Subject:

Date: No:

The VC-dimension Exercises

1. Show the following monotonicity property of VC-dimension: For every two hypothesis classes if $\mathcal{H}' \subseteq \mathcal{H}$ then $VCdim(\mathcal{H}') \leq VCdim(\mathcal{H})$

Let $\mathcal{H}' \subseteq \mathcal{H}$ be two hypothesis classes for binary classification. Since $\mathcal{H}' \subseteq \mathcal{H}$, then for every $A = \{a_1, -, a_m\} \subseteq \mathcal{X}$ we have $\mathcal{H}'_A \subseteq \mathcal{H}_A$. In particular, if A is shattered by $\mathcal{H}'$, then A is shattered by $\mathcal{H}$ as well. Thus $VCdim(\mathcal{H}') \leq VCdim(\mathcal{H})$.

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

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

We claim that $VCdim(\mathcal{H}_{=k}) = \min\{k, |\mathcal{X}| - k\}$. First, we show that $VCdim(\mathcal{H}_{=k}) \leq k$. Let $C \subseteq \mathcal{X}$ be a set of size $k+1$. Then, there doesn't exist $h \in \mathcal{H}_{=k}$ which satisfies $h(x) = 1$ for all $x \in C$. Analogously, if $C \subseteq \mathcal{X}$ is a set of size $|\mathcal{X}| - k + 1$, there is no $h \in \mathcal{H}_{=k}$ which satisfies $h(x) = 0$ for all $x \in C$. Hence, $VCdim(\mathcal{H}_{=k}) \leq \min\{k, |\mathcal{X}| - k\}$.

Let $C = \{x_1, -, x_m\} \subseteq \mathcal{X}$ be a set with of size $m \leq \min\{k, |\mathcal{X}| - k\}$. Let $(y_1, -, y_m) \in \{0,1\}^m$ be a vector of labels. Denote $\sum_{i=1}^{m} y_i$ by s. Pick an arbitrary subset $E \subseteq \mathcal{X} \setminus C$ of $k-s$ elements, and let $h \in \mathcal{H}_{=k}$ be the hypothesis which satisfies $h(x_i) = y_i$ for every $x_i \in C$, and $h(x) = 1_{[E]}$ for every $x \in \mathcal{X} \setminus C$. We conclude that C is shattered by $\mathcal{H}_{=k}$. It follows that $VCdim(\mathcal{H}_{=k}) \geq \min\{k, |\mathcal{X}| - k\}$.

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

We claim that $VCdim(\mathcal{H}_{\leq k}) = k$. First, we show that $VCdim(\mathcal{H}^{\mathcal{X}}_{\leq k}) \leq k$.

PONIX notebook

Subject:

Date: No:

Let $C \subseteq \mathcal{X}$ be a set of size $k+1$. Then, there doesn't exist $h \in \mathcal{H}_{\leq k}$ which satisfies $h(x)=1$ for all $x \in C$. Let $C = \{x_1, \ldots, x_m\} \subseteq \mathcal{X}$ be a set with of size $m \leq k$. Let $(y_1, \ldots, y_m) \in \{0,1\}^m$ be a vector of labels. This labeling is obtained by some hypothesis $h \in \mathcal{H}_{\leq k}$ which satisfies $h(x_i) = y_i$ for every $x_i \in C$, and $h(x)=0$ for every $x \in \mathcal{X} \setminus C$. We conclude that C is shattered by $\mathcal{H}_{\leq k}$. It follows that $VCdim(\mathcal{H}_{\leq k}) \geq k$.

4. We proved Sauer's lemma by providing that for every class $\mathcal{H}$ of finite VC-dimension d, and every subset A of the domain, $|\mathcal{H}_A| \leq |\{B \subseteq A : \mathcal{H} \text{ shatters } B\}| \leq \sum_{i=0}^{d} \binom{|A|}{i}$.

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 combination of $=$ and $<$.

Let $\mathcal{X} = \mathbb{R}^d$. We will demonstrate all the 4 combinations using hypothesis classes defined over $\mathcal{X} \times \{0,1\}$. Remember that the empty set is always considered to be shattered.

