1

$$\mathrm{Rad}_n(\mathcal{F}) = \mathbb{E}_{\substack{S\sim P \\ \sigma\sim \mathrm{Rad}}}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \underbrace{f(z_i)}_{f(z_0)}\sigma_i\right]$$

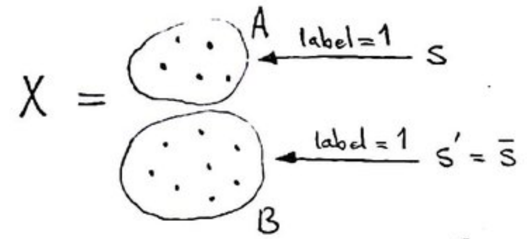
$$\overset{-1\leq f\leq 1}{=\!=} \mathbb{E}_{\substack{S\sim P \\ \sigma\sim \mathrm{Rad}}}\left[\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\right|\right] \overset{\mathrm{I}}{\leq} \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$$

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\sigma_i\right)^2\right] \overset{iid}{=} \mathbb{E}\left[\sum_{i=1}^{n}\underbrace{\sigma_i^2}_{1}\right] = n \longrightarrow \sqrt{\mathbb{E}\left[\left(\sum_{i=1}^{n}\sigma_i\right)^2\right]} = \sqrt{n} \overset{\text{Jensen}}{\geq} \mathbb{E}\left[\left|\sum_{i=1}^{n}\sigma_i\right|\right] \quad \mathrm{I}$$

2.a

X را یک مجموعه m تایی دلخواه در نظر می‌گیریم که S آنرا shatter می‌کند.

بنابراین قطعاً برای زیرمجموعه‌ای از X مانند A ، $s\in S$ ای وجود دارد که $s\cap X = A$ یعنی s به اعضای A لیبل 1 بدهد



حال به ازای s یک $s'\in S^c$ خواهیم داشت که برعکس s به $B = X - A$ لیبل 1 می‌دهد یعنی $s'\cap X = B$

بنابراین چون S ، X را shatter می‌کند و برای تمام حالات لیبل زنی، $s\in S$ پیدا می‌شود

واضح است که ما به ازای هر s ، s' ای نیز وجود دارد که حالت برعکس آن را لیبل می‌زند و قدرت لیبل زنی S^c حداقل به اندازه S است

$$\longrightarrow \forall X \quad \pi_{S^c}(X) \geq \pi_S(X)$$

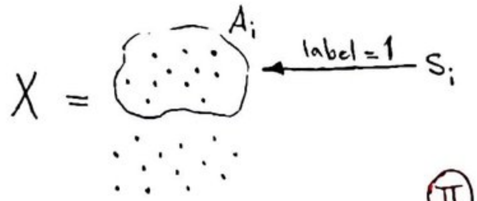
همچنین با توجه به اینکه $(S^c)^c = S$ خواهیم داشت:

$$\forall X \quad \pi_S(X) \geq \pi_{S^c}(X)$$

بنابراین اثبات کردیم که $d_{VC}(S^c) = d_{VC}(S)$ پس بعد VC برای S^c نیز محدود است.

2.b

X را مجموعه m تایی دلخواه در نظر می‌گیریم و اشتراک هر $S_i \in S$ با X را A_i می‌نامیم.
یعنی با فرض اینکه S تا k dichotomy روی X داشته باشد $A_1 \dots A_k$ را خواهیم داشت



حال با توجه به اینکه نحوه چینش در X ممکن است بدترین حالت نباشد داریم:

$$\text{(I)} \quad k \leq \Pi_S(X) \leq \Pi_S(m)$$

و به همین صورت و به دلیل کمتر بودن اندازه A_i از m داریم: $\text{(II)} \quad \Pi_T(A_i) \leq \Pi_T(m)$

به طور کلی اشتراک S, T را می‌توان به صورت $T \circ S$ نوشت یعنی T خروجی‌های S را لیبل زنی کند
بنابراین درباره اشتراک $t \in T$ و A_i ها صحبت خواهیم کرد. کل حالات لیبل زنی $T \cap S$ روی X کمتر از جمع حالات لیبل زنی T روی هر کدام از A_i هاست:

$$\underbrace{\text{کل حالات لیبل زنی } T\cap S \text{ روی } X}_{\Pi_{T\cap S}(X)} \leq \sum_{i=1}^{k} \underbrace{\Pi_T(A_i)}_{\leq \Pi_T(m) \ \text{(II)}} \leq \underbrace{k}_{\leq \Pi_S(m)\ \text{(I)}} \Pi_T(m) \longrightarrow \Pi_{T\cap S}(X) \leq \Pi_S(m)\, \Pi_T(m)$$

