فصل ۶

تمرین ۲_a) ابتدا نشان می‌دهیم VC_dim کمتر مساوی با k است پس در نتیجه باید نشون بدهیم که $k+1$ عضوی shatter نمی‌شود. فرض کنیم

$$|C|=k+1 \Rightarrow C\subseteq X$$

طبق تعریف تابع، پس

$$\Rightarrow \nexists h\in H \Rightarrow \forall x\in C \quad h(x)=1$$

C، shatter نمی‌شود

$$\Rightarrow VCdim(H)\leq k \quad ①$$

حال فرض کنیم $|C_2|=|X|-k+1$ در نتیجه

$$\nexists h\in H \Rightarrow \forall x\in C_2 \quad h(x)=0$$

$$\Rightarrow VCdim\leq |C_2|-1 \quad ②$$

$$①, ② \Rightarrow VCdim(H)\leq \min(\{k, |X|-k\}) \quad *$$

حال باید نشان دهیم که $VCdim(H)\geq \min(\{k,|X|-k\})$ فرض کنیم $|C|=m\leq *$ باشد حال فرض کنیم

حال $(y_1,\dots,y_m)\in\{0,1\}^m$ حال $\sum_{i=1}^{m} y_i = L$ داریم

حال زیر مجموعه‌ی $E\subseteq X-C$، $k-L$ عضو و فرض کنیم $h\in H$ که $h(x_i)=y_i$

برای هر $x\in C$ و $h(x)=1_E$ برای هر $x\in E$ در نتیجه C، shatter می‌شود و

$$VCdim \geq * \Rightarrow VCdim = (\{k,|X|-k\})$$

b_ فرض کنیم $C\subseteq X$ است و $|C|=k+1$ در نتیجه

$$\nexists h\in H \Rightarrow \forall x\in C \quad h(x)=1$$

$$\Rightarrow VCdim(H)\leq k$$

حال فرض کنیم $C\subseteq X$، $|C|=m\leq k$ حال فرض کنیم $(y_1,\dots,y_m)\in\{0,1\}^m$ باشد حال $h\in H$

$h(x_i)=y_i$ برای هر $x_i\in C$، و $h(x)=0$ برای هر $x\in X-C$ در نتیجه C shatter می‌شود

$$\Rightarrow VCdim = k$$

تمرین ۶) در هر h، یا x یا $\bar{x}$ یا هیچکدام از ها ظاهر می‌شوند ~~[illegible]~~
و همچنین حالت h همه صفر است ($x_i \wedge \bar{x}_i$) پس در نتیجه کل حالات می‌شود $|H^d_{con}| = 3^d + 1$

b)

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

تمرین ۱) فرض کنیم C مجموعه‌ای از اندازه d باشد که shattered می‌شود. حال ما می‌توانیم فرض کنیم $X = C$

حال با توجه تمرین ۳ در بخش ۵ برای هر الگوریتم وجود دارد یک توزیع D برای هر کدام ($\min_{h \in H} L_D(h) = 0$)

اما $(m < d)$ $E[L_D(A(S))] \ge \frac{k-1}{2k} \overset{k=\frac{d}{m}}{=} \frac{d-m}{2d}$

Scanned with CamScanner

۸.۱) b) فرض کنیم $VCdim = \infty$ باشد حال فرض کنیم الگوریتم A را داریم نشان می دهیم H PAC learner

نیست. انتخاب کنیم $\epsilon = \frac{1}{16}$ و $\delta = \frac{1}{7}$ برای هر $m \in N$ درنتیجه مجموعه ای از سایز $2m$ وجود دارد که shatter

می شود. درنتیجه وجود دارد توزیعی که $L_D(h) = 0$ $h \in H$ اما $E(L_D(A(S))) \geq \frac{1}{4}$ و

$\frac{1}{7} < \delta$ که درنتیجه

$L_D(A(S)) - \min_{h \in H} L_D(h) = \frac{1}{4} > \frac{1}{8} > \epsilon$

پس فرض خلف باطل است.

