پویا خانی ۹۹۲۱۰۲۸۳ **تمرین دوم یادگیری ماشین**

سوال ۱ – همان‌طور که میدانیم تابع f در حالت کلی در برچسب گذاری داده ها اشتباه نمیکند اما در این سوال چون نویز وجود دارد، ممکن است مرتکب اشتباه شود و به همین دلیل توزیع احتمال P(y|x) مطرح است. حال قرار است فرضیه h آن تابع f را تا حد خوبی تخمین بزند.

الف) در دو حالت فرضیه h خطا دارد، اول وقتی که تابع f درست حدس زده باشد یعنی f(x) = y اما تابع h برای این داده، نتواند تخمین درستی از f داشته باشد و برچسب‌گذاری داده x را اشتباه انجام دهد. حالت بعدی وقتی است که خود f هم به دلیل نویز، برچسب داده x را اشتباه بزند یعنی f(x) != y و در اینجا h یک تخمین خوب از f داشته باشد و او هم اشتباه f را تکرار کند. بنابراین داریم:

$$P(\,h(x) \neq y\,) = P(\,f(x) = y\,)P(\,h(x) \neq y \mid f(x) = y) + P(\,f(x) \neq y\,)P(\,h(x) \neq y \mid f(x) \neq y) = \lambda\mu + (1-\lambda)(1-\mu)$$

ب) وقتی میگوییم میخواهیم تابع h از μ مستقل بشود، یعنی در رابطه بالا باید ضریب μ صفر بشود. یعنی باید رابطه بالا را ساده کرده و ضریب آن را صفر کنیم. پس:

$$\lambda\mu + (1-\lambda)(1-\mu) = \lambda\mu + 1 - \mu - \lambda + \lambda\mu = 2\lambda\mu - \mu - \lambda + 1 = (2\lambda - 1)\mu - \lambda + 1$$

$$\rightarrow 2\lambda - 1 = 0 \rightarrow \lambda = \frac{1}{2}$$

سوال ۲ –

الف) اگر یک تابع با مشخصات زیر به نام risk تعریف کنیم:

$$risk(i) = \begin{cases} r(y_i) & if\ h(x_i) \neq y_i \\ 0 & if\ h(x_i) = y_i \end{cases}$$

آنگاه خطای داخل نمونه میشود:

$$E_{in} = \sum_{i=1}^{N} risk(i)$$

حال برای اینکه بتوانیم خطا را به صورت صریح و برحسب توابع *h* و *r* و برچسب داده یعنی *y* بیان کنیم، باید تابع ***risk*** را بر حسب این توابع بازنویسی کنیم. از آنجا که مقدار *y* برای یک داده میتواند ۱+ یا ۱- باشد، و همچنین تابع *h* نیز برای یک داده همین مقادیر را به خود میگیرد، اگر $h(x_i) = y_i$ همیشه مقدار $h(x_i)y_i$ برابر با ۱+ میشود.وقتی این اتفاق بیفتد یعنی ما هزینه ای نباید پرداخت کنیم(ریسکی وجود ندارد).برعکس وقتی $h(x_i) \neq y_i$ باشد، مقدار $h(x_i)y_i$ همیشه برابر ۱- میشود.

پس تابع ***risk*** را میتوان به شکل زیر بازنویسی کرد:

$$risk(i) = -\min(0, h(x_i)y_i) * r(y_i)$$

هم چنین اگر بخواهیم $r(y_i)$ را بر حسب نوع برچسب داده بازنویسی کنیم، داریم:

$$r(y_i) = [-\min(y_n, 0), -\min(-y_n, 0)].\left[r_{y_n=-1}, r_{y_n=1}\right]^T$$

لذا در نهایت داریم:

$$E_{in} = \sum_{i=1}^{N}(-\min(0, h(x_i)y_i) * r(y_i))[-\min(y_n, 0), -\min(-y_n, 0)].\left[r_{y_n=-1}, r_{y_n=1}\right]^T$$