حال چون هیچ فرض خاصی روی نحوه چینش X نکردیم، بدترین حالت را نیز شامل می‌شود و داریم:

$$\Pi_{T\cap S}(m) \leq \Pi_S(m)\, \Pi_T(m)$$

بنابراین چون S , T بُعد VC محدود دارند طبق لم Sauer می‌توان گفت $T\cap S$ نیز breakpoint محدود دارد

2.c

واضح است که وقتی اجتماع S, T استفاده شود قدرت لیبل زنی نیز در بهترین حالت که اشتراکی نداشته باشند، جمع خواهد شد. پس طبق union bound داریم:

$$\Pi_{S\cup T}(m) \leqslant \underbrace{\Pi_S(m)}_{\sum_{i=0}^{d_S}\binom{m}{i}} + \underbrace{\Pi_T(m)}_{\sum_{i=0}^{d_T}\underbrace{\binom{m}{i}}_{\binom{m}{m-i}}} = \sum_{i=0}^{d_S}\binom{m}{i} + \sum_{i=m-d_T}^{m}\binom{m}{i}$$

بنابراین اگر $\underbrace{m-d_T \geqslant d_S+2}_{m\geqslant d_S+d_T+2}$ باشد سمت راست نامساوی از $\underbrace{\sum_{i=0}^{m}\binom{m}{i}}_{2^m}$ کمتر می‌شود که یعنی در $m=d_S+d_T+2$

break point اتفاق می افتد $\longleftarrow$ $d_{VC}(S\cup T) = \underbrace{d_S}_{d_{VC}(S)} + \underbrace{d_T}_{d_{VC}(T)} + 1$ $\longrightarrow$ پس بعد VC برای $S\cup T$ محدود است

3.a

$$Rad_n(\mathrm{conv}(\mathcal{F})) = \mathbb{E}_{\substack{S\\ \sigma}}\left[\sup_{f\in \mathrm{conv}(\mathcal{F})} \frac{1}{n}\sum_{i=1}^{n}\sigma_i f(x_i)\right]$$

$N = |\mathcal{F}|$ (I)

$\Delta = \{\alpha\in\mathbb{R}^N \mid \sum_{i=1}^{N}\alpha_i = 1 \ ,\ \alpha_i \geqslant 0\}$

$$\overset{\mathrm{I}}{=} \mathbb{E}_{\substack{S\\ \sigma}}\left[\sup_{\alpha\in\Delta} \frac{1}{n}\sum_{i=1}^{n}\sigma_i \sum_{j=1}^{N}\alpha_j f_j(x_i)\right]$$

$$= \mathbb{E}_{\substack{S\\ \sigma}}\left[\sup_{\alpha\in\Delta} \sum_{j=1}^{N}\sum_{i=1}^{n}\frac{1}{n}\sigma_i\alpha_j f_j(x_i)\right]$$

$$= \mathbb{E}_{\substack{S\\ \sigma}}\left[\sup_{\alpha\in\Delta} \sum_{j=1}^{N}\alpha_j\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i f_j(x_i)\right)\right]$$

چون $\alpha_j \geqslant 0$, $\sum\alpha_j = 1$ بنابراین sup بیشترین وزن یعنی 1 را به بزرگترین مقدار $\frac{1}{n}\sum\sigma_i f_j(x_i)$ می دهد و بقیه را 0 می کند (II)

$$\overset{\mathrm{II}}{=} \mathbb{E}_{\substack{S\\ \sigma}}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i f(x_i)\right] = Rad_n(\mathcal{F})$$

$\boxed{3.b}$ $$Rad_n(\mathcal{F}+\mathcal{G}) = \mathbb{E}_{S,\sigma}\left[\sup_{f+g\in\mathcal{F}+\mathcal{G}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i (f+g)(x_i)\right]$$

$$= \mathbb{E}_{S,\sigma}\left[\sup_{f\in\mathcal{F},\, g\in\mathcal{G}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i f(x_i) + \frac{1}{n}\sum_{i=1}^{n}\sigma_i g(x_i)\right]$$

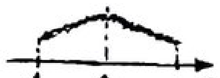
$\sup a+b \leqslant \sup a + \sup b$

$$\leqslant \underbrace{\mathbb{E}_{S,\sigma}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i f(x_i)\right]}_{Rad_n(\mathcal{F})} + \underbrace{\mathbb{E}_{S,\sigma}\left[\sup_{g\in\mathcal{G}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i g(x_i)\right]}_{Rad_n(\mathcal{G})}$$

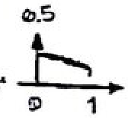
مثالی که تساوی نامساوی فوق رخ دهد :

$$\mathcal{F}_1 = \{f(x)=c \mid c\in[-1,0]\} \longrightarrow Rad_n(\mathcal{F}_1) = \mathbb{E}_\sigma\Big[\underbrace{\sup_{c\in[-1,0]} \frac{c}{n}\sum_{i=1}^{n}\sigma_i}_{\max(0,\, -\frac{1}{n}\sum_{i=1}^{n}\sigma_i)}\Big]$$

