرا که اجتماع سایه‌ها بر روی نقاطی که خارج از چهارضلعی واقع شده اند، یک پارتیشن تشکیل می‌دهد، این به این معنی است که x_4 در شکل ۲-(ب) که نقطه ای است که خارج چهار ضلعی واقع شده، دقیقا شامل یکی از آن‌ها می‌شود (شکل ۳-(c)). اگر آن سایه را با T نمایش دهیم، مجموعه نقاطی که در سایه ایجاد شده توسط T قرار دارند را با S_T نمایش می‌دهیم که $x_4 \in S_T$

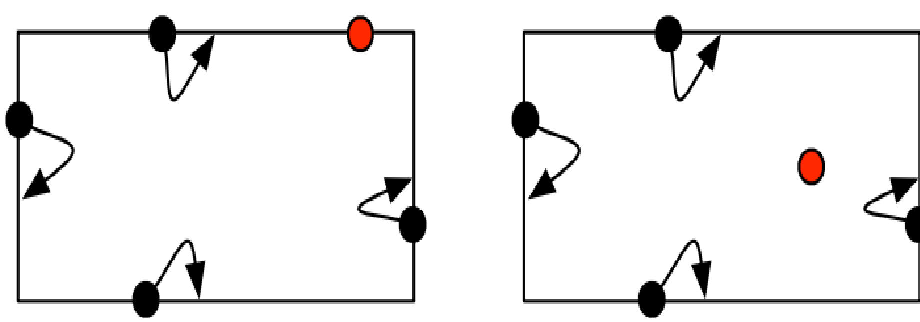
حال تفکیک هایی را در نظر بگیرید که در آن‌ها x_2 قرمز و x_1, x_3, x_5 آبی باشند. هر صفحه ای که بتواند x_2 را از $s = x_1, x_3, x_5$ جدا کند باید تمام نقاط عضو S_T را در نیم صفحه‌ی مشابه با s قرار دهد و این به این معنی است که x_4 باید همرنگ نقاط داخل s باشد(آبی). بنابراین این مجموعه از نقاط قابل $shattered$ کردن نیستند.



بنابراین حد بالای بیشترین تعداد تفکیک ها به صورت زیر به دست می‌آید:

$$
\begin{aligned}
B(N,5) \leq \sum_{i=0}^{4} \binom{N}{i} &= \binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \binom{N}{3} + \binom{N}{4} \\
&= 1 + N + \frac{N(N-2)}{2} + \frac{N(N-1)(N-2)}{6} + \frac{N(N-1)(N-2)(N-3)}{24} \\
&= \frac{1}{24}N^4 - \frac{1}{12}N^3 + \frac{11}{24}N^2 - \frac{7}{12}N + 1
\end{aligned}
$$

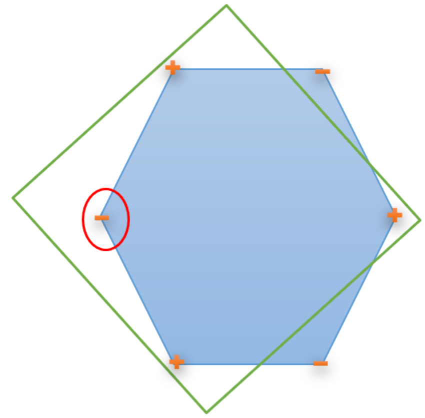
پاسخ سوال ۴ آ) ثابت میکنیم هیچ پنج نقطه‌ای $shattered$ نمی‌شود. یک $shattering$ مناسب برای پنج نقطه باید به ما اجازه دهد که تمام پنج نقطه را و همچنین چهار نقطه را بدون نقطه پنجم انتخاب کنیم. اما کوچکترین مستطیلی که حول پنج نقطه میتوان داشت تنها با چهار نقطه که هر کدام یکی از اضلاع مستطیل را مشخص می‌کنند، تعیین می‌شود. بنابراین واضح اس که نقطه پنجم یا روی یک ضلع قرار می‌گیرد و یا داخل مستطیل قرار می‌گیرد. این مسئله مانع انتخاب چهار نقطه بدون پنجمین نقطه می‌شود. بنابراین ادعا اثبات میشود.



