5.2

کلاس مستطیل های تراز محور را با H_d نشان دهید. از آنجایی که H_5 زیر مجموعه H_2 ، خطای تقریبی کلاس H_5 کوچکتر است. با این حال، پیچیدگی H_5 بیشتر است، بنابراین ما انتظار داریم که خطای تخمینی آن بزرگتر باشد. بنابراین، اگر تعداد محدودی از نمونه های آموزشی داشته باشیم، ترجیح می دهیم کلاس کوچکتر را یاد بگیریم.

6.2

در نظر میگیرم که $VCdim(H_{=k}) = \min\{k,|X| - k\}$ ابتدا نشان می دهیم که $VCdim(H{=}k)\leq k$. فرض کنید $C\subseteq X$ مجموعه ای به اندازه $k+1$ باشد. سپس، $h \in H{=}k$ وجود ندارد که $h(x) = 1$ را برای تمام $x\in C$ برآورده کند. به طور مشابه، اگر $C\subseteq X$ مجموعه‌ای از اندازه $|X|-k+1$ است، هیچ $h\in H{=}k$ وجود ندارد که $h(x)=0$ را برای همه $x\in C$ برآورده کند. بنابراین، $VCdim(H_{=k}) \leq \min\{k, |X| - k\}$.

فقط باید نشان دهیم $VCdim(H_{=k}) \geq \min\{k, |X| - k\}$.فرض کنید $C = \{x1,\dots,xm\} \subseteq X$

مجموعه ای با اندازه $m \leq \min\{k,|X| - k\}$. فرض کنید $(y_1, \dots, y_m) \in \{0, 1\}^m$ بردار برچسب ها باشد .

$\sum_{i=1}^{m} yi$ را نشان با S نشان می دهیم

یک زیرمجموعه دلخواه $E \subseteq X \setminus C$ از عناصر $k-s$ را انتخاب کنید، و اجازه دهید $h \in H{=}k$ فرضیه ای باشد که $h(xi) = yi$ را برای هر $xi \in C$ و $h(x) = 1[E]$ را برآورده می کند. هر $x \in X \setminus C$. نتیجه می گیریم که C توسط $H{=}k$ shatter می شود. بنابراین $VCdim(H{=}k) \geq \min\{k, |X| - k\}$.

b.

ادعا می کنیم که $VCdim(H{\leq}k) = k$. ابتدا نشان می‌دهیم که $VCdim(H^X_{\leq k}) \leq k$

. فرض کنید $C \subseteq X$ مجموعه ای به اندازه $k+1$ باشد. سپس، $h\in H{\leq}k$ وجود ندارد که $h(x)=1$ را برای تمام $x\in C$ برآورده کند.

فقط باید نشان دهیم $VCdim(H{\leq}k) \geq k$. فرض کنید $C = \{x1,\dots,xm\} \subseteq X$ مجموعه‌ای با اندازه $m \leq k$ باشد. فرض کنید $(y1,\dots,ym) \in \{0,1\}$ بردار برچسب‌ها باشد. این برچسب‌گذاری توسط برخی فرضیه $h\in H{\leq}k$ بدست میاد که $h(xi)=yi$ را برای هر $xi \in C$ و $h(x)=0$ را برای هر $x \in X \setminus C$ برآورده می‌کند. نتیجه می گیریم که C توسط $H{\leq}k$ shatter می شود. بدین ترتیب $VCdim(H{\leq}k) \geq k$.

6.4

فرض کنید $X = R^d$ ما نشان میدهیم که تمام ترکیبات این ۴ کلاس h رو 1 , 0 تعریف میشوند .