اگر $\frac{1}{n}\sum\sigma_i$ را z بنامیم می‌دانیم که z حول 0 تقارن دارد



پس $\max(0, -\frac{1}{n}\sum\sigma_i)$ شبیه



می‌شود

که $\mathbb{E}$ آن نصف $\mathbb{E}[|z|]$ است. پس داریم

$$\mathbb{E}_\sigma\left[\max(0, -\frac{1}{n}\sum_{i=1}^{n}\sigma_i)\right] = \frac{1}{2}\mathbb{E}_\sigma\left[\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\right|\right]$$

$$\mathcal{G}_1 = \{g(x)=c' \mid c'\in[0,1]\} \xrightarrow{\text{مشابه } \mathcal{F}_1} Rad_n(\mathcal{G}_1) = \mathbb{E}_\sigma\left[\max(0, \frac{1}{n}\sum_{i=1}^{n}\sigma_i)\right]$$

$$= \frac{1}{2}\mathbb{E}_\sigma\left[\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\right|\right]$$

$$Rad_n(\mathcal{F}_1+\mathcal{G}_1) = \mathbb{E}_\sigma\left[\sup_{c''\in[-1,1]} \frac{c''}{n}\sum_{i=1}^{n}\sigma_i\right] = \mathbb{E}_\sigma\left[\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\right|\right]$$

پس نشان دادیم $Rad_n(\mathcal{F}_1+\mathcal{G}_1) = Rad_n(\mathcal{F}_1) + Rad_n(\mathcal{G}_1)$ و تساوی برقرار شد

$\boxed{3.c}$

$$\mathrm{Rad}_n(\mathcal{F}+g) = \mathbb{E}_{\substack{S\\ \sigma}}\left[\frac{1}{n}\sup_{g,f\in\{g\},\mathcal{F}}\sum_{i=1}^{n}\big(f(x_i)+g(x_i)\big)\sigma_i\right]$$

$$\leqslant \underbrace{\mathbb{E}_{\substack{S\\ \sigma}}\left[\frac{1}{n}\sup_{f\in\mathcal{F}}\sum_{i=1}^{n} f(x_i)\sigma_i\right]}_{\mathrm{Rad}_n(\mathcal{F})} + \underbrace{\mathbb{E}_{\substack{S\\ \sigma}}\left[\left|\frac{1}{n}\sum_{i=1}^{n} g(x_i)\sigma_i\right|\right]}_{\mathbb{E}_{\substack{S\\ \sigma}}\left[\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\sigma_i\right)^2}\right] \overset{\text{Jensen}}{\leqslant} \sqrt{\mathbb{E}_{\substack{S\\ \sigma}}\left[\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\sigma_i\right)^2\right]}}$$

$$= \sqrt{\mathbb{E}_{\substack{S\\ \sigma}}\left[\frac{1}{n^2}\sum \underbrace{g(x_i)^2}_{\leqslant \|g\|_\infty^2}\underbrace{\sigma_i^2}_{1}\right]} = \sqrt{\frac{\|g\|_\infty^2}{n}}$$

$$\longrightarrow \mathrm{Rad}_n(\mathcal{F}+g) \leqslant \mathrm{Rad}_n(\mathcal{F}) + \frac{\|g\|_\infty}{\sqrt{n}}$$

4.a

می‌خواهیم نشان دهیم که $d_{VC}(H) = 3$

circles in $\mathbb{R}^2$

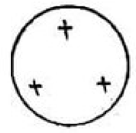
lower bound

ابتدا نشان می‌دهیم مثالی وجود دارد که 3 سمپل را در فضای $\mathbb{R}^2$ بچینیم و H بتواند آن را shatter کند

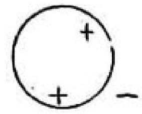
یعنی به ازای تمام $2^3 = 8$ حالت ممکن برای لیبل‌ها، H بتواند سمپل‌ها را دسته‌بندی کند

به هر صورت که نقاط را در feature space قرار دهیم، حالات زیر متصور است

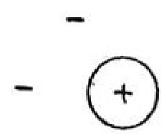
- هر 3 نقطه لیبل + — 1 حالت



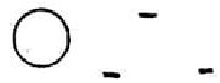
- دو نقطه لیبل + — 3 حالت



- یک نقطه لیبل + — 3 حالت



- تمام نقاط لیبل − — 1 حالت

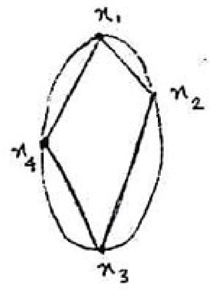


هر 8 حالت دسته‌بندی شد

$d_{VC}(H) \geq 3$ (I)

upper bound

حال باید نشان دهیم که اگر 4 سمپل را در فضای $\mathbb{R}^2$ بچینیم، به هر نحوی که این چیدمان را انجام دهیم، H نمی‌تواند تمام $2^4=16$ حالت ممکن برای لیبل‌ها را دسته‌بندی کند. (این 4 نقطه را x_1، x_2، x_3، x_4 می‌نامیم)



