TEOREMA NFL Y DIMENSION VC
PROBLEMAS

ESTUARDO A. DÍAZ H., 16110

Equation 5.2: $\underset{S\sim\mathcal{D}^m}{\mathbb{E}}\left[L_{\mathcal{D}}\left(A(S)\right)\right] \geq 1/4$

Lemma 1 (B1)**.** Let Z be a random variable that takes values in $[0,1]$. Assume that $\mathbb{E}[Z] = \mu$. Then, for any $a \in (0,1)$

$$\mathbb{P}[Z > 1-a] \geq \frac{\mu - (1-a)}{a}$$

This also implies that for every $a \in (0,1)$,

$$\mathbb{P}[Z > a] \geq \frac{\mu - a}{1-a} \geq \mu - a$$

- **(Cap 5, Ejercicio 1)** Prove that Equation (5.2) suffices for showing that

  $$\mathbb{P}\left[L_{\mathcal{D}}(A(S)) \geq 1/8\right] \geq 1/7$$

  Proof. Let θ be a random variable that takes values in $[0,1]$ and whose expectation satisfies $\mathbb{E}[\theta] \geq 1/4$. By Lemma B1, with $Z = \theta$ and $a = \frac{1}{8}$, we have $\mathbb{P}[\theta > \frac{1}{8}] \geq \frac{1}{4} - \frac{1}{8} = \frac{1}{8} \geq 1/7$. Notice that for any algorithm A and sample $S \sim D^m$, $L_{\mathcal{D}}\left(A(S)\right)$ is a random variable that takes values in $[0,1]$ and, by the Equation (5.2), $\underset{S\sim\mathcal{D}^m}{\mathbb{E}}\left[L_{\mathcal{D}}\left(A(S)\right)\right] \geq 1/4$. Hence $\mathbb{P}\left[L_{\mathcal{D}}(A(S)) \geq 1/8\right] \geq 1/7$. $\square$

- **(Cap 5, Ejercicio 2)** Assume you are asked to design a learning algorithm to predict whether patients are going to suffer a heart attack. Relevant patient features the algorithm may have access to include blood pressure (BP), body-mass index (BMI), age (A), level of physical activity (P), and income (I). You have to choose between two algorithms; the first picks an axis aligned rectangle in the two dimensional space spanned by the features BP and BMI and the other picks an axis aligned rectangle in the five dimensional space spanned by all the preceding features.
  1. **Explain the pros and cons of each choice.**
     R: The first algorithm requires less samples in order to find the appropriate rectangle, but this rectangle may not represent fairly the data since it is taking into account all the variables, for example the age (A) might be a deterministic variable over the probability of suffering a heart attack. On the other hand, the second algorithm might give a more precise solution, but it would require more sample data to train properly.
  2. **Explain how the number of available labeled training samples will affect your choice.**
     R: If there are enough sample data, then it would be best to choose the second algorithm. If the sample data is to restricted tough, it is better to choose the first algorithm to being able to train it properly, even if its not as accurate as the second one.

- **(Cap 6, Ejercicio 5) 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, $VCdim(\mathcal{H}^d_{rec}) = 2d$.

  Proof. ($\leq$) Let us consider a set $A = \{x_1, \dots, x_{2d+1}\} \subset \mathbb{R}^d, d \in \mathbb{Z}^+$. Now, for each coordinate $i = 1, \dots, d$, consider the canonical projection π_i, where $\pi_i((a_1, \dots, a_i, \dots, a_n)) = (0, \dots, a_i, \dots, 0)$. We can calculate $\underset{x\in A}{min}\{\pi_i(x)\}$ and $\underset{x\in A}{max}\{\pi_i(x)\}$ for every $i = 1, \dots, d$. Such points completely determine a rectangle R_A in $\mathbb{R}^d$ and notice there are at most $2d$ points of A touching the boundary of R_A. Also, by definition, R_A is the smallest rectangle such that $A \subset R_A$. That is, there is at