ب) وقتی حرف از کمینه کردن یک عبارت به میان می‌آید، یعنی باید مشتق گرفته شده و برابر صفر در نظر بگیریم. چون تابع خطای خارج‌نمونه بر حسب فرضیه *h* است، لذا از آن بر حسب *h* مشتق جزئی میگیریم:

$$\frac{d\, E_{out}(h)}{d\, h} = E[2(h(x) - y)] = 2E[h(x) - y] = 2E_x\left[E_{y|x}[h(x) - y]\right]$$

توجه شود آخرین برابری بخاطر این اتفاق افتاد که امید ریاضی اولیه بر حسب متغیر های *x* و *y* بودند و برای جدا کردن آن ها میتوانیم امید ریاضیِ امید ریاضی شرطی را جایگذاری کنیم.

حال میدانیم که برای صفر کردن عبارت بالا، نیاز داریم که $\left[E_{y|x}[h(x) - y]\right]$ صفر شود. لذا داریم:

$$E_{y|x}[h(x) - y] = E_{y|x}[h(x)] - E_{y|x}[y] = h(x) - E[y|x] = 0$$
$$\rightarrow h(x) = E[y|x] = h^*(x)$$

توجه شود برابری هایلایت شده به این دلیل اتفاق افتاد که *h(x)* مستقل از *y* است لذا $E_{y|x}[h(x)] = h(x)$.

برای اثبات قسمت دوم سوال، از رابطه $y = h^*(x) + \in$ امیدریاضی میگیریم.پس داریم:

$$E[y] = E[h^*(x)] + E[\in(x)] = E\left[E[y|x]\right] + E[\in(x)] = E[y] + E[\in(x)] \rightarrow E[\in(x)] = 0$$

سوال سوم-

الف) ابتدا هر ۱۶ حالت مختلف ۴ نقطه را بررسی میکنیم:

X1	*X2*	*X3*	*X4*	*OK?*
+1	+1	+1	+1	**YES**
+1	+1	+1	-1	**YES**
+1	+1	-1	+1	**YES**
+1	+1	-1	-1	**YES**
+1	-1	+1	+1	**YES**
+1	-1	+1	-1	**YES**
+1	-1	-1	+1	**YES**
+1	-1	-1	-1	**NO**
-1	+1	+1	+1	**YES**
-1	+1	+1	-1	**YES**
-1	+1	-1	+1	**YES**
-1	+1	-1	-1	**NO**
-1	-1	+1	+1	**YES**
-1	-1	+1	-1	**NO**
-1	-1	-1	+1	**NO**
-1	-1	-1	-1	**NO**

حال با تشکیل جدول مربوط به *S1* و *S2* داریم:

name	*#*	*X1*	*X2*	*X3*	X4
S1	*3*	+1	-1	-1	+1
		-1	+1	-1	+1
		-1	-1	+1	+1
$S2^+$	*4*	+1	+1	+1	+1
		+1	+1	-1	+1
		+1	-1	+1	+1
		-1	+1	+1	+1
$S2^-$	*4*	+1	+1	+1	-1
		+1	+1	-1	-1
		+1	-1	+1	-1
		-1	+1	+1	-1

همانطور که میبینیم مقدار آلفا برابر ۳ و مقدار بتا برابر ۴ است که مجموع *alpha + 2beta = 11*

حال درستی نامساوی مفروض را بررسی میکنیم:

We know that → $B(N,1) = 1,\ B(1,K>1) = 2$

So $\quad B(2,2) = 3 \ \leq B(1,1) + B(1,2) = 1 + 2$

$B(2,3) = 4 \ \leq B(1,2) + B(1,3) = 2 + 2$

$B(3,2) = 4 \ \leq B(2,1) + B(2,2) = 1 + 3$

$B(3,3) = 7 \ \leq B(2,2) + B(2,3) = 3 + 4$

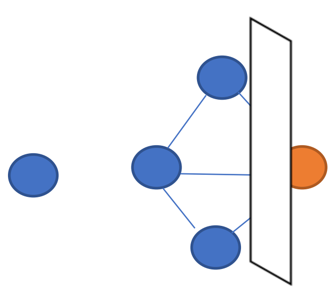
$B(4,2) = 5 \ \leq B(3,1) + B(3,2) = 1 + 4$