اگر این 4 نقطه روی مرز یک convex hull قرار نداشته باشند، از همان ابتدا معلوم است که حداقل یکی از آن‌ها بین دیگران قرار می‌گیرد و به این صورت دایره‌ای وجود نخواهد داشت که این نقطه را مخالف 3 نقطه دیگر لیبل بزند. پس فرض می‌کنیم که 4 نقطه روی مرز یک convex hull قرار دارند، حال فرض می‌کنیم که ترتیب قرار گرفتن نقاط روی convex hull مطابق شکل رسم شده باشد.
می‌دانیم که در چهارضلعی تشکیل شده حتماً یک جفت از زوایای روبه‌رو جمعشان کمتر مساوی $180°$ است (در این شکل $\hat{x}_1 + \hat{x}_3 \leq 180°$)
حال می‌خواهیم ثابت کنیم که هیچ دایره‌ای نخواهیم داشت که x_1 و x_3 را شامل شود اما هیچ کدام x_2 یا x_4 را شامل نشود. فرض کنیم x_2 بخواهد خارج این دایره باشد پس بنابراین کمان دایره که روبه‌روی x_2 است و بین x_1 و x_3 قرار دارد زاویه‌ای کمتر از $180 - \hat{x}_2$ خواهد داشت.
در حقیقت تمام نقاط بیرون این دایره، زاویه‌ای که با x_1 و x_3 می‌سازند کمتر از $180-\hat{x}_2$ خواهد بود و تمام نقاط داخل دایره، زاویه‌ای که با x_1 و x_3 می‌سازند بیشتر از $180 - \hat{x}_2$ خواهد بود. از آن جا که $\hat{x}_1+\hat{x}_3 \leq 180°$ پس $\hat{x}_2+\hat{x}_4 \geq 180°$ ← $x_4 \geq 180-\hat{x}_2$
بنابراین x_4 داخل دایره قرار می‌گیرد. به همین صورت نیز اگر x_4 بخواهد خارج دایره باشد، x_2 داخل آن قرار می‌گیرد.

پس نشان دادیم که امکان ندارد x_1 و x_3 را یک لیبل بزنیم و x_2 و x_4 را لیبلی مخالف آن ← $d_{VC}(H) < 4$ (II)

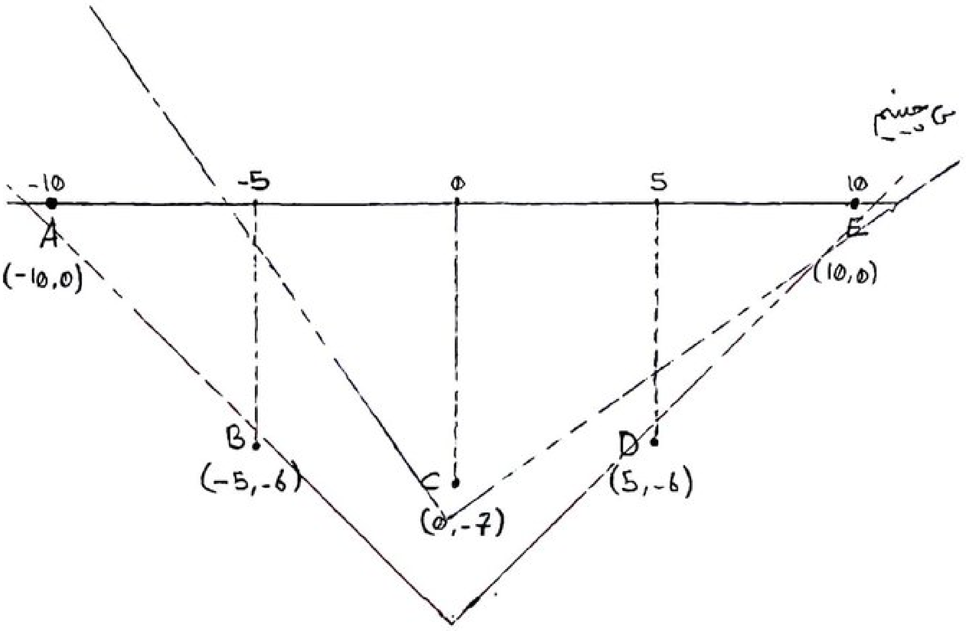
(I)(II) $\Longrightarrow \boxed{d_{VC}(H) = 3}$

4.b

می‌خواهیم نشان دهیم که $d_{VC}(\mathcal{H}) = 5$

$\mathcal{H}$: squares in $\mathbb{R}^2$

lower bound — 5 نقطه را به شکل زیر در صفحه می‌چینیم



می‌توان نشان داد که تمام 32 حالت لیبل زنی توسط مربع را می‌توان روی این چینش نقاط به دست آورد. به عنوان نمونه 2 حالت چالشی‌تر

