Learning Theory - Homework 1

Alexandru Mocanu, SCIPER 295172

19 March 2019

1 Exercise 5.1

1.1 Statement

Prove that Equation (5.2) suffices for showing that $P\left[L_D(A(S)) \geq 1/8\right] \geq 1/7$.

1.2 Solution

From equation 5.2 we have that $\mathbb{E}\left[L_D(A'(S))[\right] \geq 1/4$ for any algorithm A' and a well-chosen distribution D. We want to prove that $P\left[L_D(A(S)) \geq 1/8\right] \geq 1/7$. Note that we can write $\theta = L_D(A(S))$ and then, using F as the cdf of D, we get:

$$p = P(\theta \geq 1/8) = \int_{1/8}^{1} F(\theta)d\theta \geq \int_0^1 \theta F(\theta)d\theta - \int_0^{1/8} \theta F(\theta)d\theta \geq \mathbb{E}[\theta] - \frac{1}{8}\int_0^{1/8} F(\theta)d\theta \geq \geq \frac{1}{4} - \frac{1}{8}(1-p) = \frac{1}{8} + \frac{1}{8}p \quad (1)$$

From this we get that $p \geq 1/7$, which is exactly what we wanted to prove.

2 Exercise 6.2

2.1 Statement

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

2.2 Solution

In what follows, for each subproblem we use $\mathcal{H}$ for the hypothesis classes for a shorter notation.

1. Given that any $h \in \mathcal{H}$ assigns exactly k values of 1 for elements from $\mathcal{X}$, we can not have $C \subseteq \mathcal{X}$ with $|C| > k$ such that $\mathcal{H}$ "shatters" C, as we could not do an all-one classification. Therefore, $VCdim(\mathcal{H}) \leq k$.

If $|\mathcal{X}| \geq 2k$, we can always find some $h \in \mathcal{H}$ such that for a given $C \subseteq \mathcal{X}$, $|C| = k$, we obtain one of the 2^k classifications of the elements of C. Therefore, $VCdim(\mathcal{H}) = k$ for $|\mathcal{X}| \geq 2k$.

If $|\mathcal{X}| < 2k$, as we are restricted to have exactly k ones, we can have at most $|\mathcal{X}| - k$ zeros, so $VCdim(\mathcal{H}) \leq |\mathcal{X}| - k$, but like in the previous case, we can achieve this upper bound.

In conclusion, $VCdim(\mathcal{H}) = \min\{k, |\mathcal{X}| - k\}$.

2. If $\mathcal{X} > 2k+1$, for $|C| = 2k+2$, we can not have $(k+1)$ elements classified as 1 and the other $(k+1)$ classified as 0. However, for $|C| = 2k+1$, if we have more than k ones associated to C, we will implicitly have fewer than $(k+1)$ zeros associated to C and vice versa. Therefore, we can find $h \in \mathcal{H}$ to do any classification on C.

We are also restricted by the size of $|\mathcal{X}|$ with respect to k, so $VCdim(\mathcal{H}) = \min\{|\mathcal{X}|, 2k+1\}$

3 Exercise 6.5

3.1 Statement

VC-dimension of axis aligned rectangles in $\mathbb{R}^d$: Let $\mathcal{H}^d_{rec}$ be the class of axis aligned rectangles in $\mathbb{R}^d$. We have already seen that $VCdim(\mathcal{H}^2_{rec}) = 4$. Prove that in general, $VCdim(\mathcal{H}^d_{rec}) = 2d$.

3.2 Solution

Suppose we have more than $2d$ points from $\mathbb{R}^d$. For each axis, we select the points with the minimum and the maximum coordinates. We therefore get a box defined by at most $2d$ points. The other points lie inside the box. Therefore, we can not classify the points defining the box with 1 and the inner points with 0 at the same time. This implies that $VCdim(\mathcal{H}) \leq 2d$. For $2d$ points, choosing them as $(0, ..., 0, \pm 1, 0, ..., 0)$ i.e. one-hot vectors, we see that for any subset of these points there is a box containing only them, so $VCdim(\mathcal{H}) = 2d$.

4 Exercise 6.8

4.1 Statement

It is often the case that the VC-dimension of a hypothesis class equals (or can be bounded above by) the number of parameters one needs to set in order to define each hypothesis in the class. For instance, if H is the class of axis aligned rectangles in $\mathbb{R}^d$, then VCdim(H) = 2d, which is equal to the number of parameters used to define a rectangle in $\mathbb{R}^d$. Here is an example that shows that this is not always the case. We will see that a hypothesis class might be very complex and even not learnable, although it has a small number of parameters.