$B(4,3) = 11 \ \leq B(3,2) + B(3,3) = 4 + 7$

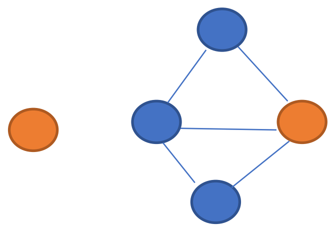
ب) در فضای سه بعدی، یک چهارضلعی با رئوسِ هر زیر مجموعه ای از ۵ نقطه موجود میسازیم. دو حالت برای نقطه پنجم در نظر میگیریم:

حالت اول: نقطه پنجم درون فضای چهارضلعی ایجاد شده قرار میگیرد: این فرض ممکن نیست زیرا در این حالت، نقطه پنجم + ۳ نقطه از ۴ نقطه انتخاب شده، خود یک چهارضلعی جدید میسازند و نقطه آخر قطعا بیرون از چهارضلعی می افتد.

حالت دوم: نقطه پنجم بیرون فضای چهارضلعی قرار میگیرد: با برهان خلفی که در حالت قبل اثبات کردیم، قطعا میدانیم نقطه پنجم باید بیرون از چهارضلعی ایجاد شده توسط ۴ نقطه انتخاب شده باشد. میدانیم در فضای دو بعدی، جدا کننده نقاط یک خط بود و میدانیم این جدا کننده در فضای سه بعدی یک صفحه است. حال دایکوتومی ای را در نظر میگیریم که ۳ نقطه از ۴ نقطه انتخابی که به نقطه پنجم نزدیک تر هستند، برچسب +۱ بگیرند و نقطه چهارم برچسب -۱ بگیرد. حال اگر مدل ما بخواهد ۵ نقطه را *shattered* کنم یعنی باید نقطه بیرون افتاده، بتواند هم برچسب +۱ هم -۱ به خود بگیرد یعنی صفحه ای وجود داشته باشد که در دو حالت بتواند نقاط را جدا کند. حالت اول را حالتی در نظر میگیریم که نقطه بیرون افتاده، برچسب یکسانی با سه نقطه نزدیک به خود بگیرد که در شکل زیر نشان میدهیم این ممکن است و نقضی صورت نمیگیرد:



حال حالت دیگر حالتی است که نقطه بیرون افتاده، با دورترین راس از چهارضلعی بخواهد برچسب یکسان داشته باشد. طبق شکل زیر میبینیم همچین چیزی ممکن نیست پس ۵ نقطه نمیتواند در فضای ۳ بعدی توسط مدل پرسپترون *shattered* شود.



پس *breakpoint* مدل پرسپترون در فضای سه بعدی، ۵ است.

حال برای قسمت دوم سوال داریم:

$$B(N,5) \leq \sum_{i=0}^{4} \binom{N}{i} = \binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \binom{N}{3} + \binom{N}{4}$$

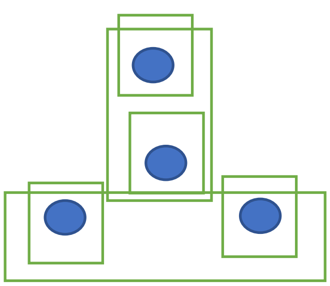
$$= 1 + N + \frac{N(N-1)}{2} + \frac{N(N-1)(N-2)}{6} + \frac{N(N-1)(N-2)(N-3)}{24}$$

$$= \; 24 + 24N + \; 12N^2 - 12N + 4N^3 - 12N^2 + 8N + N^4 - 6N^3 + 10N^2 - 6N$$

$$= \frac{N^4 - 2N^3 + 10N^2 + 14N + 24}{24}$$

سوال ۴ – مفهوم *breakpoint* یعنی کمترین تعداد نقاطی که نتوانیم هیچ چینشی از آن ها در آن فضا را پیدا کنیم، که تمام حالت های مختلف را بتوان با دسته بند *shattered* کرد.