A	B	C	D	E
0	0	1	0	1

و

A	B	C	D	E
1	0	1	0	1

رسم شده است

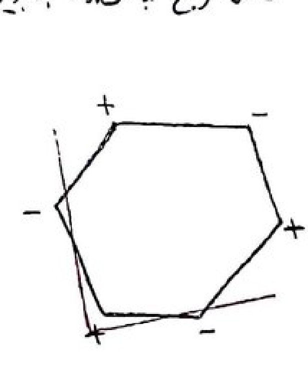
(I) $d_{VC}(\mathcal{H}) \geq 5$

upper bound — حال باید نشان دهیم اگر 6 سمپل را در صفحه بچینیم، به هر نحوی که این چیدمان را انجام دهیم، $\mathcal{H}$ نمی‌تواند تمام $2^6 = 64$ حالت ممکن را لیبل زنی کند. (این نقاط را به ترتیب x_1 تا x_6 می‌نامیم)

اگر این 6 نقطه روی مرز یک convex hull قرار نداشته باشند از همان ابتدا معلوم است که حداقل یکی از آن ها بین دیگران قرار دارد و اگر مربعی، نقاط بیرونی را در بر بگیرد حتماً نقطه داخلی را هم در بر می‌گیرد

حال اگر 6 نقطه روی مرز یک convex hull قرار گیرند نیز تشکیل یک 6 ضلعی می‌دهند که با توجه به تقارن و شکل مربع، امکان ندارد 3 رأس یکی در میان داخل مربع قرار گیرند و هیچ کدام از رئوس دیگر داخل مربع نباشند. بنابراین نشان دادیم نمی‌توان تمام 2^6 حالت ممکن را به دست آورد

$d_{VC}(\mathcal{H}) < 6$ (II)



$\xrightarrow{(I),(II)} d_{VC}(\mathcal{H}) = 5$

5.a

فرم اصلی مسئله

$$\min_{\alpha, b, \xi} \frac{1}{2} \sum_{i=1}^{m} \alpha_i^2 + C \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \quad y_i \left(\left(\sum_{j=1}^{m} \alpha_j y_j x_j \right) \cdot x_i + b \right) \geq 1 - \xi_i \qquad \alpha_i \geq 0, \ \xi_i \geq 0 \ , \ i \in [m]$$

حال اگر تعریف کنیم
$$x'_i = \begin{bmatrix} y_1 (x_1 \cdot x_i) \\ \vdots \\ y_m (x_m \cdot x_i) \end{bmatrix}$$
می توان مسئله را به فرم زیر بازنویسی کرد :

بازنویسی مسئله

$$\min_{\alpha, b, \xi} \frac{1}{2} \|\alpha\|_2^2 + C \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \quad y_i (\alpha \cdot x'_i + b) \geq 1 - \xi_i \qquad \alpha_i \geq 0 \ , \ \xi_i \geq 0 \ , \ i \in [m]$$

که همان فرم اصلی SVM است و به جای $\vec{w}$ ، $\vec{\alpha}$ دارد

5.b

lagrange variables $\longrightarrow p_i \ , \ q_i \ , \ r_i \geq 0$

$$L(\alpha, b, \xi, p, q, r) = \frac{1}{2} \|\alpha\|_2^2 + C \sum_{i=1}^{m} \xi_i - \underbrace{\sum_{i=1}^{m} p_i \left(y_i \left[\left(\sum_{j=1}^{m} \alpha_j y_j x_j \right) \cdot x_i + b \right] + \xi_i - 1 \right)}_{(\sum \alpha_i y_i x_i) \cdot (\sum p_i y_i x_i) + \sum p_i y_i b + \sum p_i \xi_i - \sum p_i}$$

$$- \sum_{i=1}^{m} q_i \xi_i - \sum_{i=1}^{m} r_i \alpha_i$$

$$\nabla_{\alpha_i} L = 0 \longrightarrow \alpha_i = (y_i x_i) \cdot \left(\sum_{j=1}^{m} p_j y_j x_j \right) + r_i$$

$$\nabla_b L = 0 \longrightarrow \sum_{i=1}^{m} p_i y_i = 0$$

$$\nabla_{\xi_i} L = 0 \longrightarrow p_i + q_i = C$$

5.b ادامه

$$\xrightarrow{\text{جاگذاری}} L(\alpha, b, \xi, p, q, r) = \frac{1}{2}\|\alpha\|_2^2 + \cancel{C\sum_{i=1}^{m}\xi_i} - \sum_{i=1}^{m}\alpha_i(\alpha_i - r_i)$$

