Machine Learning Exercises_Chapter 6

Mohammadjavad Abbaspour

The VC-Dimension

6.2. 1) We must show that
1. $VC\text{-}Dimension(H) \leq \min\{k, |X|-k\}$
2. $VC\dim(H) \geq \min\{k, |X|-k\}$

, then we prove that $VC\dim(H) = \min\{k, |X|-k\}$.

At first we show (1). Suppose k is the min of the two (without reducing the totality of the issue). choose $C = \{c_n\}_{n=1}^{n=k+1}$, cause of there are $k+1$ elements with lable 1 in C, we can't have $(1,\dots,1)$ labeling for C. Also if we assume $C \subseteq X$ is the set of size $|X|-k+1$, there is no $h \in H$ which satisfies $h(x)=0$ for all $x \in C$.

Hence we show (1).

It's left to show (2). Let $C = \{c_1, c_2, \dots, c_m\}$, $m = \min\{k, |X|-k\}$ and C be a subset of X with labeling $\{y_1, \dots, y_m\}$ produces by h $(h(c_i) = y_i \ \forall i \in 1,\dots,m)$. choose points in X/C to get 1 label. we know that there are at most k elements which get labeled 1. Therefore $VC\dim(H) \geq m$. Hence we show (2).

Then (1), (2) give the result that $VC\dim(H) = \min\{k, |X|-k\}$.

2) We claim that $VCdim(H) = k$. We must show
1. $VCdim(H) \geq k$
2. $VCdim(H) \leqslant k$

At first we show (1). Let $C = \{c_n\}_{n=1}^{n=k+1}$ and $Y = \{y_n\}_{n=1}^{n=k+1}$ be labeling for C.
we have $h(C_i) = 1$ if $y_i = 1$ and $h(c_i) = 0$ otherwise. Ther are at most k points with label 1.

(2): Let $C = \{c_1, \dots, c_k\} \subset X$. There are more than k points in C, Cause of it there is no $h \in H$ that assigns label 1 to any numbers of C.

Hence $VCdim(H) = k$.

6.4. Let $X=\mathbb{R}^2$. we demonstrate all the 4 combinations over $X\times\{0,1\}$. we know that $VCdim(H)=2$. ($|\{B\subseteq A : B \text{ shattered by } H\}| = b$, $|H_A|=a$, $\sum\binom{|A|}{i}=c$)

1) $(<,<)$: Let $C=\{c_1=(1,0), c_2=(2,0)\}$. then $a=|\{(1,1),(0,0)\}|=2$, $b=|\{\emptyset,\{c_1\},\{c_2\}\}| = 3$

$c=1+2+1=4$

2) $(<,=)$: Let $C=\{c_1=(1,0), c_2=(1,1), c_3=(1,-1)\}$, then:

$a=|\{\{1,1,1\},\{0,0,0\},\{1,0,0\},\{0,0,1\},\{0,1,1\},\{1,1,0\}\}|=6$

$b=|\{\emptyset,\{c_1\},\{c_2\},\{c_3\},\{c_1,c_2\},\{c_1,c_3\},\{c_2,c_3\}\}|=7$

$c=1+3+3=7$

3) $(=,<)$: Let $C=\{c_1=(0,0), c_2=(1,0)\}$ then: $\begin{cases} a=|\{(1,0),(1,1)\}|=2 \\ b=|\{\emptyset,\{c_2\}\}|=2 \\ c=1+2+1=4 \end{cases}$

4) $(=,=)$: Let $C=\{c_1=(1,0), c_2=(1,1)\}$ then: $\begin{cases} a=4 \text{ (C shatters by H)} \\ b=|\{\emptyset,\{c_1\},\{c_2\},\{c_1,c_2\}\}|=4 \\ c=1+2+1=4 \end{cases}$

6.6. (a): For each Variables $x_1, x_2, \dots, x_d$ $(d \geq 2)$, we assume that each $h \in H$ is determined by deciding whether x_i, $\bar{x}_i$ or it might missing from h.