- $(<, =)$, Let $d \geq 2$ and consider the class $\mathcal{H} = \{1_{[\|x\|_2 \leq r]} : r \geq 0\}$ of concentric balls. The VC-dimension of this class is 1. To see this, we first observe that if $x \neq (0, \ldots, 0)$, then $\{x\}$ is shattered. Second, if $\|x_1\|_2 \leq \|x_2\|_2$, then the labeling $y_1 = 0$ and $y_2 = 1$ is not obtained by any hypothesis in $\mathcal{H}$. Let $A = \{e_1, e_2\}$, where e_1, e_2 are the first two elements of the standard basis of $\mathbb{R}^d$. Then, $\mathcal{H}_A = \{(0,0), (1,1)\}$, $\{B \subseteq A : \mathcal{H} \text{ shatters } B\} = \{\emptyset, \{e_1\}, \{e_2\}\}$, and $\sum_{i=0}^{d} \binom{|A|}{i} = 3$.
- $(=, <)$, Let $\mathcal{H}$ be the class of axis-aligned rectangles

PONIX notebook

Subject:

Date: No:

in $\mathbb{R}^2$. We have seen that the VC-dimension of $\mathcal{H}$ is 4. Let $A = \{x_1, x_2, x_3\}$, where $x_1 = (0,0)$, $x_2 = (1,0)$, $x_3 = (2,0)$. All the labelings except $(1,0,1)$ are obtained. Thus $|\mathcal{H}_A| = 7$, $|\{B \subseteq A : \mathcal{H} \text{ shatters } B\}| = 7$, and $\sum_{i=0}^{d} \binom{|A|}{i} = 8$.

- $(<, <)$, Let $d \geq 3$ and consider the class $\mathcal{H} = \{\text{sign}\langle w, x\rangle : w \in \mathbb{R}^2\}$ of homogenous halfspaces. Let $A = \{x_1, x_2, x_3\}$, where $x_1 = e_1$, $x_2 = e_2$ and $x_3 = (1,1,0,\ldots,0)$. Note that all the labelings except $(1,1,-1)$ and $(-1,-1,1)$ are obtained. It follows that $|\mathcal{H}_A| = 6$, $|\{B \subseteq A : \mathcal{H} \text{ shatters } B\}| = 7$ and $\sum_{i=0}^{d} \binom{|A|}{i} = 8$.
- $(=, =)$, Let $d = 1$, and consider the class $\mathcal{H} = \{\mathbb{1}_{[x \geq t]} : t \in \mathbb{R}\}$ of thresholds on the line. We have seen that every singleton is shattered by $\mathcal{H}$, and that every set of size at least 2 is not shattered by $\mathcal{H}$. Choose any finite set $A \subseteq \mathbb{R}$. Then each of the three terms in Sauer's inequality equals $|A| + 1$.

6. VC-dimension of Boolean conjuctions: Let $\mathcal{H}^d_{con}$ be the class of Boolean conjuctions over the variables $x_1, \ldots, x_d$ $(d \geq 2)$. We already know that this class is finite and thus (agnostic) PAC learnable. In this question we calculate $VCdim(\mathcal{H}^d_{con})$.

1. Show that $|\mathcal{H}^d_{con}| \leq 3^d + 1$.
2. Conclude that $VCdim(\mathcal{H}) \leq d \log 3$.
3. Show that $\mathcal{H}^d_{con}$ shatters the set of unit vectors $\{e_i : i \leq d\}$.
4. (**) Show that $VCdim(\mathcal{H}^d_{con}) \leq d$.

Hint: Assume by contradiction that there exists a set $C = \{c_1, \ldots, c_{d+1}\}$ that is shattered by $\mathcal{H}^d_{con}$. Let $h_1, \ldots, h_{d+1}$ be hypothesis in $\mathcal{H}^d_{con}$ that satisfy

PONIX notebook

Subject:

Date: No:

$$\forall i,j \in [d+1],\ h_i(c_j) = \begin{cases} 0 & i=j \\ 1 & \text{otherwise} \end{cases}$$

For each $i \in [d+1]$, h_i (or more accurately, the conjuction that corresponds to h_i) contains some literal l_i which is false on c_i and true on c_j for each $j \neq i$. Use the Pigeonhole principle to show that there must be a pair $i < j \leq d+1$ such that l_i and l_j use the same x_k and use the fact to derive a contradiction to the requirements from the conjuctions h_i, h_j.