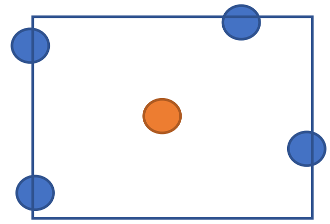
الف) با توجه به تعریف بالا، میگوییم وجود دارد ۴ نقطه ای که بتوانیم همه حالت های آن را ایجاد کنیم و دسته بند آن هارا دسته بندی کند. برای مثال:



میبینیم که هر حالتی از برچسب گذاری آن هارا بخواهیم میتوانیم توسط این دسته بند ایجاد کنیم.

اما هیچ ۵ نقطه ای وجود ندارد که بتوان آن هارا در صفحه چید و دسته بند بتواند همه حالت های آن را دسته بندی کند. به این صورت اثبات شهودی میکنیم که با ۴ نقطه که از مرکز ثقل نقاط دور تر هستند، یک مستطیل با شرایط مسئله میسازیم(نقاط روی ضلع ها) و نقطه بعدی که نزدیک ترین به مرکز ثقل هست را برچسب مخالف با بقیه نقاط میزنیم.

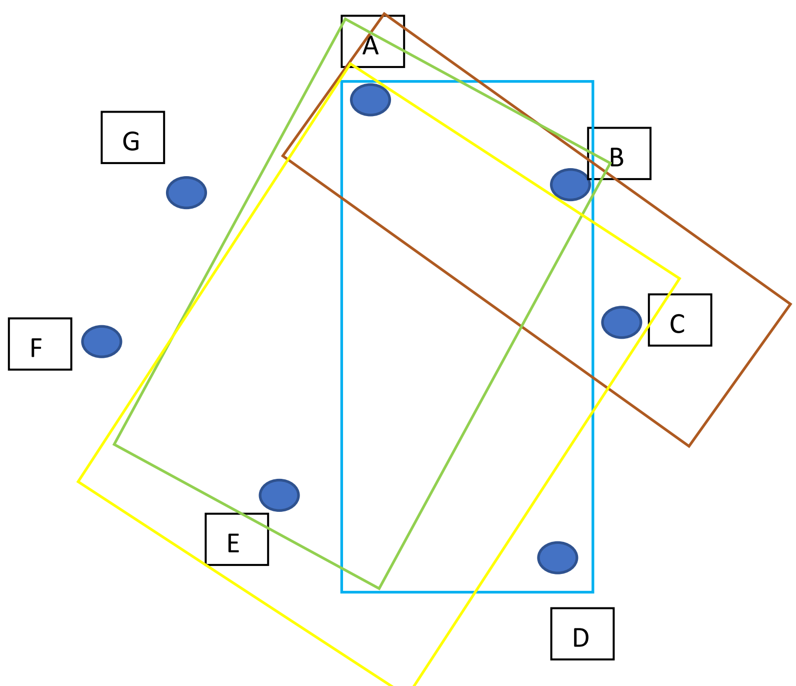
یعنی همچین شکلی میشه:



پس میفهمیم break point = 5 است. لذا داریم:

$$\mathcal{M}_{\mathcal{H}}(N) \le B(N,5) \le \sum_{i=0}^{4} \binom{N}{i} = \frac{N^4 - 2N^3 + 10N^2 + 14N + 24}{24}$$

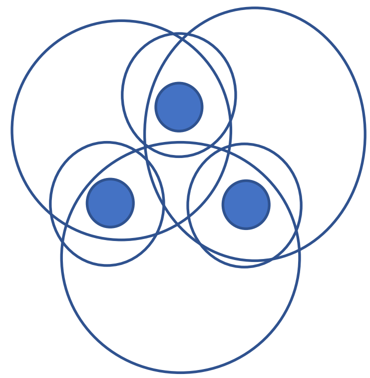
ب) فرض میکنیم یک هفت ضلعی داریم. با این ۷ ضلعی به راحتی میتوان هر زیرمجموعه ۰ یا ۱ یا ۲ یا ۶ تایی از نقاط را در برگرفت. برای ۳ نقطه، تمام *setup* ها معادل با *ABC*، *ABD*، *ABE* و یا *ACE* هستند. شکل زیر چرخاندن مستطیل را نشان میدهد که با آن بتوانیم این سه نقاط را احاطه کنیم.



