Assignment 5

Applied Machine Learning

Fatemeh Saberi Khomami | 400422114 | fatima.saberi@gmail.com

Question 6.2

Given some finite domain set, X, and a number $k \leq |X|$, figure out the VC-dimension of each of the following classes (and prove your claims):

6.2.1

$H_{=k}^{X} = \{h \in \{0,1\}^{X} : |x : h(x) = 1| = k\}$: that is, the set of all functions that assign the value 1 to exactly k elements of X.

Answer: Let C be the subset of X which will assist us to determine the $VCdim$ of H. For the $C \subseteq X$ with at least k members, H will shatter C, because for all the 2^k subsets of C, there will exist a $h \in H$ which assigns 1 to the t members of the corresponding subset and $k - t$ members of $X - C$. Therefore, $VCdim(H) \geq k$. Regarding that H is the class of all functions that assign the value 1 to exactly k elements of X, for $C \in X$ which has $k+1$ members, H will not shatter C, because for the subset $\{c_1, c_2, \ldots, c_{k+1}\}$ it fails to assign 1 to all of them; hence, it fails to shatter C with $k+1$ members.
We can conclude that $VCdim(H_{=k}) = k$

6.2.2

$H_{at-most-k} = \{h \in \{0,1\}^X \ : \ |x \ : \ h(x) = 1| \leq k \ \ or \ |\{x \ : \ h(x) = 0\}| \leq k\}.$

Answer: Regarding that H is the class of all functions that assign 1 to at most k elements of X, for $C \in X$ which has $k+1$ members, H will not shatter C, because for the subset $\{c_1, c_2, \ldots, c_{k+1}\}$ it fails to assign 1 to all of them; hence, it fails to shatter C with $k+1$ members. $(VCdim(H) \neq k+1)$
For $C \subseteq X$ with k members, H will shatter C, because for all the 2^k subsets of C, there will exist a $h \in H$ which assigns 1 to at most k elements of X. So, all the subsets of C like: $\{\emptyset\}, \{c_1\}, \ldots, \{c_1, c_2\}, \ldots, \{c_1, c_2, c_3\}, \ldots, \{c_1, c_2, \ldots, c_k\}$ will have the chance to be addressed by the learning rule h. Therefore, $VCdim(H) \geq k$, and since $VCdim(H) \neq k+1$, we can conclude that it's:
$VCdim(H_{at-most-k}) = k$

Question 6.4

We proved Sauer's lemma by proving that for every class H of finite VC-dimension d, and every subset A of the domain,

$$|H_A| \leq |\{B \subseteq A \ : \ H \ shatters \ B\}| \leq \sum_{i=0}^{d} \binom{|A|}{i}.$$

Show that there are cases in which the previous two inequalities are strict (namely, the $\leq$ can be replaced by $<$) and cases in which they can be replaced by equalities. Demonstrate all four combinations of $=$ and $<$.

Answer:

- for the ($=$ & $<$) case we can give an example with $H_{interval}$ in which the $VCdim(H_{interval}) = 2$, and if we assume $A = \{c_1, c_2, c_3\}$, $|H_A|$ would be 4 because it can not produce a function for the following subsets: $\{c_2\}, \{c_3\}, \{c_2, c_3\}, \{c_1, c_3\}$, but $H_{interval}$ will shatter $B \subseteq A$ which is $\{c_1, c_2\}$ and $|H_B| = 2^{|B|} = 4$. Moreover, $\sum_{i=0}^{2} \binom{3}{i} = 7$, so $4 = 4 < 7$.

Question 6.9

Let H be the class of signed intervals, that is, $H = \{h_{a,b,s} : a \leq b, s \in \{-1, 1\}\}$ where

$$h_{a,b,c}(x) = \begin{cases} s & if x \in [a,b] \\ -s & if x \notin [a,b] \end{cases}$$

Calculate $VCdim(H)$.

Answer: First, we will show that $VCdim(H) \geq 2$. Assume $C = \{c_1, c_2,\ s.t\ c_1 < c_2\}$. For the 4 possible subsets of C, we'll have:

- **(- , -) :** $c_1 < c_2 < a$
- **(+ , +) :** $a < c_1 < c_2 < b$
- **(+ , -) :** $a < c_1 < b < c_2$
- **(- , +) :** $c_1 < a < c_2 < b$

Second, we'll investigate whether $VCdim(H) \geq 3$. Assume $C = \{c_1, c_2, c_3\ s.t\ c_1 < c_2 < c_3\}$. For **(+ , - , +)** , no valid intervals exist in H. So, $VCdim(H) \neq 3$. Therefore, $VCdim(H) = 2$.

Question 6.10

Let H be a class of functions from X to $\{0, 1\}$.

6.10.1

Prove that if $VCdim(H) \geq d$, for any d, then for some probability distribution D over $X \times \{0, 1\}$, for every sample size, m,

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

Hint: Use Exercise 5.3 in Chapter 5.

Answer: Based on what has been stated in exercise 5.3 in chapter 5, for a class of functions from X to $\{0, 1\}$,

- There exists a function $f : X \rightarrow \{0, 1\}\ with\ L_D(f) = 0$
- for $|X| \geq km$, $\mathbb{E}_{S \sim D^m}[L_D(A(S))] \geq \frac{1}{2} - \frac{1}{2k}$

Here, $|X| = d$, so, $d \geq km \Rightarrow k \leq \frac{d}{m}$. We know that H will shatter classes with d members in $|X|$, and since all of these classes are PAC learnable, we can assume $k = \frac{d}{m}$

$$\mathbb{E}_{S\sim D^m}[L_D(A(S))] \geq \frac{1}{2} - \frac{1}{2k}$$
$$\mathbb{E}_{S\sim D^m}[L_D(A(S))] \geq \frac{1}{2} - \frac{m}{2d}$$
$$\mathbb{E}_{S\sim D^m}[L_D(A(S))] \geq \frac{d-m}{2d}$$

Due to the No-Free-Lunch theorem, there exist $h \in H$ in which $L_D(h) = 0$, So the ERM learning strategy would return 0 for $min_{h\in H} L_D(h)$. Therefore, we can conclude that:

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

6.10.2

Prove that for every H that is PAC learnable, $VCdim(H) < \infty$. (Note that this is the implication $3 \to 6$ in Theorem 6.7.)

Answer: We only need to prove that for $VCdim(H) = \infty$, H is not PAC learnable. According to no free lunch theorem, we know that If someone can explain every phenomenon, his explanations are worthless. Hence, when H can shatter any class, it is not possible to learn each and every class with that H with low error. It would be some classes that $h \in H$ fails to address without low error-rate. Therefore, if $VCdim(H)$ is infinite, H is not PAC learnable.
We can conclude that if $VCdim(H)$ is finite, H is PAC learnable.