B 4.9

$$H = \{ h_{a,b,s} : a \leq b ,\ s \in \{-1,1\} \} \qquad h_{a,b,s} \begin{cases} s & x\in[a,b] \\ -s & \text{o.w} \end{cases}$$

y_1	y_2	y_3	a	b	s
1	1	1	$a<x_1$	$x_3<b$	1
1	1	-1	$a<x_1$	$x_2<b<x_3$	1
1	-1	1	$x_1<a<x_2$	$x_2<b<x_3$	-1
1	-1	-1	$a<x_1$	$x_1<b<x_2$	1
-1	1	1	$a<x_1$	$x_1<b<x_2$	-1
-1	1	-1	$x_1<a<x_2$	$x_2<b<x_3$	1
-1	-1	1	$a<x_1$	$x_2<b<x_3$	-1
-1	-1	-1	$a<x_1$ / $a<x_2$	$x_2<b<x_3$ / $x_3<b$	-1

حداکثر 3 تایی را shatter می‌کند.

$\longrightarrow$ shatter ۳ تایی ✓ $\longrightarrow$

۴ تایی باشد:

$$x \in \{(x_1,x_2,x_3,x_4) : x_1<x_2<x_3<x_4\}$$

اگر برچسب ها مقادیر زیر باشند:

$$y_1=1,\ y_2=-1,\ y_3=1,\ y_4=-1$$

$$\Downarrow$$

$$\text{vcdim} = \underline{\underline{3}} \ ✓$$

T.17

الف) خیر، چرا که h_S وابسته به (x_i, y_i) ها وابسته بوده و بر حسب آن ها تعیین می شود بنابراین $l(h_S,(x_i,y_i))$ به ازای تمام x_i, y_i ها مستقل نیست

ب) بله، چرا که تمام h_S ها ثابت هستند و هیچ کدام از $(x_i,y_i)\in T$ ها ناشی باشد بنابراین θ ها تابع x_i, y_i مورد نظر می باشد و از نقاط دیگر نمونه مستقل هستند شکل می گیرد.

ج) با توجه به بالا

$$P\left(|L_S(h_S)-L_D(h_S)| \leq \epsilon\right) \geq 1-\delta \Rightarrow m \geq m_H(\epsilon,\delta) = \frac{\log \frac{2|H|}{\delta}}{2\epsilon^2}$$

$$\rightarrow m_S = \frac{\log\left(\frac{2K}{\delta}\right)}{2\epsilon_1^2} \rightarrow \epsilon_1 = \sqrt{\frac{\log \frac{2K}{\delta}}{2m_S}} \ ✓$$

منظور از کمترین ϵ، کمترین مقدار اختلاف بین True risk

د) با توجه به اینکه θ ها از یکدیگر مستقل اند، با توجه به Hoeffding inequality

$$P\{|L_D(h_S)-L_T(h_S)| \leq 2e^{-2m_T\epsilon_2^2} = \delta$$

$$\rightarrow \epsilon_2 = \sqrt{\frac{\log(2/\delta)}{2m_T}} \ ✓$$

هـ)

$$m = m_S = m_T \rightarrow \epsilon_1 = \sqrt{\frac{\log(2/\delta)+\log(K)}{2m}} \ ,\ \epsilon_2 = \sqrt{\frac{\log(2/\delta)}{2m}}$$

$$\rightarrow \epsilon_2^2 + \frac{\log K}{2m} = \epsilon_1^2 \Rightarrow \epsilon_1 > \epsilon_2$$

و) ϵ_2 به ازای یک m, δ کمتر از ϵ_1 می باشد بنابراین با استفاده از مجموعه T کران بهتری برای الگوریتم با استفاده از $L_T(h_S)$ حاصل می شود

B ۱۵.۲

$$\rho = \max_i \|x_i\| \quad , \quad B = \{\, \|w\| \;,\; y_i \langle w, x_i \rangle \geq 1 \,\} = \frac{1}{\gamma}$$

$$w^{(T)} \rightarrow \text{خروجی} \Rightarrow \forall i \;,\; y_i (w^{(T)}, x_i) > 0$$

$\gamma \rightarrow$ margin $\Rightarrow$ فاصله نزدیکترین x به hp برابر γ است.