۱۱-۱) فرض می کنیم $VCdim(H_i) \leq d$ و $r \geq d$ است حال فرض کنیم که $H = \cup H_i$, $k \in [d]$

همچنین طبق راهنمایی فرض کنیم $\tau_H(k) = 2^k$ حال با توجه به تابع داریم

$$\tau_H(k) \leq \sum_{i=1}^{r} \tau_{H_i}(k)$$

حال با توجه به لم Sauer داریم $\tau_H(k) \leq r m^d$ حال با گرفتن log داریم:

$$k < d \log m + \log r \Rightarrow k < 4d \log 2d + 2 \log r$$

Scanned with CamScanner

تمرین ۶) ابتدا نشان می‌دهیم که مجموعه‌ی ۳ عضوی {۱, ۲, ۳}، shatter می‌شود به وسیله‌ی H داریم:

۱	۲	۳	a	b	s
−	−	−	۰٫۵	۳٫۵	
−	−	+	۲٫۵	۳٫۵	
−	+	−	۱٫۵	۲٫۵	
−	+	+	۱٫۵	۴٫۵	
+	−	−	۰٫۵	۱٫۵	
+	−	+	۱٫۵	۲٫۵	
+	+	−	۰٫۵	۲٫۵	
+	+	+	۰٫۵	۳٫۵	

2^3

حال که نشان دادیم که $VCdim \geq 3$ است حال فرض کنیم $C=\{x_1, x_2, x_3, x_4\}$ و فرض کنیم که $x_1 < x_2 < x_3 < x_4$ در این صورت نمی‌توانیم لیبل $y_1 = y_3 = -1$ و $y_2 = y_4 = 1$ را به وسیله‌ی h در H بسازیم در نتیجه $VCdim(H) = 3$

تمرین ۷) در هر

(فصل ۹)

۱- ابتدا بردارهای $s_1, \dots, s_m$ را تعریف می کنیم. حال با توجه به راهنمای سوال می نیمم کردن خطا معادل با می نیمم کردن چندجمله ای $\sum_{i=1}^{m} s_i$ هست با شرایط

$$\forall i \in [m] \quad w^T x_i - s_i \le y_i$$

$$-w^T x_i - s_i \le -y_i$$

حال فرض کنیم $A = [X \; -I_m ; -X \; -I_m]$ ماتریسی باشد که $X_i = x_i$ برای هر $i \in [m]$، $A \in \mathbb{R}^{2m \times (m+d)}$

حال فرض کنیم $v \in \mathbb{R}^{d+m}$ برداری باشد با مقادیر $(w_1, \dots, w_d, s_1, \dots, s_m)$. $b \in \mathbb{R}^{2m}$ را بردار $(y_1, \dots, y_m, -y_1, \dots, -y_m)$ تعریف می کنیم. در نتیجه فرض کنیم $c \in \mathbb{R}^{d+m}$ برداری باشد که $(0_d, 1_m)$. حال می نیمم سازی خطا به شکل یک مسئله بهینه سازی خطی به شکل زیر می شود:

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

۳- با توجه به راهنمایی فرض کنیم $d=m$ باشد و برای هر $i \in [m]$ فرض کنیم $x_i = e_i$. برای $i = 1, \dots, d$ فرض کنیم $y_i = 1$ لیبل های x_i باشد. حال $w^{(t)}$ خروجی الگوریتم Perceptron. فرض کنیم که $sign(0) = -1$ است. حال با یک حساب ساده می بینیم که برای هر $i \in [d]$ داریم $w_i = \sum_{j<i} e_j$. در نتیجه برای هر $i \in [d]$ داریم $\langle w_i, x_i \rangle = 0$. در نتیجه تمام $x_1, \dots, x_d$ کلاسیفای (دسته بندی) نشده اند. همچنین $w = (1, \dots, 1)$ هر شرایط ذکر شده در سوال را برآورده می کند.