$$\cancel{-\sum_{i=1}^{m} p_i y_i b} + \sum_{i=1}^{m} p_i \cancel{- \sum_{i=1}^{m} p_i \xi_i} \cancel{- \sum_{i=1}^{m} q_i \xi_i} - \sum_{i=1}^{m} r_i \alpha_i$$

$$= \frac{1}{2}\|\alpha\|_2^2 - \sum_{i=1}^{m}\alpha_i^2 + \sum_{i=1}^{m} p_i = -\frac{1}{2}\|\alpha\|_2^2 + \sum_{i=1}^{m} p_i$$

$$\overset{\mathrm{I}}{=} -\frac{1}{2}\left\| \sum_{j=1}^{m} p_j y_j x'_j + r \right\|_2^2 + \sum_{i=1}^{m} p_i$$

$$\mathrm{I} \quad \alpha = \begin{bmatrix} \sum_{j=1}^{m} p_j y_j \overbrace{y_1 x_1 \cdot x_j}^{x'_{j,1}} + r_1 \\ \vdots \end{bmatrix}$$

بنابراین در فرم dual باید عبارت به دست آمده را با قید های $0 \leq p_i \leq C$ و $r_i \geq 0$ و $\sum_{i=1}^{m} p_i y_i = 0$ بیشینه کرد

5.c

حالت 1 : اگر منظور بیشینه کردن $\frac{1}{2}\|\alpha\|_1^2$ باشد داریم :

$$L(\alpha, b, \xi, p, q, r) = \frac{1}{2}\|\alpha\|_1^2 + C\sum_{i=1}^{m}\xi_i - \left(\sum_{i=1}^{m}\alpha_i y_i x_i\right)\cdot\left(\sum_{i=1}^{m} p_i y_i x_i\right) - \sum_{i=1}^{m} p_i y_i b$$

$$- \sum_{i=1}^{m} p_i \xi_i + \sum_{i=1}^{m} p_i - \sum_{i=1}^{m} q_i \xi_i - \sum_{i=1}^{m} r_i \alpha_i$$

$$\nabla_{\alpha_i} L = 0 \rightarrow 2\left(\sum_{j=1}^{m}\alpha_j\right) - (y_i x_i)\cdot\left(\sum_{j=1}^{m} p_j y_j x_j\right) - r_i = 0$$

$$\rightarrow \sum_{j=1}^{m} p_j y_j y_i x_i \cdot x_j = 2\sum_{j=1}^{m}\alpha_j - r_i \quad \circledast$$

$$\nabla_b L = 0 \rightarrow \sum_{i=1}^{m} p_i y_i = 0$$

$$\nabla_{\xi_i} L = 0 \rightarrow p_i + q_i = C$$

$\boxed{5.c}$ ادامه

$$\xrightarrow{\text{جایگذاری}} L(\alpha, b, \xi, p, q, r) = \frac{1}{2}\|\alpha\|_1^2 + \cancel{C\sum_{i=1}^{m}\xi_i} - \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i y_i p_j y_j x_i \cdot x_j$$

$$-\cancel{\sum_{i=1}^{m} p_i y_i b} - \cancel{\sum_{i=1}^{m} p_i \xi_i} + \sum_{i=1}^{m} p_i - \cancel{\sum_{i=1}^{m} q_i \xi_i} - \sum_{i=1}^{m} r_i \alpha_i$$

$$= \frac{1}{2}\|\alpha\|_1^2 \underbrace{- \sum_{i=1}^{m}\alpha_i \left(2\sum_{j=1}^{m}\alpha_j - r_i\right)}_{-2\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j}_{\|\alpha\|_1^2} + \cancel{\sum_{i=1}^{m}\alpha_i r_i}} + \sum_{i=1}^{m} p_i - \cancel{\sum_{i=1}^{m} r_i \alpha_i}$$

$$= \frac{-3}{2}\|\alpha\|_1^2 + \sum_{i=1}^{m} p_i$$

بنابراین در فرم dual عبارت به‌دست آمده را با قید های $0 \leq p_i \leq C$ و $r_i \geq 0$ و $\sum_{i=1}^{m} p_i y_i = 0$ بیشینه می‌کنیم

حالت 2 : اگر منظور بیشینه کردن $\|\alpha\|_1$ باشد داریم:

$$L(\alpha, b, \xi, p, q, r) = \|\alpha\| + C\sum_{i=1}^{m}\xi_i - \left(\sum_{i=1}^{m}\alpha_i y_i x_i\right)\cdot\left(\sum_{i=1}^{m} p_i y_i x_i\right) - \sum_{i=1}^{m} p_i y_i b$$

$$-\sum_{i=1}^{m} p_i \xi_i + \sum_{i=1}^{m} p_i - \sum_{i=1}^{m} q_i \xi_i - \sum_{i=1}^{m} r_i \alpha_i$$