Thus, $|H^d| = 3^d + 1$

(b): Let $VCdim(H) = \alpha \Rightarrow |H| \geq 2^\alpha \longrightarrow \alpha \leq \log_2 (|H|)$ then: $VCdim(|H|) \leq \log(3^d + 1) \leq d \log(3)$

(C): Let $C = \{e_i : i \leq d\}$ and $\{y_1, \dots, y_d\}$ are labels, now we show $VCdim(H) \geq d$.

$$\begin{cases} \text{if } \forall_{1 \leq i \leq d} \ y_i = 1 \rightarrow h(x) = 1 \\ \text{if } \forall_{1 \leq i \leq d} \ y_i = 0 \rightarrow h(x) = 0 \end{cases}$$

gives the correct labeling.

for other cases we have to sort them like: $y_1 = y_2 = \dots = y_m = 1$ $(m < d)$ and 0-label for others.

Let $h(x) = x_1 \wedge \dots \wedge x_m$. Hence H is shattering d elements: $VCdim(H) \geq d$.

(d): We must to show that $VC\dim(H) \leq d$.

(Proof of contradiction). Let $VC\dim(H) > d \rightarrow H$ shatters a set of $d+1$ elements.

Assume $C = \{c_1, \dots, c_{d+1}\}$ and $h_i(C_j) = 0$ for $i = j$ $(1 \leq i \leq d+1)$. we have l_i in h_i that is false for c_i and true for C_j; Let L_i in h_i and l_j in h_j:

$$\begin{cases} 1)\ l_i = l_j \rightarrow h_i(C_i) = 0 \wedge h_j(C_i) = 0 \\ 2)\ l_i \neq l_j \rightarrow h_i(C_k) = \neg h_j(C_k) \end{cases}$$

Thus $VC\dim(H) \leq d$.

(e) The size of this class is $2^d + 1$, therefore: $VC\dim(H) \leq \lfloor \log(2^d + 1) \rfloor = d$

then we must to show that a d points set is shattered by H:

Proof. Let $C = \{C_1, \dots, C_d\}$ and $C_K = (1,1,1,\dots,1,\underset{\text{k'th point is 0}}{0},1,\dots,1,1)$ and L is a set of negetive labels. we assume that $h(x) = x_{i_1} \wedge x_{i_2} \wedge x_{i_3} \wedge \dots \wedge x_{i_m}$ if $L = \emptyset$ then $h(x) = 1$: $\big(h(C_i) = 0 \leftrightarrow i \in L\big)$ hence, $\boxed{VC\dim(H^d) = d}$

Therefore: $VCdim(H) \leq \lfloor \log(2^d+1) \rfloor = d$

then we must to show that a d points set is shattered by H:

Proof: Let $C = \{c_1, \dots, c_d\}$ and $c_K = (1,1,1,\dots,1,0,1,\dots,1,1)$ and L is a set of negetive labels. we assume that $h(x) = x_{i_1} \wedge x_{i_2} \wedge x_{i_3} \wedge \dots \wedge x_{i_m}$ if $L = \emptyset$ then $h(x) = 1$: $\left(h(c_i) = 0 \leftrightarrow i \in L\right)$ hence, $\boxed{VCdim(H^d) = d}$

k'th point is 0

6.9. Let $C = \{0, 1, 2\}$, we have to show that C is shattered by H. The table below show this:

a	b	s	y_1	y_2	y_3
4	5	1	-1	-1	-1
$\frac{1}{2}$	$\frac{3}{2}$	1	-1	1	-1
0	2	1	1	1	1
1	2	1	-1	1	1
1	2	-1	1	-1	-1
0	1	1	1	1	1
$\frac{1}{2}$	$\frac{3}{2}$	-1	1	-1	1

(I)

Therefor $VCdim(H) \geq 3$. Let $C = \{c_1, c_2, c_3, c_4\}$ $(c_1 < c_2 < c_3 < c_4)$, then the labling $y_1 = -1$, $y_2 = 1$, $y_3 = -1$, $y_4 = 1$ can not be obtained by any $h \in H$.

