فصل 6

آرین صابونچی

2) a)

فرض می کنیم $C \subseteq X$ طوری که $|C| = k+1$

$\rightarrow \nexists\, h \in H$ s.t $\forall x \in C$ $h(x)=1$

فرض می کنیم $C \subseteq X$ طوری که $|C| = |X| - k + 1$

$\rightarrow \nexists\, h \in H$ s.t $\forall x \in C$ $h(x)=0$

$VC\,dim(H) \leq \min\{k, |X|-k\}$ (I)

فرض می کنیم $C \subseteq X$ طوری که $|C| = m \leq \min\{k, |X|-k\}$

همچنین $(y_1, \dots, y_m) \in \{0,1\}^m$ و $s = \sum y_i$

$E \subseteq X \setminus C$ & $h \in H$ $h(x_i) = y_i$

$\forall x_i \in C$ $h(x) = I_E$ $\forall x \in X \setminus C$

$\Rightarrow C$ is shattered by H

$\Rightarrow VC\,dim(H) \geq \min\{k, |X|-k\}$ (II)

(I), (II) $\Rightarrow VC\,dim(H) = \min\{k, |X|-k\}$

(b(2

فرض می کنیم $C \subseteq X$ طوری که $|C| = k+1$

$\rightarrow \nexists\, h \in H$ s.t $\forall x \in C \ h(x) = 1$

$\Rightarrow VCdim(H) \leq k$

فرض می کنیم $C = \{x_1, \dots, x_m\} \subseteq X$ طوری که $|C| = m \leq k$

و همینطور $(y_1, \dots, y_m) \in \{0,1\}^m$

$\exists\, h \in H$ s.t $\forall x_i \in C \ h(x_i) = y_i$ & $h(x) = 0$ $(\forall x \in X \setminus C)$

$\rightarrow VCdim(H) \geq k$

$\Rightarrow VCdim(H) = k$

---

4) فرض می کنیم $X = \mathbb{R}^d$

همچنین $H = \{ I_{\|x\|_2 \leq r} : r \geq 0 \}$ & $d \geq 2$

if $x \neq (0, \dots, 0)$ then $\{x\}$ is shatterred

if $\|x_1\|_2 \leq \|x_2\|_2 \Rightarrow y_1 = 0$ & $y_2 = 1$ is'nt obtained by any $h \in H$

$\mathcal{H}_A = \{(0,0),(1,1)\} \Leftarrow A=\{e_1,e_2\}$ فرض می‌کنیم

$\Rightarrow \{B \subseteq A : \text{"}\mathcal{H}\text{" shatteres "}B\text{"}\} = \{\emptyset, \{e_1\}, \{e_2\}\}$

$\Rightarrow \sum_{i=0}^{d} \binom{|A|}{i} = 3$

Let $\mathcal{H}$ be the class of axis-aligned rectengle in $\mathbb{R}^2$.

می‌دانیم که VC-dim $\mathcal{H}$ ، 4 است.

Let $A\{x_1, x_2, x_3\}$ where $x_1 = (0,0)$, $x_2 = (1,0)$, $x_3 = (2,0)$

all the lablings except $(1,0,1)$ are obtained

thus $|\mathcal{H}_A| = 7$, $|\{B \subseteq A : \text{"}\mathcal{H}\text{" shatterres "}B\text{"}\}| = 7$

and $\sum \binom{|A|}{i} = 8$

let $d \geq 3$ and consider class $\mathcal{H} = \{\text{sign}\langle w, x\rangle : w \in \mathbb{R}^d\}$

of homogereous half spaces we will prove in a theorem in CH9 that vcdim of this class is $\underline{d}$.

$\text{vcdim}(\mathcal{H}) \geq 3$

$\{e_1, e_2, e_3\}$ is shatterred.

let $A = \{x_1, x_2, x_3\}$ where $x_1 = e_1$, $x_2 = e_2$ & $x_3 = (1,1,0)$ &

all the lablings except $(1,1,-1)$ and $(-1,-1,1)$

are obtained. it follows that $|H_A| = 6$.

$|\{B \subseteq A : \text{"}H\text{" shatterres "}B\text{"}\}| = 7$

and $\sum \binom{|A|}{i} = 8$

let $d = 1$ and conside $H = \{I_{[x, t]} \; t \in \mathbb{R}\}$

every singetion is shatterred by H

and every set of size at least 2 is'nt

shatterred by H. choose any finite set $A \subseteq \mathbb{R}$

then each of the three terms in sauser's inequality

equals $|A| + 1$.