$$\nabla_{\alpha_i} L = 0 \longrightarrow 1 - (y_i x_i)\cdot\left(\sum_{j=1}^{m} p_j y_j x_j\right) - r_i = 0$$

$$\longrightarrow \sum_{j=1}^{m} p_j y_j y_i \, x_i \cdot x_j = 1 - r_i$$

$$\nabla_b L = 0 \longrightarrow \sum_{i=1}^{m} p_i y_i = 0$$

$$\nabla_{\xi_i} L = 0 \longrightarrow p_i + q_i = C$$

$\boxed{5.c}$ ادامه

$$\xrightarrow{\text{جاگذاری}} L(\alpha, b, \xi, p, q, r) = \|\alpha\|_1 + \cancel{C\sum_{i=1}^{m}\xi_i} - \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i y_i p_j y_j x_i \cdot x_j$$

$$- \cancel{\sum_{i=1}^{m} p_i y_i b} - \cancel{\sum_{i=1}^{m} p_i \xi_i} + \sum_{i=1}^{m} p_i - \cancel{\sum_{i=1}^{m} q_i \xi_i} - \sum_{i=1}^{m} r_i \alpha_i$$

$$= \cancel{\|\alpha\|_1} \underbrace{- \sum_{i=1}^{m}\alpha_i(1-r_i)}_{\underbrace{-\sum_{i=1}^{m}\alpha_i}_{\cancel{\|\alpha\|_1}} + \cancel{\sum_{i=1}^{m}\alpha_i r_i}} + \sum_{i=1}^{m} p_i - \cancel{\sum_{i=1}^{m} r_i \alpha_i}$$

$$= \sum_{i=1}^{m} p_i$$

بنابراین در فرم dual عبارت به دست آمده را با قیدهای $0 \leqslant p_i \leqslant C$ و $\sum_{i=1}^{m} p_i y_i = 0$ بیشینه می کنیم

$\boxed{6}$

طبق صورت سوال در فضای $\mathbb{R}^n$ قرار داریم و اعضای مجموعه S را x_i می‌نامیم. همچنین بعد j ام هر x_i را به صورت x_i^j نمایش می‌دهیم.

دستگاه معادلات زیر را می‌نویسیم:

$$\begin{bmatrix} x_1^1 & \cdots & x_m^1 \\ \vdots & & \vdots \\ x_1^d & \cdots & x_m^d \\ 1 & \cdots & 1 \end{bmatrix}_{(d+1)\times m} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix} = \vec{0}$$

طبق فرض صورت سوال $m \geqslant d+2$ است و بنابراین ماتریس wide خواهد شد. می‌دانیم که چون ماتریس wide بعد را کاهش می‌دهد حتماً Nullity بیشتر از 0 دارد و بنابراین دستگاه معادلات نوشته شده دارای جواب غیر بدیهی است. این جواب را $\lambda'_1 \dots \lambda'_m$ می‌نامیم.

حال ادعا می‌کنیم که اگر S را به S^+ و S^- افراز کنیم به صورتی که تمام x_i های دارای ضریب λ'_i مثبت در S^+ باشند و به همین صورت اگر تمام x_i های دارای ضریب λ'_i منفی عضو S^- باشند، آنگاه پوش محدب S^+ و S^- اشتراک خواهد داشت.

طبق سطر آخر دستگاه معادلات داریم
$$\sum_{i=1}^{m} \lambda'_i = 0 \longrightarrow \sum_{i:\, x_i \in S^+} \lambda'_i = -\sum_{i:\, x_i \in S^-} \lambda'_i = \Lambda$$

طبق d سطر اول دستگاه معادلات داریم
$$\sum_{i=1}^{m} \lambda'_i x_i = 0 \longrightarrow \sum_{i:\, x_i \in S^+} \lambda'_i x_i = -\sum_{i:\, x_i \in S^-} \lambda'_i x_i$$

بنابراین نقطه ای مانند x' هم عضو پوش محدب S^+ است هم پوش محدب S^-

$$x' \triangleq \sum_{i:\, x_i \in S^+} \frac{\lambda'_i}{\Lambda} x_i = \sum_{i:\, x_i \in S^-} \frac{-\lambda'_i}{\Lambda} x_i$$

$\circledast$ توسط Λ ضرایب را نرمالایز کردیم که جمع‌شان 1 شود

بنابراین نشان دادیم S دارای دو زیرمجموعه مجزا (افراز) است که پوش محدب آن ها تداخل دارد

$\boxed{7}$ $\mathcal{H} = \{x \to \lceil \sin(\theta x) \rceil \mid \theta \in \mathbb{R}\}$

کافی است مجموعه ای m عضوی از سمپل ها مانند $x_1 \dots x_m$ مثال بزنیم که به ازای هر m ، عضوی از $\mathcal{H}$ باشد که آن را shatter کند. فرض می کنیم labeling دلخواه این مجموعه به صورت $y_1 \dots y_m$ باشد. $(y \in \{0,1\})$