least one point p in the interior of R_A. Therefore, it is not possible to find a rectangle that classify as 1 the points of the boundary of R_A and classify p as -1. Hence, $VCdim(\mathcal{H}^d_{rec}) \leq 2$.
($\geq$) On the other hand, consider the set $B = \{(0, \dots, a_i, \dots, 0) \in \mathbb{R}^d : a_i \in \{1, -1\}, i = 1, \dots, d\}$, the set of centers of the faces of a cube centered in the origin. Let $\sigma = (0, \dots, 0) \in \mathbb{R}^d$ represent the origin. Consider an arbitrary subset C of B. Recall that a rectangle can be fully determined by two opposite vertices. Now, consider the following points: $p_1 = \sum_{i=1}^{d} \min_{x \in C \cup \{\sigma\}} \{\pi_i(x)\}$ and $p_2 = \sum_{i=1}^{d} \max_{x \in C \cup \{\sigma\}} \{\pi_i(x)\}$. This two points define the vertices of a rectangle R_C that contains C but not contains any point of $B \setminus C$, therefore $\mathcal{H}$ shatters B. Hence $VCdim(\mathcal{H}^d_{rec}) = 2d$. □

- **(Cap 6, Ejercicio 7)** We have shown that for a finite hypothesis class $\mathcal{H}$, $VCdim(\mathcal{H}) \leq log(|\mathcal{H}|)$. However, this is just an upper bound. The VC-dimension of a class can be much lower than that:
  (1) **Find an example of a class $\mathcal{H}$ of functions over the real interval $X = [0, 1]$ such that $\mathcal{H}$ is infinite while $VCdim(\mathcal{H}) = 1$.**
  R: Consider $\mathcal{H} = \{h_a : a \in [0, 1]\} \ni h_a(x) = \begin{cases} 1 & x = a \\ -1 & x \neq a \end{cases}$. Clearly $VCdim(\mathcal{H}) = 1$.
  (2) **Give an example of a finite hypothesis class $\mathcal{H}$ over the domain $X = [0, 1]$, where $VCdim(\mathcal{H}) = log_2(|\mathcal{H}|)$.**
  R: Consider $\mathcal{H} = \{h_1, h_{-1}\}$, where $h_s(x) = s$. Notice that
  $VCdim(\mathcal{H}) = 1 = log_2(2) = log_2(|\mathcal{H}|)$.

- **(Cap 6, Ejercicio 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

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

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

Claim: $VCdim(\mathcal{H}) = 3$.

Proof. Let us consider the set $A = \{x, y, z\} \subset \mathbb{R} \ni x < y < z$. We show that $\mathcal{H}$ can shatter A: For the empty set, consider $h_{x-1,z+1,-1}$, it maps x, y, z to -1. Similary, $h_{x-1,z+1,1}$ maps x, y, z to 1. Now, for each element $a \in A$, we can find an closed disc of radius ϵ such that $D_\epsilon(a) \cap A = a$, i.e. a closed neighbourhood of a that not touches any other point of A. It follows that $h_{a-\epsilon,a+\epsilon,1}$ maps a to 1 and A^c to -1. Similarly, $h_{a-\epsilon,a+\epsilon,-1}$ maps a to -1 and A^c to 1. Therefore $VCdim(\mathcal{H}) \geq 3$. Now consider the set $B = \{w, x, y, z\} \subset \mathbb{R} \ni w < x < y < z$. There is no $h_{a,b,s} \in \mathcal{H}$ that can map w, y to 1 and x, z to -1, because that would imply that the whole interval $[w, y] \subseteq [a, b]$ and $x \in [w, y] \implies x \in [a, b] \implies h_{a,b,s}(x) = 1$. Therefore $\mathcal{H}$ can not shatter B. Hence, $VCdim(\mathcal{H}) = 3$. □