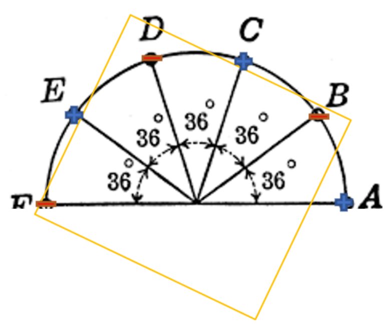
$$
m_H(N) \leq \frac{1}{24}N^4 - \frac{1}{12}N^3 + \frac{11}{24}N^2 - \frac{7}{12}N + 1
$$

ب) نقطه شکست در اینجا ۸ نقطه است که این تعداد قابل $shattered$ کردن با مستطیل های قابل چرخش در فضای $\mathbb{R}^2$ نیستند.تا چهار نقطه که در قسمت قبل نشان داده شد، حال برای پنج نقطه به سادگی میتوان نشان داد که اگر نقاط روی یک پنج ضلعی منتظم قرار بگیرند قابل $shattered$ شدن

خواهند بود. حال به بررسی شش نقطه می پردازیم که اگر شش نقطه را در فضا به صورت یک شش ضلعی منتظم در نظر بگیریم، در این شش‌ضلعی نمی‌توان مستطیلی کشید که تمام دسته‌بندی‌ها برای سه نقطه را در بر بگیرد (حالتی که برچسب‌های شش نقطه به صورت یکی در میان قراربگیرند، قابل $shattered$ کردن نیست).



اما این موضوع برای یک هفت‌ضلعی برقرار است، یعنی می‌توان مستطیلی پیدا کرد که تمام دسته‌بندی‌ها را در یک هفت‌ضلعی در بر بگیرد. به طور کلی، با توجه به اینکه یکی از بهترین راه ها برای بررسی نقطه شکست گذاشتن نقاط روی یک چندضلعی منتظم است و همچنین با توجه به اینکه یک مستطیل از دو سری خط موازی تشکیل شده است؛ اگر تعداد نقاط زوج باشد که چندضلعی منتظم حاصل دارای اضلاع موازی است و $shattered$ کردن آن نقاط توسط مستطیل‌های قابل چرخش در فضای دوبعدی غیر ممکن خواهد بود اما اگر تعداد نقاط فرد باشد و روی یک چند ضلعی منتظم قرار گیرند، قابل $shattered$ شدن با مستطیل قابل چرخش در فضا خواهند بود. اما اگر به جای در نظر گرفتن شش نقطه روی یک شش ضلعی منتظم آن‌ها را روی یک نیم دایره با فواصل مساوی قرار دهیم میبینیم که شش نقطه نیز قابل $shattered$ شدن با مستطیل‌های قابل چرخش در فضای دو بعدی خواهد بود. اما این موضوع برای هشت نقطه برقرار نیست.



$$m_H(N) \leq \sum_{i=0}^{7} \binom{N}{i} = \frac{1}{5040}N^7 - \frac{1}{360}N^6 + \frac{1}{45}N^5 - \frac{5}{72}N^4 + \frac{157}{720}N^3 + \frac{13}{180}N^2 + \frac{319}{420}N + 1$$

پ) نقطه شکست این دسته بند چهار نقطه است. برای $shattered$ کردن چهار نقطه باید این چهار نقطه روی یک دایره قرار بگیرند. می‌توانیم چهار نقطه را به صورتی در نظر بگیریم که یک دایره یکتا وجود داشته باشد که تمام این نقاط روی آن قرار بگیرند اما از آنجا که هر زیر مجموعه سه‌تایی از این نقاط هم همین دایره را تشکیل میدهند(با کشیدن یک دایره دور هر تعداد نقطه‌ای می‌توانیم آنها را به دسته مثبت اختصاص دهیم)، بنابراین نمی‌توانیم

هیچ کدام از این سه نقطه را به دسته مثبت اختصاص دهیم بدون این که نقطه چهارم هم داخل دایره قرار بگیرد.