ادعا می کنیم اگر پارامتر θ را به صورت $y'_i = 1 - y_i$ و $\theta = \pi\left(1 + \sum_{i=1}^{m} 2^i y'_i\right)$ انتخاب کنیم و سمپل ها را نیز به صورت $x_i = \{2^{-i}\}$ انتخاب کنیم، آنگاه این حالت از $\mathcal{H}$ به ازای هر m ، هر labeling دلخواهی را shatter می کند. برای هر سمپل x_j داریم:

$$\theta x_j = \pi\left(1 + \sum_{i=1}^{m} 2^i y'_i\right) 2^{-j} = \pi\left(2^{-j} + \sum_{i=1}^{m} 2^{i-j} y'_i\right)$$

$$= \pi\Big(2^{-j} + \underbrace{\sum_{i=1}^{j-1} 2^{i-j} y'_i}_{\underbrace{\sum_{i=-1}^{1-j} 2^i y'_{-i}}_{\sum_{i=1}^{j-1} 2^{-i} y'_i}} + y'_j + \underbrace{\sum_{i=j+1}^{m} 2^{i-j} y'_i}_{\underbrace{\sum_{i=1}^{m-j} 2^i y'_{i+j}}_{\text{یک عبارت زوج است که چون } \sin(2k\pi)=0 \text{ بی اثر است و حذفش می کنیم}}}\Big) = \pi\left(2^{-j} + y'_j + \sum_{i=1}^{j-1} 2^{-i} y'_i\right)$$

حال اگر در $\sum_{i=1}^{j-1} 2^{-i} y'_i$ تمام y'_i ها را که یا 0 اند یا 1 ، 1 فرض کنیم داریم:

$$\theta x_j < \pi\left(y'_j + \sum_{i=1}^{j} 2^{-i}\right) < \pi(y'_j + 1)$$

همچنین اگر $2^{-j} + \sum 2^{-i} y'_i$ را که بزرگتر مساوی 0 است حذف کنیم داریم:

$$\theta x_j > \pi y'_j$$

حال اگر لیبل واقعی x_j برابر 0 باشد یعنی $y_j = 0$ $\longrightarrow$ $y'_j = 1$ $\longrightarrow$ $\pi < \theta x_j < 2\pi$ $\longrightarrow$ $-1 < \sin(\theta x_j) < 0$

$$\longrightarrow \lceil \sin(\theta x_j) \rceil = 0$$

به طور مشابه اگر لیبل واقعی x_j برابر 1 باشد یعنی $y_j = 1$ $\longrightarrow$ $y'_j = 0$ $\longrightarrow$ $0 < \theta x_j < \pi$ $\longrightarrow$ $0 < \sin(\theta x_j) < 1$

$$\longrightarrow \lceil \sin(\theta x_j) \rceil = 1$$

$$\longrightarrow d_{VC}(\mathcal{H}) = \infty$$

$\boxed{8}$

طبق لم Sauer می دانیم که اگر $d_{VC}(\mathcal{H})=d$ آنگاه $\mathcal{H}$ مجموعه D تایی از سمپل ها را حداکثر $O(D^d)$ مدل می تواند لیبل بزند . فرض می کنیم مجموعه ای D تایی از سمپل ها داشته باشیم که $MAJ_k(\mathcal{H})$ آنها shatter کند .

حال اگر k تا عضو از $\mathcal{H}$ (با فرض مجاز بودن تکرار) برداریم ، $MAJ_k(\mathcal{H})$ در بهترین حالت که هر h_i حالت labeling جدیدی را اضافه کند حداکثر $O((D^d)^k)$ مدل می تواند لیبل بزند .

چون فرض کردیم $MAJ_k(\mathcal{H})$ D تا سمپل را shatter می کند

$$\longrightarrow D^{dk} \geqslant 2^D \longrightarrow D \leqslant kd \underbrace{\lg_2^D}_{\leqslant\sqrt{D}} \longrightarrow D \leqslant k^2 d^2$$

به ازای $D\geqslant 16$

حال نامساوی به دست آمده را برای تقریب بهتر از $\lg_2^D$ استفاده می کنیم

$$\left.\begin{array}{l} D \leqslant (kd)^2 \longrightarrow \lg_2^D \leqslant 2\lg_2^{kd} \\ D \leqslant kd\lg_2^D \end{array}\right\} \longrightarrow D \leqslant kd\left(2\lg_2^{kd}\right) \longrightarrow D \in O\left(kd\lg kd\right)$$

پس مجموعه D تایی که بتواند توسط $MAJ_k(\mathcal{H})$ shatter شود باید $D\in O(kd\lg kd)$ باشد

$$\longrightarrow d_{VC}\left(MAJ_k(\mathcal{H})\right) \in O\left(kd\lg kd\right)$$

در برخی سوال ها با علی بابایی همفکری داشتم