$$\Rightarrow \gamma = \min_i \frac{|\langle w^{(T)}, x_i \rangle|}{\|w^{T}\|} = \min_i \frac{y_i \langle w^{(T)}, x_i \rangle}{\|w^{T}\|} = \min_{i \in \rho} y_i \|x_i\| \cos\theta_i \quad *$$

$$B = \min \{\, \|w\| \;,\; \forall i \quad y_i \langle w, x_i \rangle \geq 1 \,\} \rightarrow \forall i \quad y_i \langle w^*, x_i \rangle \geq 1$$

$$w^* = \frac{B}{\|w^T\|} w^T \Rightarrow \|w^*\| = B \qquad \forall i \quad \frac{y_i \langle w^*, x_i \rangle}{B} = \frac{y_i \|x_i\| \cos\theta_i}{B} \quad **$$

$$*, ** \quad \gamma = \frac{1}{B} \rightarrow B = \frac{1}{\gamma} \Rightarrow T \leq (RB)^2 = \left(\frac{\rho}{\gamma}\right)^2 \checkmark$$

T.۸

الف) با توجه به فرض realizability و با توجه به اینکه خروجی کوچکترین مستطیل دربرگیرنده نقطه‌های مثبت است و اینکه empirical loss مثل مانند درست تشخیص داده شده است، الگوریتم ERM است.

ب)

$$R^* = R(a_1^*, b_1^*, a_2^*, b_2^*)$$

$$R(S) \subset R^*$$

با توجه به محدود بودن S

(R_S خروجی تنها با توجه به S)

مانند hint R_i را تعریف می‌کنیم، برای بدست آوردن خطا ۴ مجموعه نقاط زیر را تعریف می‌کنیم

$$F_i = \{ S|_x \;,\; S|_x \cap R_i = \emptyset \} \longrightarrow \text{مجموعه نقاط دیده نشده در } S \longrightarrow D^m(F_i) = (1 - \varepsilon/4)^m \leq e^{-m\varepsilon/4}$$

(نسبت به R_i هایی که بزرگتر R^* هستند)

(اشتراک گرفتن درون R_i نیست)

(با توجه به R_i ها، احتمال نقطه $\varepsilon/4$ است)

$$D^m (S : L(A(S)) > \varepsilon) \leq D^m(\cup F_i) \leq \sum_{i=1}^{4} D^m(F_i) \leq 4 e^{-m\varepsilon/4} \rightarrow m \leq \frac{4 \log(4/\delta)}{\varepsilon}$$

T11 $h(x) = w^T x$, $\ell(h(x), y) = \|h(x) - y\|$

$$\min f(w) = L_S(w) = \frac{1}{m} \sum \ell(h(x_i), y_i) = \frac{1}{m} \sum |w^T x_i - y_i|$$

الف) مسئله بهینه‌سازی خطی نیست چرا که توابع $|y_i - w^T x_i|$ برای داده‌های x_i, y_i (ثابت شده) خطی نیست (قدر مطلق خطی نیست)

ب) مسئله محدب و اکید نیست چرا که توابع $|y_i - w^T x_i|$ محدب بوده (شکل زیر) و مجموع توابع محدب، محدب باشد.

$\forall x_1, x_2 \; \forall \alpha \in (0,1)$: $f(\alpha x_2 + (1-\alpha) x_1) \le \alpha f(x_1) + (1-\alpha) f(x_2)$

ج) خیر، هر کدام از توابع در $y_i = w^T x_i$ شکستگی دارند و مشتق ندارند

د) فرض کنید که جواب بهینه مسئله $c_0 = \min_{w \in R^d} |f(w)|$ در w^* ای باشد.

$\rightarrow \min |f(w)| = |f(w^*)| = c_0 = c(w^*) \le |f(w^*)| = c(w)$

$\Rightarrow c_0 = c(w^*) \le c(w)$ و $-c(w) \le f(w) \le c(w)$

$\Leftarrow$ برای مسئله $\min_{w \in R^d, \, c \in R} c$ به شرط $f(w) \le c$ و $-f(w) \le c$ ، پاسخ در w^* و $c^* = \min c$ باشد

$\Leftarrow$ مسئله $\min_{w \in R^d} |f(w)|$ معادل $\min c$ به شرط $c \ge f(w)$ و $c \ge -f(w)$ باشد