Consider the domain $\mathcal{X} = \mathbb{R}$, and the hypothesis class

$$\mathcal{H} = \{x \mapsto \lceil sin(\theta x) \rceil : \theta \in \mathbb{R}\} \tag{2}$$

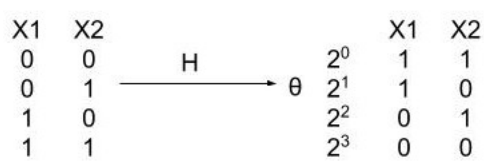
(here we take $\lceil -1 \rceil = 0$). Prove that $VCdim(\mathcal{H}) = \infty$.

4.2 Solution

In order for $VCdim(\mathcal{H}) = \infty$, we have to show that $\forall d \in \mathbb{N}, \exists C \subseteq \mathcal{X}$, $|C| = d$, such that $\mathcal{H}$ "shatters" C.

We fix $d \in \mathbb{N}$ and build the set $C = \{X_1, X_2, ..., X_d\}$ with binary representations $X_j = 0.\underbrace{0...0}_{2^{d-j}}\underbrace{1...1}_{2^{d-j}}\underbrace{0...0}_{2^{d-j}}\underbrace{1...1}_{2^{d-j}}...$, so that element X_j is composed from a repetition of alternating 0/1 blocks of length 2^{d-j}.

We use the fact that for a binary represented number $x = 0.x_1x_2...$, we get $\lceil sin(2^m \pi x) \rceil = 1 - x_m$. Considering now $\theta = \{2^{2^d}\pi, 2^{2^{d-1}}\pi, ..., 2^0\pi\}$, we obtain all the 2^d possible classifications of C. To visualize it, we look at how C is mapped under the previous classification for $d = 2$:



d can be chosen arbitrarily, so we conclude that $VCdim(\mathcal{H}) = \infty$.

5 Exercise 6.9

5.1 Statement

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

$$h_{a,b,s}(x) = \begin{cases} s & \text{if } x \in [a, b] \\ -s & \text{if } x \notin [a, b] \end{cases} \tag{3}$$

Calculate $VCdim(\mathcal{H})$.

5.2 Solution

For the interval class, we had a VC-dimension of 2. Consider $C \in \mathbb{R}$ with 3 points $x_1 \leq x_2 \leq x_3$. For the interval class, we could not do a classification (1,-1,1), but this is possible now with a classifier $h_{a,b,-1}$ which has $x_1 < a \leq x_2 \leq b < x3$.

However, for 4 points $x_1 \leq x_2 \leq x_3 \leq x_4$, we can not find a classifier in $\mathcal{H}$ to give the classification (-1,1,-1,1), as this requires to have two disjoint intervals $[a, b]$ and $[c, d]$ either for $s = -1$ with $a \leq x_1 \leq b < x_2 < c \leq x_3 \leq d < x_4$ or for $s = 1$ with $x_1 < a \leq x_2 \leq b < x_3 < c \leq x_4 \leq d$.

In conclusion, $VCdim(\mathcal{H}) = 3$.

6 Exercise 7.3

6.1 Statement

- Consider a hypothesis class $\mathcal{H} = \cup_{n=1}^{\infty} \mathcal{H}_n$, where for every $n \in \mathbb{N}$, $\mathcal{H}_n$ is finite. Find a weighting function $w : \mathcal{H} \to [0, 1]$ such that $\sum_{h \in \mathcal{H}} w(h) \leq 1$ and so that for all $h \in \mathcal{H}$, $w(h)$ is determined by $n(h) = \min\{n : h \in \mathcal{H}_n\}$ and by $|\mathcal{H}_{n(h)}|$.
- (*) Define such a function w when for all n, $\mathcal{H}_n$ is countable (possibly infinite).

6.2 Solution

- If we set $w(h) = \frac{6}{(\pi n(h))^2} \frac{1}{|\mathcal{H}_{n(h)}|}$, we get

$$\sum_{h \in \mathcal{H}} w(h) \leq \sum_{n=1}^{\infty} \frac{6}{(\pi n)^2} \frac{1}{|\mathcal{H}_n|} |\mathcal{H}_n| = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} = 1, \tag{4}$$

  so we get a valid set of weights.

- Each set $\mathcal{H}_n$ is countable, so within $\mathcal{H}_n$, there is a mapping from $h \in \mathcal{H}_n$ to $\mathbb{N}$. Considering this, we associate an index i to each hypothesis $h_i \in \mathcal{H}_n$. Given $h_i \in \mathcal{H}_{n(h)}$, we compute $w(h_i) = \left(\frac{6}{\pi^2}\right)^2 \frac{1}{n(h)^2} \frac{1}{i^2}$. Therefore

$$\sum_{h \in \mathcal{H}} w(h) \leq \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \left(\frac{6}{\pi^2}\right)^2 \frac{1}{n^2} \frac{1}{i^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \frac{6}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} = 1, \tag{5}$$

  so we again get a valid set of weights.