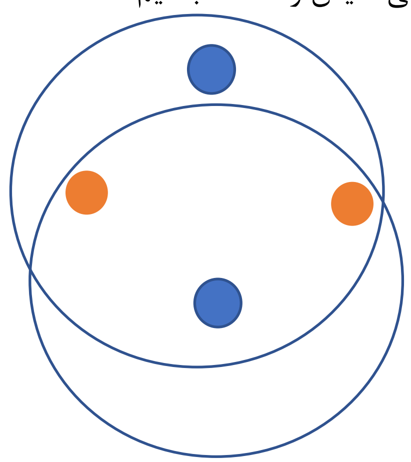
به طور مشابه ، میتوانیم مستطیل را جوری بچرخانیم که چهار نقطه را احاطه کند. اما هیچ وقت نمیتوانیم هر ۸ نقطه ای را داشته باشیم و همه حالت ها را بتوانیم دسته بندی کنیم. یعنی VC dimension = 7 لذا breakpoint = 8 است.پس داریم:

$$\mathcal{M}_{\mathcal{H}}(N) \le B(N,8) \le \sum_{i=0}^{7} \binom{N}{i} = \binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \binom{N}{3} + \binom{N}{4} + \binom{N}{5} + \binom{N}{6}$$

ج) در مورد دایره ها، واضح است که سه نقطه را میتوانیم داشته باشیم که هر دایکوتومی اش را بتوانیم داشته باشیم:



اما هیچ ۴ نقطه ای را نمیتوانیم تمام دایکوتومی هایش را داشته باشیم:



لذا *VC dimension=3* و *breakpoint = 4* است. پس داریم:

$$\mathcal{M}_{\mathcal{H}}(N) \leq B(N,4) \leq \sum_{i=0}^{3}\binom{N}{i} = \binom{N}{0} + \binom{N}{1} + \binom{N}{2} = 1 + N + \frac{N(N-1)}{2}$$
$$= \frac{N^2 + N + 2}{2}$$

سوال ۵: با جایگذاری مقادیر مفروض در سوال در سمت راست نامساوی هافدینگ داریم:

$$2 * 1 * \ e^{-2*25*10^{-4}*N} \leq \ 0.03 \ \rightarrow \ e^{-5*10^{-3}*N} \ \leq \ 0.015 \ \rightarrow \ -5 * 10^{-3} * N \ \leq \ln 0.015$$
$$\rightarrow N \ \geq \ \frac{4.2}{0.005} = \ 84$$

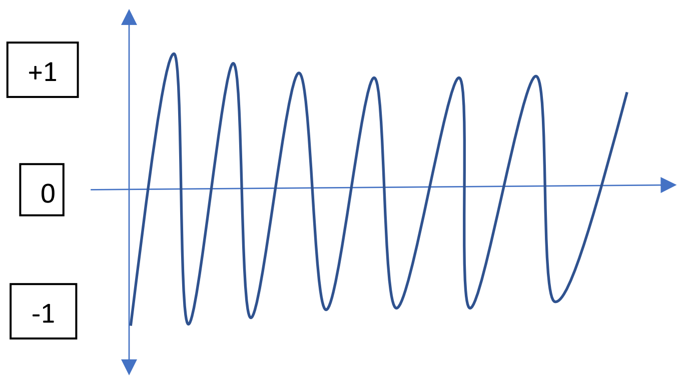
$$2 * 10 * e^{-2*25*10^{-4}*N} \le 0.03 \rightarrow e^{-5*10^{-3}*N} \le 0.0015 \rightarrow -5 * 10^{-3} * N \le \ln 0.0015 \rightarrow N \ge \frac{6.5}{0.005} = 1301$$

$$2 * 100 * e^{-2*25*10^{-4}*N} \le 0.03 \rightarrow e^{-5*10^{-3}*N} \le 0.00015 \rightarrow -5 * 10^{-3} * N \le \ln 0.00015 \rightarrow N \ge \frac{8.8}{0.005} = 2222$$