میدانیم که کلاس تهی شتر میشود

فرض کنید $d>2$ بنابراین کلاس $H = \{1_{[\|x\|_2\leq r]} : r \geq 0\}$ خواهد بود و vc 1 میشود برای مشاهده این ما نشان میدهیم که اگه X همه مقادیر 0 نباشد بنابراین x شتر میشود

م، اگر $\|x_1\|_2 \le \|x_2\|_2$، پس برچسب گذاری $y1 = 0, y2 = 1$ است
با هیچ فرضیه ای در H به دست نمی آید. فرض کنید $A = \{e1, e2\}$ ، که در آن
e1، e2 دو بردار یکه استاندارد R^d هستند. سپس،
$H_A = \{(0,0),(1,1)\}$, $\{B \subseteq A : H \text{ shatters } B\} = \{\emptyset,\{e_1\},\{e_2\}\}$, $\sum_{i=0}^{d}(|A| \ , i) = 3$

فرض کنید H کلاس مستطیل های تراز محور در R^2 باشد. ما
دیدیم که بعد VC مجموعه H برابر 4 است فرض کنید , $A = \{x1, x2, x3\}$ که در آن
$x_1 = (0, 0), x_2 = (1, 0), x_3 = (2, 0)$
تمام برچسب ها به جز (1، 0، 1) به دست می آیند. بنابراین، $|H_A| = 7$ ،
$|\{B \subseteq A : H \text{ shatters } B\}| = 7$
,$\sum_{i=0}^{d}(|A| \ , i) = 8$

$d \ge 3$ را در نظر بگیرید و کلاس $H = \{\text{sign}\langle w,x\rangle : w \in Rd\}2$ نیمه فاصله های همگن را در نظر بگیرید
که VC این کلاس d است. $\text{VCdim}(H) \ge 3$ بنابراین مجموعه $\{e1, e2, e3\}$ shatter شده است، فرض
کنید $A = \{x_1,x_2,x_3\}$ که در آن
$x_1 = e_1$, $x_2 = e_2$, and $x_3 = (1,1,0,\dots,0)$
. توجه داشته باشید که تمام برچسب ها به جز (1،1،-1) و (-1،-1،1) به دست می آیند. نتیجه می شود
$|H_A| = 6$, $|\{B \subseteq A : H \text{ shatters } B\}| = 7$ $\sum_{i=0}^{d}(|A| \ , i) = 8$

$d = 1$ را فرض کنید و کلاس $H = \{1_{[x\ge t]} : t \in R\}$ را در نظر بگیرید که threshold ها میباشند
هر مجموعه یکی باshatter H می‌شود، و هر مجموعه‌ای از اندازه حداقل 2 باshatter H نمی‌شود. هر
مجموعه محدود $A \subseteq R$ را انتخاب کنید. سپس هر یک از سه عبارت در «نابرابری ساور» برابر با
$|A|+1$

6.6

هر h در کنار همه h های منفی با هر x_i تخمین زده میشود هر $x_i, \bar{x}_i$ or none در رابطه ظاهر شود را در
بر میگیرد بنابراین
$|H_{con}^{d}| = 3^d + 1.$

$\text{VCdim}(H^d) \le \lfloor \log(|H^d|) \rfloor \le 3 \log d$

فرض کنید $J \subseteq [d]$ یک زیر مجموعه از ایندکس ها باشد . نشان دادیم که لیبل گذاری هر جز e_i مثبت
میباشند . اگه $J = [d]$, ما همه فرضیات مثبت را انتخاب میکنیم . اگر $J = \emptyset$, همه فرضیات منفی را

انتخاب میکنیم .. فرض کنید Ø زیر مجموعه j و زیر مجموعه ,[d] نباشد فرض کنید h کلاس مربوط به ضرب بولین ها باشد بنابران
$h(e_j) = 1$ if $j \in J$, and $h(e_j) = 0$ otherwise.

بر اساس تضاد فرض کنید که یک مجموعه $C = \{c1,...cd\}$، وجود داردکه برای آن $|H_C| = 2^d+1$. $h1$
$h_1,...,h_{d+1},$ و $l_1,...,l_{d+1}$ را مانند راهنمایی تعریف کنید. با اصل کبوتر، در میان $l_1,...,l_{d+1}$ حداقل یک متغیر دو بار رخ می دهد. فرض کنید که l1 و l2 به یک متغیر متعلق هستند. ابتدا فرض کنید $l2 = l1$. سپس، l1 در c1 درست است بنابرانی l2 در c1 درست است. با این حال، این با فرضیات ما در تضاد است. حال فرض کنید $l_1 \neq l_2$ در این مورد $h_1(c_3)$ منفی است، زیرا l2 روی c_3 مثبت است. این دوباره با فرضیات ما در تضاد است.