با توجه به این، مسئله $\min L_S(w)$ را بازنویسی کنیم:

for x_i, y_i , $f_i(w) = \frac{1}{m}(w^T x_i - y_i)$

$$\min_{w \in R^d} |f_i(w)| = \min \frac{1}{m} |w^T x_i - y_i| \equiv \min c_i \quad st: \; c_i \ge \frac{1}{m}(w^T x_i - y_i), \; c_i \ge -\frac{1}{m}(w^T x_i - y_i)$$

چون $|f_i(w)|$ ها در $\sum f_i(w)$ از یکدیگر مستقل هستند، کمینه کل جمله با کمینه هر جمله به صورت مجزا محاسبه می‌شود

c_i ها $\leftarrow \sum c_i$ ، ، ،

$$\min L_S(w) = \min \sum |f_i(w)| = \sum \min |f_i(w)|$$

$$\min f(w) = \mathbb{1} c \quad st \quad \begin{bmatrix} y \\ -y \end{bmatrix} \ge -\begin{bmatrix} c \\ c \end{bmatrix} + \frac{1}{m} \begin{bmatrix} w \\ -w \end{bmatrix}^T \begin{bmatrix} x \\ x \end{bmatrix} \checkmark$$

ب)

رابطه‌ی مشابه‌ای برای $m_H(\epsilon,\delta)$ است و آنکه $m \geq m_H(\epsilon,\delta)$ تا به ϵ, δ مطلوب برسیم.

چه قدر درست آمده است مورد نظر شما دیگر
نیاز به دو ترم مشابه به این ثابت

در تمرین قبل دیدیم که $d = \text{vcdim}(H) = \infty$ بین ما در به قضیه ۶.۸ ثابت می‌داریم:

$$C_1 \frac{\log(1/\delta)}{\epsilon} \leq m_H(\epsilon,\delta) \leq C_2 \frac{\log(\frac{1}{\epsilon})+\log(\frac{1}{\delta})}{\epsilon}$$

ما هم با دادای شخص تری به دست آوردیم.

B.13.1

If $m \geq m_H(\epsilon)$,

$$\mathop{E}_{S \sim D^m}[L_D(A(S))] \leq \min_{h \in H} L_D(h) + \epsilon$$

since $A(S) \in H \rightarrow \theta = L_D(A(S)) - \min_{h \in H} L_D(h) \geq 0$

$$P(\theta \geq E(\theta)/\delta) \leq \frac{E(\theta)}{E(\theta)/\delta} = \delta$$

$\rightarrow$ $\underset{1-\delta}{\text{احتمال حداقل}}$ $\rightarrow \theta \leq E(\theta)/\delta$

اگر $m \geq m_H(\epsilon\delta)$ $\rightarrow$ طبق (ا) $E(\theta) \leq \epsilon\delta \rightarrow \epsilon \leq \frac{\epsilon\delta}{\delta} \geq \frac{E(\theta)}{\delta} \geq \theta$ ✓ اثبات شد.

$K = \lceil \log_2(\frac{2}{\delta}) \rceil$ داده را به $k+1$ تکه تقسیم، که k تکه اول $m_H(\epsilon/2)$ نشان داده k تکه اول را با A، h_i خروجی A برای i باشد.

استقلال داریم که برای $\frac{1}{2}$ داریم: $L_D(h_i) - \min_{h \in H} L_D(h) > \frac{\epsilon}{2}$ و از آنجایی که این k تا مستقل از هم پس $L_D(h_i) - \min L_D(h) \leq \frac{\epsilon}{2}$ داریم

$\frac{\delta}{2}$ است پس با احتمال $1-\frac{\delta}{2}$ داریم که $\min_{h \in K} L_D(h_i) \leq \min L_D(h) + \frac{\epsilon}{2}$ است، حال با استفاده از ERM برای $\{h_1, \dots, h_K\}$

و اگر تعداد نمونه را برابر با $\frac{\log(\frac{2k}{\delta})}{\epsilon^2}$ باشد داریم که $L_D(\hat{h}) \leq \min_{i \in K} L_D(h_i) + \frac{\epsilon}{2}$ و پس complexity overall برابر است با:

$$m_H(\epsilon/2)\lceil \log_2(2/\delta) \rceil + 8\left\lceil \frac{\log(4/\delta) + \log(\lceil \log_2(2/\delta) \rceil)}{\epsilon^2} \right\rceil$$ ✓