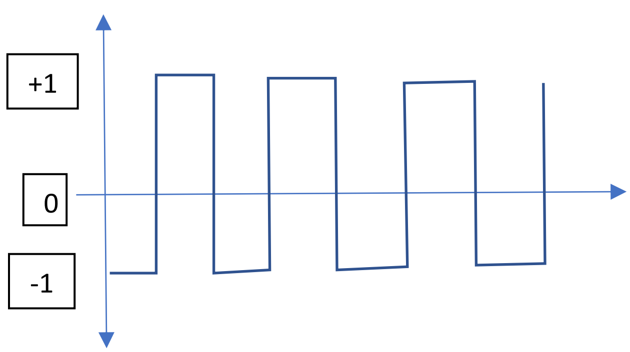
سوال ۶ –

آ)

ابتدا بخشی از تابع *sin(ax+b)* را رسم میکنیم:



حال تابع *sign(sin(ax+b))* را رسم میکنیم:



برای راحتی محاسبات، تابع *sign(sin x)* را در نظر میگیریم. برچسب نقاط *x , 2x , 3x , 4x* را به ترتیب *+,-,+,+* در نظر میگیریم. پس داریم:

$$sinx, sin2x, sin4x \ge 0 \,, sin3x < 0$$

حال مقدار *sin 4x* را با روابط مثلثاتی بر حسب زاویه *2x* بازنویسی میکنیم:

$$\sin 4x = 2\sin 2x \cos 2x \ \geq 0 \ \rightarrow \cos 2x \ \geq 0 \ \rightarrow 1 - 2sin^2 x \ \geq 0 \ \rightarrow \ sin^2 x \ \leq \ {}^1/_2$$

حال مقدار *sin 3x* را بر حسب زاویه *x* بازنویسی میکنیم:

$$\sin 3x = 3\sin x - 4\,sin^3\,x \ < 0 \ \rightarrow 3 - 4sin^2 x \ < 0 \ \rightarrow sin^2 x \ > \ {}^3/_4$$

که به نقض میرسیم. پس مدل ما امکان *shattered* کردن این ۴ زاویه را ندارد.

<u>ب)</u>

اگر *k* نقطه که *k* هر عددی بتواند باشد توسط مدل *shattered* شود، و *k+1* نقطه نتوانند *shattered* شوند، پس بعد *VC* مدل برابر *k* میشود. اگر نقاط را به فرمت $x_i = m^{-i}$ در نظر بگیریم(*m* یک عدد طبیعی زوج است) که برچسب های آن ها بتوانند به صورت دلخواه ۱+ و ۱- شوند، و مقدار *b* را صفر در نظر بگیریم، وقتی مقدار *a* به شکل زیر باشد، داریم:

$$h(x_j) = sign\left(sin\left(m^{-j}\pi + \sum_{i=1}^{k} (1 - y_j) m^{i-j} \frac{\pi}{2} \right) \right)$$

اگر مقدار برچسب نقطه ای ۱+ باشد یا *i* از *j* بزرگ تر باشد، تغییری در حاصل ایجاد نمیشود پس شرایط گفته شده را روی سیگما اعمال میکنیم:

$$h(x_j) = sign\left(sin\left(m^{-j}\pi + (1 - y_j)\frac{\pi}{2} + \ \pi \sum\nolimits_{i<j, y\neq+1} m^{i-j} \right) \right)$$

چون شرط سیگما آن است که *i* از *j* کوچک تر باشد، لذا سیگما همیشه کوچک تر از ۱ است.

حال دو حالت را بررسی میکنیم:

حالت اول: اگر برچسب داده *i* برابر ۱+ باشد، *sign* مثبت میشود لذا $h(x_i) = +1$

حالت دوم: اگر برچسب داده *i* برابر ۱- باشد، *sign* منفی میشود لذا $h(x_i) = -1$

لذا میبینیم که همه نقاط گفته شده به خوبی *shattered* میشوند بدون وابستگی به مقدار *k* یعنی*k* میتواند بینهایت نیز باشد. پس میفهمیم بعد *VC* مدل ما بینهایت است.