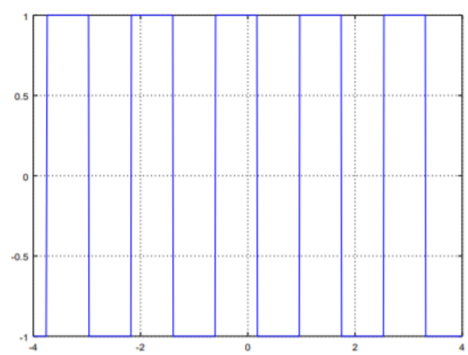
$$m_H(N) \leq \frac{1}{6}N^3 + \frac{5}{6}N + 1$$

پاسخ سوال ۵

$$2Me^{-2\epsilon^2 N} \leq 0.03 \Rightarrow \begin{cases} M = 1 & N \geq 839.9 \Rightarrow minN = 840 \\ M = 10 & N \geq 1300.4 \Rightarrow minN = 1301 \\ M = 1000 & N \geq 2221.5 \Rightarrow minN = 2222 \end{cases} \tag{۱}$$

با توجه به مقادیر بالا میتوان گفت برای آن که $generalizationerror$ ثابت نگه داشته شود، با زیاد شدن تعداد فرضیات مختلف در فضای فرضیه تعداد نمونه‌های مورد نیازمان هم به صورت لگاریتمی بیشتر میشود.

پاسخ سوال ۶ آ) با توجه به شکل زیر می‌توان دید که تابع مورد نظر یک موج مربعی با اندازه ۱ و دوره تناوب $2\pi cos(tan^{-1}(a))$ است. با ثابت کردن a, b میتوان به یک تابع از این کلاس رسید و دید که نقاط $x, 2x, 3x, 4x$ قابل $shattered$ کردن با این تابع نیستند.



ب) بعد VC این کلاس فرضیه بی‌نهایت است.

اثبات: نشان می‌دهیم برای هر l مجموعه نقاط $\{x_1, x_2, ..., x_l\}$ که $x_i = 10^{-i}$ قابل $shattered$ کردن هستند. با در نظر گرفتن هر برچسب‌دهی $\{y_1, y_2, ..., y_l\}$ و قرار دادن $a = \pi(1 + \sum_{i=1}^{l} \frac{(1-y_i)10^i}{2})$ و $b = 0$ خواهیم داشت:

$$\begin{aligned} h(x_j) &= sign\Big(sin\Big(10^{-j}\pi(1 + \sum_{i=1}^{l} \frac{(1-y_i)10^i}{2})\Big)\Big) \\ &= sign\Big(sin\Big(10^{-j}\pi + \sum_{i=1}^{l}(1-y_i)10^{i-j}\frac{\pi}{2})\Big)\Big) \end{aligned}$$

برای هر $y_i = 1$ عبارت مربوطه در $summation$ برای آن مقدار صفر به خود می‌گیرد. همچنین، هرگاه $i > j$ ضریب صحیحی از $\frac{\pi}{2}$ به تابع sin اضافه

میشود که باعث تغییری در مقدار آن نخواهد شد. بنابراین این عبارات قابل حذف از مجموع هستند.

$$\begin{aligned} h(x_j) &= sign\Big(sin\Big(10^{-j}\pi + \sum_{i:i<j, y_i=-1}(1-y_i)10^{i-j}\frac{\pi}{2})\Big)\Big) \\ &= sign\Big(sin\Big(10^{-j}\pi + (1-y_j)\frac{\pi}{2} + \sum_{i:i<j, y_i=-1} 2*10^{i-j}\frac{\pi}{2})\Big)\Big) \\ &= sign\Big(sin\Big(10^{-j}\pi + (1-y_j)\frac{\pi}{2} + \pi\sum_{i:i<j, y_i=-1} 10^{i-j})\Big)\Big) \end{aligned}$$

چرا که $sin(\pi + x) = -sin(x)$.

مقدار جمع عبارت آخر همیشه از یک کمتر خواهد بود ، بنابراین اگر $y_j = 1$ آرگومان تابع سینوسی بین صفر و π خواهد بود. بنابراین تابع سینوسی مقدار مثبتی خواهد داشت و $h(x_j) = 1 = y_j$. حال اگر $y_j = -1$ آرگومان سینوس بین π و 2π خواهد بود و تابع سینوسی مقادیر منفی خواهد گرفت و $h(x_j) = -1 = y_j$. درنتیجه $h(x_j) = y_j, \forall j$ و مجموعه $\{10^{-1}, 10^{-2}, ..., 10^{-l}\}$ میتواند برای هر مقدار l ، $shattered$ شود و بعد VC بی‌نهایت می‌باشد.

پاینده باشید