(a (6

$\text{vcdim}(H^d_{con}) \leq \log |H^d_{con}| \leq 3 \log d$ (b

let $I \subseteq [d]$ be a subset of indices (c

we will show that the labeling in which exactly the elements $\{e_j\}_{j \in J}$ are positive is obtained if $I = [d]$ picks all positive hypothesis h_{empty}

if $J = \emptyset$, pick the all negative hypothesis $x_1 \wedge \bar{x}_1$. Assume now that $\emptyset \subset J \subset [d]$

let h be hypothesis which corresponds to the boolean Conjumention $\bigwedge_{j \in J} x_j$. then $h(e_j) = 1$ if $j \in J$, and $h(e_j) = 0$ O.w.

So $C = \{e_j\}_{j=1}^d$ is shatterered by H^d_{con} thus $H^d_{con} \geq d$.

(d 6

Assume by contradiction that there exists a set $C=\{c_1,\dots,c_{d+1}\}$ for which $|H_C|=2^{d+1}$ define $h_1 \dots h_{d+1}$ and $l_1 \dots l_{d+1}$ as in the hint. by the pigen hole principle, among $l_1,\dots l_{d+1}$ at least one variable occus twice. Assume that l_1 and l_2 correspond to the some variable Assume first that $l_1=l_2$. then l_1 is true on c_1 since l_2 is true on c_1. Howerer this contradicts our assumptions. Assume now that $l_1 \neq l_2$ in this case $h_1(c_3)$ is negative, since l_2 is positive on c_3. This again Contradicts our assumptions.

(e

First we abserve that $|H'| = 2^d + 1$. thus
$vc\,dim(H') \leq \log|H| = d$

Let $C=\{(1,\dots,1)-e_j\}_{j=1}^{d} = \{(0,1,\dots,1)\},\dots(1,\dots,1,0)\}$

let $J \subseteq [d]$ be a subset of indices we will show that the labeling in which exactly the elements $\{(1,\dots,1)-e_j\}_{j\in J}$ are negative is obtained by the boolean conjunction $\bigwedge_{j\in J} x_j$. Finally if $J=\emptyset$, pick all positive hypothesis $h_\emptyset$. So C shattered.

let $C=\{1,2,3\}$ (9

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

$\Rightarrow \mathrm{VCdim}(H) \geq 3$

let $C=[x_1,x_2,x_3,x_4\}$ and assume that $x_1<x_2<x_3<x_4$ then labeling $y_1=y_3=-1$, $y_2=y_4=1$ is'nt obtained by any hypothesis in H.

3 Thus $\Rightarrow \mathrm{VCdim}(H)\leq 3 \Longrightarrow \boxed{\mathrm{VCdim}(H)=3}$

(a (10

assume that $m<d$, since otlerwise the statement is meaningless. let c be a shatterred set of size d. we assume that $X = C$ (since we can always choose distributions which are concentrated on C).

H contains all the functions from C to $\{0,1\}$. accoarding to Section 5, for every algorithm there exists a distribution D. for which $\min_{h\in H} L_D(h) = 0$ but $E(L_D(A(S))) \geq \frac{k-1}{2k} = \frac{d-m}{2d}$.

(b

assume that $vc\dim(H) = \infty$. let A be a learning algorithm. we show that A fails to PAC learn H. choose $\varepsilon = 1/16$, $\delta = \frac{1}{14}$. for any $m \in N$, there exists a shatterred set of size $d = 2$. Applying the above, we obtain that there exists a distributeon D for which $\min_{h\in H} L_D(h) = 0$,

4

h

but $E(L_D(A(S))) \geq \frac{1}{4}$. Applying Lemma B1

that with probabilty at least $\frac{1}{7} > \delta$,

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

---

assume that $\forall i \in [r], vcdim(H_i) = d \geq 3$ (11

(a

let $H = \cup_{i=1}^{T} H_i$. let $k \in [d]$, such that $\tau_H(k) = 2^k$

we will show that $k \leq 4d \log(2d) + 2 \log r$.

$\Rightarrow \tau_H(k) \leq \sum \tau_{H_i}(k)$

since $d \geq 3$, by applying sauser's lemma on each of the terms τ_{H_i}, we obtain $\tau_H(k) < r m^d$

it follows that $k < d \log m + \log r$

$\Rightarrow k < 4d \log(2d) + 2 \log r$

(b

a direct application of the result above yields a weaker bound. we need to emply a more careful analyisis.