ابتدا مشاهده می کنیم که $|H'| = 2^d + 1$ بنابراین، $VCdim(H') \le \lfloor \log(|H|) \rfloor = d$.
ما اثبات را با نمایش یک مجموعه شترشده با اندازه d کامل می کنیم. فرض کنید
$C=\{(1,1,...,1)-e_j\}^d_{j=1} =\{(0,1,...,1),...,(1,...,1,0)\}$.

فرض کنید $J \subseteq [d]$ زیر مجموعه ای از ایندکس ها باشد. نشان خواهیم داد که برچسب گذاری دقیقاً در $\{(1,1,...,1)-e_j\}_{j\in J}$ منفی است. برای لحظه $J \neq \emptyset$ فرض کنید. سپس برچسب‌گذاری با ضرب بولین ها به دست می‌آید. در نهایت، اگر $J = \emptyset$، فرضیه همه مثبت $h_\emptyset$. را انتخاب کنید.

6.9

ما ثابت می کنیم که$VCdim(H) = 3$. را انتخاب کنید. فرض کنید$C = \{1, 2, 3\}$
جدول زیر نشان می دهد که C توسط H شتر شده است.

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

6.10

a.

فرض میکنیم که m < d ، زیرا در غیر این صورت گزاره بی معنی است. فرض کنید C یک مجموعه shatter به اندازه d باشد. X = C را از توزیع‌هایی انتخاب میکنیم که روی C متمرکز هستند. توجه داشته باشید که H شامل تمام توابع از C به {0,1} است. با فرض اینکه برای هر الگوریتم، یک توزیع D وجود دارد که برای آن$\min_{h\in H} L_D(h) = 0$

$$\mathbb{E}[L_{\mathcal{D}}(A(S))] \geq \frac{k-1}{2k} \underbrace{=}_{k=\frac{d}{m}} \frac{d-m}{2d}$$

b.

فرض کنید که .VCdim(H) = ∞ فرض کنید A یک الگوریتم یادگیری باشد. نشان می‌دهیم که A نمی‌تواند H را یاد بگیرد ε= 1/16 . δ= 1/14,را انتخاب کنید.

برای هر$m \in N$ ، یک مجموعه shatter شده به اندازه d = 2m وجود دارد. با اعمال موارد فوق، دریافتیم که توزیعی D وجود دارد که برای آن$\min_{h\in H} L_D(h) = 0$

، اما$E[L_D(A(S))] \geq 1/4$ است. بنابراین با احتمال$\delta < 1/7$:

$L_D(A(S)) - \min_{h\in H} L_D(h) = L_D(A(S)) \geq 1/8 > \varepsilon$.

6.11

a

فرض کنید که برای هر$i \in [r]$ ، .VCdim(Hi) = d ≥ 3 فرض کنید:

$$\mathcal{H} = \bigcup_{i=1}^{r} \mathcal{H}_i$$

فرض کنید $k\in[d]$ بطوریکه $k \leq 4d \log(2d) + 2 \log r$ با تعریف تابع رشد داریم

$$\tau_{\mathcal{H}}(k) \leq \sum_{i=1}^{r} \tau_{\mathcal{H}_i}(k)$$

از آنجایی که$d \geq 3$ ، با اعمال لم سائر در مورد هر یک از عبارتτ_{H_i} ، به دست می آوریم

$\tau_H(k) < rm^d$

نتیجه این است که $k < d \log m + \log r$. لم A.2 نشان میدهد که $k < 4d \log(2d) + 2 \log r$.

b

فرض کنیم که $VCdim(H1) = VCdim(H2) = d$. فرض کنید $H = H1 \cup H2$. فرض کنید k یک عدد صحیح مثبت باشد به طوری که