There fore $VCdim(H) \leq 3$, Hence $(1),(2) \rightarrow VCdim(H) = 3$.

6.10.(1) Let C be a shattered set of size d and $m<d$. H_C contains all functions from C to $\{0,1\}$. By exercise 3:

$$\begin{cases} \exists f \in H_C : L_D(f)=0 \\ E_{S\sim C^m}[L_D(A(S))] \geq \frac{1}{2} - \frac{1}{2k} \quad \text{for } k \leq \frac{d}{m} \end{cases}$$

ther exists a D for which $\min_{h\in H} L_D(H)=0$ then:

$$E_{S\sim D^m}[L_D(A(S))] \geq \frac{1}{2} - \frac{m}{2d} = 0 + \frac{d-m}{2d}$$

$$E_{S\sim D^m}[L_D(A(S))] \geq \min_{h\in H} L_D(H) + \frac{d-m}{2d}$$

(2) we suppuse that $VC\dim(H)=\infty \rightarrow$ for any set with size of m there exists a shattered set of size $2m=d$, then we have:

$$E_{S\sim D^m}[L_D(A(S))] \geq \frac{2m-m}{4m} = \boxed{\frac{1}{4}}$$

then by using Markov's result:

$$P_{S\sim D^m}\left[L_D(A(S)) \geq \frac{1}{8}\right] \geq \frac{E[L_D(A(S))] - \frac{1}{8}}{1-\frac{1}{8}} = \frac{1}{7}$$

Hence if we choosing $\varepsilon \leq \frac{1}{8}$, $\delta \leq \frac{1}{7}$, This violates PAC-learning. Therefore H is not PAC-learnable.

6.11. (a) We assume that for each $i \in [r]$, $\text{vc dim}(H_i) = d \geqslant 3$.

Let $H = \bigcup_{i=1}^{r} H_i$ and $k \in [d] \rightarrow \tau_H(k) = 2^k$. By definition of growth functio

$$\tau_H(K) \leqslant \sum_{i=1}^{r} \tau_{H_i}(K).$$

By applying Sauer's Lemma on each of the τ_{H_i} we obtain:

$$\boxed{\tau_H(k) < r m^d} \rightarrow k < d \log m + \log r \xrightarrow[\text{Lemma}]{} k < 4d \log(2d) + 2 \log r$$

We have that $\tau_{\cup H}(k) = 2^k$ then: $2^k < rk^d \Rightarrow k < d\log k + \log r \Rightarrow$

$$\xrightarrow[\text{A.2. Lemma}]{} k < d\log k + \log r \Rightarrow k < 4d\log(2d) + 2\log r$$

then we can say that a set with k elemens was shattered by UH. Hence:

$$\text{VC dim}\left(\bigcup_{i=1}^{r} H_i\right) \leq k < 4d\log(2d) + 2\log r.$$

(2) proof that: $VCdim(H_1 \cup H_2) \leq 2d+1$

we must to show that $H_1 \cup H_2$ is not shattered by a set of size $2d+2$

Hence we show $\tau_{H_1 \cup H_2}(k) < 2^k$ for $k \geq 2d+2$.

By using Sauer's Lemma:

$$\tau_{H_1 \cup H_2}(k) \leq \tau_{H_1}(k) + \tau_{H_2}(k) \leq \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=0}^{d} \binom{k}{i} = \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=0}^{d} \binom{k}{k-i} = \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=k-d}^{k} \binom{k}{i}$$

$(k \geq 2d+2)$

$$\leq \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=d+2}^{d} \binom{k}{i} < \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=d+1}^{k} \binom{k}{i} = \sum_{i=0}^{k} \binom{k}{i} =$$

$$= 2^k.$$

$$\boxed{\text{Hence, } \tau_{H_1 \cup H_2}(k) \leq 2^k}$$