~~As~~ we ~~assumption~~ assume that $vcdim(H_1) = vcdim(H_2) = \cancel{2}d$

$\cancel{2}$

let $H_1 \cup H_2$. let k be a positive integer such that $k \geq 2d+2$. we show that $\tau_H(k) < 2^k$. by Sauser's lemma:

$$
\begin{aligned}
\tau_H(k) &\leq \tau_{H_1}(k) + \tau_{H_2}(k) \\
&\leq \sum \binom{k}{i} + \sum \binom{k}{i} \\
&= \sum \binom{k}{i} + \sum \binom{i}{k-i} \\
&= \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=k-d}^{k} \binom{k}{i} \\
&\leq \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=d+2}^{k} \binom{k}{i} \\
&\leq \sum \binom{k}{i} + \sum_{i=d+1}^{k} \binom{k}{i} = \sum \binom{k}{i} = 2^k.
\end{aligned}
$$

chapter 9

(1

Define avector of auxiliay variables $\xi = (\xi_1, \dots, \xi_m)$ following the hint, minimizing the emeprical risk is equivalent to minimizing the linear objective $\sum \xi_i$ under following constains:

$\forall i \in [m] \quad w^T x_i - \xi_i \leq y_i \; , \; -w^T x_i - \xi_i \leq -y_i .$

it is left to translate the above into matrix from. Let $A \in \mathbb{R}^{2m \times (m+d)}$ be the matrix

$A = [X - I_m ; -X - I_m]$, where $X_i = x_i$ for every $i \in [m]$ let $v \in \mathbb{R}^{d+m}$ be the vector of variables $(w_1, \dots, w_d, \xi_1, \dots, \xi_m)$ define $b \in \mathbb{R}^{2m}$ to be the vector $b = (y_1, \dots, y_m, y_1, \dots, y_m)^T$.

finnally let $c \in \mathbb{R}^{d+m}$ be the vector $c = (0_d, 1_m)$

it follows that the optimaziton problem of minimizing the empeial risk can be empressd as the following LP : min $c^T v$ s.t $Av \leq b$

~~following the min~~

---

following the hint, let $d=m$, and for every $i \in [m]$, let $x_i = e_i$. let us agree that $\text{sign}(0) = -1$. for $i=1,\dots,d$, let $y_i = 1$ be the label of x_i. Denote by $w^{(t)}$ the weight vector which is maintained by the preceptron. (3

A simple inductive argument shows that for every $i \in [d]$, $w_i = \sum_{j<i} e_j$.

it follows that for every $i \in [d]$, $\langle w(i), x_i \rangle = 0$

Hence all the $x_1,\dots,x_d$ are misclassifid then we obtain the vector $w=(1,\dots,1)$ which is consistenst with $(x_1,\dots,x_m)$. we also note that the vector $w^* = (1,\dots,1)$ satisfies the requirments listedin the question.

~~o~~ consider all positive examples of the form $(\alpha, \beta, 1)$, where $\alpha^2+\beta^2+1 \leq R^2$. (4

observe that $w^* = (0, 0, 1)$ satisfies $y \langle w^*, x \rangle \geq 1$ for all such (x, y), we show a sequence of $\mathbb{R}^2$ examples on which the preceptron makes $\mathbb{R}^2$ mistakes.

the idea of the construction is to start with the examples $(\alpha_1, 0, 1)$ where $\alpha_1 = \sqrt{R^2-1}$, on round t let the new example be such that the followings conditions hold (a) $\alpha^2+\beta^2+1 = R^2$

(b) $\langle w_t, (\alpha, \beta, 1) \rangle = 0$

~~Az~~ as long as we can satisfy both conditons, the Preceptron will ~~o~~ continue to err

observe that, by induction, $w^{(t-1)} = (a, b, t-1)$ for some scalors a, b. obsserve also that $\|w_{t-1}\|^2 = (t-1) R^2$.

that $a^2+b^2+(t-1)^2=(t-1)R^2$

let us rotate $w^{(t-1)}$ the z axis so that it is of the form $(a,0,t-1)$ and we have $a=\sqrt{(t-1)R^2-(t-1)^2}$ choose $\alpha=-\frac{t-1}{a}$

Then for every B: $\langle (a,0,t-1),(\alpha,B,1)\rangle=0$

we need to verify than $\alpha^2+1\leq R^2$, because if it is ~~then~~ True then we can choose $B=\sqrt{R^2-\alpha^2-1}$ $\Rightarrow \alpha^2+1=\frac{(t-1)^2}{\alpha^2}+1$

$$=\frac{(t-1)^2}{(t-1)R^2-(t-1)^2}+1=R^2\frac{1}{R^2-(t-1)}\leq R^2$$

where the last inequatity assumes $R^2\geq t$.