$k \geq 2d + 2$. نشان می دهیم که $\tau_H(k) < 2^k$. با لم سائر خواهیم داشت :

$$
\begin{aligned}
\tau_{\mathcal{H}}(k) &\leq \tau_{\mathcal{H}_1}(k) + \tau_{\mathcal{H}_2}(k) \\
&\leq \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=0}^{d} \binom{k}{i} \\
&= \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=0}^{d} \binom{k}{k-i} \\
&= \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=k-d}^{k} \binom{k}{i} \\
&\leq \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=d+2}^{k} \binom{k}{i} \\
&< \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=d+1}^{k} \binom{k}{i} \\
&= \sum_{i=0}^{k} \binom{k}{i} \\
&= 2^k .
\end{aligned}
$$

9.1

بردار متغیرهای کمکی $s = (s1, . . . , sm)$ را تعریف کنید. به دنبال راهنمایی، به حداقل رساندن خطای تجربی معادل به حداقل رساندن $\sum_{i=1}^{m} si$ تحت محدودیت زیر میباشد :

$(\forall i\in[m])\ w^T x_i - s_i \leq y_i\ ,\ -w^T x_i - s_i \leq -y_i$

باید مقدار فوق را به ماتریس تبدیل کنیم فرض کنید $A \in R^{2m\times(m+d)}$ ماتریس $A = [X\ -I_m; -X\ -I_m]$ باشد بطوریکه برای هر i عضو m داشته باشیم $X_i \rightarrow = x_i$

همچنین فرض کنید بردار $v \in R^{d+m}$ بردار مقادیر باشد $(w_1,\ldots,w_d,s_1,\ldots,s_m)$

و $b \in R^{2m}$ را به صورت $b = (y_1,\ldots,y_m,-y_1,\ldots,-y_m)^T$ تعریف میکنیم

و $c \in R^{d+m}$ را به صورت $c = (0_d;1_m)$ تعریف میکنیم

نتیجه می شود که مسئله بهینه سازی به حداقل رساندن ریسک تجربی را می توان به صورت LP زیر بیان کرد:

$\min c^T v$
s.t. $Av \le b$.

9.3

فرض کنید $d = m$، و برای هر $i \in [m]$، $x_i = e_i$. فرض میکنیم $\mathrm{sign}(0) = -1$. برای
$i = 1, ..., d$ اجازه دهید $y_i = 1$ برچسب x_i باشد. بردار وزنی پرسپترون حفظ می شود را با $w^{(t)}$ نشان دهید. یک آرگومان استقرایی ساده نشان می دهد که برای هر $i \in [d]$ داریم:

$$\mathbf{w}_i = \sum_{j<i} \mathbf{e}_j$$

بنابراین، برای هر $i \in [d]$، $\langle w^{(i)}, x_i \rangle = 0$. از این رو، تمام موارد x_1، . . .، x_d به خطا کلسیفای می‌شوند (و سپس بردار $w = (1,...,1)$ را بدست می آوریم که با $x_1,...,x_m$ سازگار است). همچنین توجه می کنیم که بردار $w^\star = (1,...,1)$ الزامات ذکر شده در سوال را برآورده می کند.

9.4

تمام نمونه‌های مثبت شکل $(\alpha,\beta,1)$ را در نظر بگیرید که $\alpha^2+\beta^2+1 \le R^2$ و $w^\star = (0, 0, 1)$ را در نظر بگیرید یه طوریکه برای هر x, y داشته باشیم $y\langle w^\star, x\rangle \ge 1$. ما دنباله‌ای از نمونه‌های R^2 را نشان می‌دهیم که Perceptron در آنها R^2 خطا را انجام می‌دهد.

ایده ساخت و ساز این است که با مثال ها شروع شود $(\alpha_1،0،1)$ که $\alpha_1 = \sqrt{R^2 - 1}$
حالا، در دور t نمونه جدید چنین باشد :

$$\alpha^2 + \beta^2 + 1 = R^2 \qquad \text{(a)}$$

$$\langle w_t,(\alpha,\beta,1)\rangle=0 \qquad \text{(b)}$$

تا زمانی که بتوانیم هر دو شرط را برآورده کنیم، پرسپترون به خطا خود ادامه خواهد داد. ما نشان خواهیم داد که تا زمانی که $t \le R^2$ بتوانیم این شرایط را برآورده کنیم.