5. Consider the class $\mathcal{H}^d_{mcon}$ of monotone Boolean conjunctions over $\{0,1\}^d$. Monotonically here means that the conjuctions do not contain negations. As in $\mathcal{H}^d_{con}$ the empty conjuction is interpreted as the all-positive hypothesis. We augment $\mathcal{H}^d_{mcon}$ with the all-negative hypothesis h^-. Show that $VCdim(\mathcal{H}^d_{mcon}) = d$.

1. Each hypothesis, besides the all-negative hypothesis, is determined by deciding for each variable x_i, whether x_i, $\bar{x}_i$ or none of which appear in the corresponding conjuction. Thus, $|\mathcal{H}^d_{con}| = 3^d + 1$.

2. $VCdim(\mathcal{H}^d_{con}) \leq \lfloor \log(|\mathcal{H}^d_{con}|) \rfloor \leq 3\log d$.

3. We prove that $\mathcal{H}^d_{con} \geq d$ by showing that the set $C = \{e_j\}_{j=1}^d$ is shattered by $\mathcal{H}^d_{con}$. Let $J \subseteq [d]$ be a subset of indices. We will show that the labeling in which exactly the elements $\{e_j\}_{j \in J}$ are positive is obtained. If $|J| = [d]$, pick the all-positive hypothesis h_{empty}. If $J = \phi$, pick the all-negative hypothesis $x_1 \wedge \bar{x}_1$. Assume now that $\emptyset \subsetneq J \subsetneq [d]$. Let h be the hypothesis which corresponds to the boolean

Subject:

Date: No:

conjuction $\wedge_{j \in J} x_j$. Then, $h(e_j) = 1$ if $j \in J$, and $h(e_j) = 0$ otherwise.

4. Assume by contradiction that there exists a set $C = \{c_1, \ldots, c_{d+1}\}$ for which $|\mathcal{H}_C| = 2^{d+1}$. Define, $h_1, \ldots, h_{d+1}$ and $l_1, \ldots, l_{d+1}$ as in the hint. By the Pigeonhole principle, among $l_1, \ldots, l_{d+1}$, at least one variable occurs twice. Assume w.l.o.g. that l_1 and l_2 correspond to the same variable. Assume first that $l_1 = l_2$. Then l_1 is true on c_1 since l_2 is true on c_1. However, 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.

5. First, we observe that $|\mathcal{H}'| = 2^d + 1$. Thus,

$$\mathrm{VCdim}(\mathcal{H}') \leq \lfloor \log(|\mathcal{H}|) \rfloor = d$$

We will complete the proof by exhibiting a shattered set with size d. Let $C = \{(1, 1, \ldots, 1) - e_j\}_{j=1}^{d} =$ $\{(0, 1, \ldots, 1), \ldots, (1, 1, \ldots, 1, 0)\}$

Let $J \subseteq [d]$ be a subset of indices. We will show that the labeling in which exactly the elements $\{(1, 1, \ldots, 1) - e_j\}_{j \in J}$ are negative is obtained. Assume for the moment that $J \neq \emptyset$. Then the labeling is obtained by the boolean conjuction $\wedge_{j \in J} x_j$. Finally, if $J = \emptyset$, pick the all-positive hypothesis $h_\emptyset$.

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

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

calculate $\mathrm{VCdim}(\mathcal{H})$.

PONIX notebook

Subject:

Date: No:

We prove that $VCdim(\mathcal{H}) = 3$. Choose $C = \{1,2,3\}$. The following table shows that C is shattered by $\mathcal{H}$.

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

We conclude that $VCdim(\mathcal{H}) \geq 3$. Let $C = \{x_1, x_2, x_3, x_4\}$ and assume w.l.o.g. that $x_1 < x_2 < x_3 < x_4$. Then the labeling $y_1 = y_3 = -1$, $y_2 = y_4 = 1$ is not obtained by any hypothesis in $\mathcal{H}$. Thus, $VCdim(\mathcal{H}) \leq 3$.