مشاهده کنید که با استقرا، برای برخی از اعداد a, b ($w^{(t-1)} = (a, b, t - 1)$. همچنین مشاهده کنید که $\|w_{t-1}\|^2 = (t - 1)R^2$ (این از اثبات کران خطا پرسپترون، جایی که نابرابری‌ها با تساوی برقرار می‌شوند، به دست می‌آید). یعنی : $a^2+b^2+(t-1)^2 =(t-1)R^2$

فرض کنید $w^{(t-1)}$را حول محور Z بچرخانیم z به گونه ای که به شکل $(a,0,t-1)$ و خواهیم داشت :

$$a = \sqrt{(t-1)R^2 - (t-1)^2}$$

$$\alpha = -\frac{t-1}{a}\ .$$

و برای هر β : $\langle(a, 0, t-1), (\alpha, \beta, 1)\rangle = 0$ و در اینصورت فقط باید ثابت کنیم: $\alpha^2+1 < R^2$
زیرا اگر این رابطه برقرار باشد

$$\begin{aligned}\alpha^2+1 &= \frac{(t-1)^2}{a^2}+1 = \frac{(t-1)^2}{(t-1)R^2-(t-1)^2}+1 = \frac{(t-1)R^2}{(t-1)R^2-(t-1)^2}\\ &= R^2\frac{1}{R^2-(t-1)}\\ &\leq R^2\end{aligned}$$

که این نشان میدهد $R^2 \geq t$

9.6

در این سوال کلاس نیم فاصله ها در R^{d+1} را با L_{d+1} نشان می دهیم (a). فرض کنید که
$A = \{x_1,\dots,x_m\} \subseteq R^d$ توسط B_d. shatter شده است. سپس،
$\forall y = (y_1,\dots,y_d) \in \{-1,1\}^d$ وجود دارد یک $B_{\mu,r} \in B$ s.t. برای هر i $B_{\mu,r}(x_i) = y_i$
بنابراین، برای μ و r بالا، عبارت زیر برای هر $i \in [m]$ برقرار است:

$$\operatorname{sign}\left((2\mu;-1)^T(\mathbf{x}_i;\|\mathbf{x}_i\|^2) - \|\mu\|^2 + r^2\right) = y_i$$

هر جا که نشان دهنده الحاق برداری است. برای هر $i \in [m]$ ، فرض کنید $\phi(x_i) = (x_i; \|x\|^2_i)$.هاف
اسپیس $h \in L_{d+1}$ را تعریف کنید که مطابق با $w = (2\mu; -1)$, و $b = \|\mu\|^2 - r^2$ است. دهد که برای هر
$i \in [m]$ داریم : $h(x_i) = y_i$
در مجموع، اگر $A = \{x_1,\dots,x_m\}$ توسط B shatter شود، $\phi(A) := \{\phi(x_1), \dots, \phi(x_m)\}$ توسط L shatter می شود. نتیجه می گیریم که:
$d + 2 = VCdim(L_{d+1}) \geq VC(B_d)$.

مجموعه C را متشکل از بردارهای یکه $e_1, \dots, e_d$ در نظر بگیرید. فرض کنید $A \subseteq C$. نشان می دهیم که یک نقطه وجود دارد به طوری که همه بردارها در A برچسب مثبت دارند، در حالی که بردارها در $C \setminus A$ برچسب منفی دارند. مرکز را به صورت زیر تعریف می کنیم:

$\mu = \sum_{e \in A}$

توجه داشته باشید که برای هر بردار واحد درA ، فاصله آن تا مرکز $\sqrt{|A| - 1}$. همچنین، برای هر بردار واحد خارج Aفاصله آن تا مرکز $\sqrt{|A| + 1}$ و در نهایت فاصله اصلی $\sqrt{|A|}$ میباشد. بنابراین اگر $0 \in A$ باشد $r = \sqrt{|A| - 1}$ در غیر اینصورت $r = \sqrt{|A|}$

نتیجه می گیریم که مجموعه C توسط B_d shatter شده است. در مجموع، ما فقط نشان دادیم که $VCdim(B_d) \geq d + 1$.