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

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

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

2. Prove that for every $\mathcal{H}$ that is PAC learnable, $VCdim(\mathcal{H}) < \infty$.

1. We may assume that $m < d$, since otherwise the statement is meaningless. Let C be a shattered set of size d. We may assume w.l.o.g. that $\mathcal{X} = C$ (since we can always choose distributions which are concentrated on C). Note that $\mathcal{H}$ contains all the functions from C to $\{0,1\}$. For every algorithm, there exists a distribution $\mathcal{D}$, for which $\min_{h \in \mathcal{H}} L_{\mathcal{D}}(h) = 0$, but $E[L_{\mathcal{D}}(A(S))] \geq \frac{k-1}{2k} = \frac{d-m}{2d}$

$$\Downarrow$$

$$k = \frac{d}{m}$$

PONIX notebook

Subject:

Date: No:

2. Assume that $VCdim(\mathcal{H}) = \infty$. Let A be learning algorithm. We show that A fails to PAC learn $\mathcal{H}$. Choose $\epsilon = \frac{1}{16}$, $\delta = \frac{1}{14}$.
For any $m \in \mathbb{N}$, there exists a shattered set of size of $d = 2m$.
Applying the above, we obtain that there exists a distribution D for which $\min_{h \in H} L_D(h) = 0$, but $E[L_D(A(S))] \geq 1/4$.
With probability at least $1/7 > \delta$, $L_D(A(S)) - \min_{h \in H} L_D(h) = L_D(A(S)) \geq 1/8 > \epsilon$.

11. VC of union: Let $\mathcal{H}_1, \ldots, \mathcal{H}_r$ be hypothesis classes over some fixed domain set $\mathcal{X}$. Let $d = \max_i VCdim(\mathcal{H}_i)$ and assume for simplicity that $d \geq 3$.

1. Prove that $VCdim(\bigcup_{i=1}^{r} \mathcal{H}_i) \leq 4d\log(2d) + 2\log(r)$

Hint: Take a set of k examples and assume they are shattered by the union class. Therefore, the union class can produce all 2^k possible labelings on these examples. Use Sauer's lemma to show that the union class cannot produce more than rk^d labelings. Therefore $2^k < rd^k$.

We may assume w.l.o.g. that for each $i \in [r]$, $VCdim(\mathcal{H}_i) = d \geq$

3. Let $\mathcal{H} = \bigcup_{i=1}^{r} \mathcal{H}_i$. Let $k \in [d]$, such that $\tau_{\mathcal{H}}(k) = 2^k$. We will show that $k \leq 4\log(2d) + 2\log r$.

By definition of the growth function, we have

$$\tau_{\mathcal{H}}(k) \leq \sum_{i=1}^{r} \tau_{\mathcal{H}_i}(k)$$

Since $d \geq 3$, by applying Sauer's lemma on each of the terms $\tau_{\mathcal{H}_i}$, we obtain $\tau_{\mathcal{H}}(k) < rm^d$.

It follows that $k < d\log m + \log r$.

Lemma: Let $a \geq 1$ and $b > 0$. Then: $x \geq 4a\log(2a) + 2b \Rightarrow x \geq a\log(x) + b$

This lemma implies that $k < 4d\log(2d) + 2\log r$

Prove that for $r=2$ it holds that $VCdim(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 2d+1$

We may assume that $VCdim(\mathcal{H}_1) = VCdim(\mathcal{H}_2) = d$. Let $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$. Let k be a positive integer such that $k \geq 2d+2$. We show that $\tau_{\mathcal{H}}(k) < 2^k$. By Sauer's lemma

$$
\begin{aligned}
\tau_{\mathcal{H}}(k) &\leq \tau_{\mathcal{H}_1}(k) + \tau_{\mathcal{H}_2}(k) \\
&\leq \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=0}^{d} \binom{k}{i} \\
&= \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=0}^{d} \binom{k}{k-i} \\
&= \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=k-d}^{k} \binom{k}{i} \\
&\leq \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=d+2}^{k} \binom{k}{i} \\
&< \sum_{i=0}^{d} \binom{k}{i} + \sum_{i=d+1}^{k} \binom{k}{i} \\
&= \sum_{i=0}^{d} \binom{k}{i} = 2^k
\end{aligned}
$$