AML-Assignment-1 - Lucian Istrati

lucianistrati

April 2022

1 Exercise 1

1. **(0.5 points)** Give an example of a finite hypothesis class $\mathcal{H}$ with VCdim($\mathcal{H}$) = 2022. Justify your choice.

2 Solution 1

Having A = e1,e2,...,en the orthonormal basis of R^n, we define a finite hypothesis class H as follows: $H = \{h(w,0) : R^n -> 1, 1, h(w,0) = sign(\Sigma w_i * x_i) | w_i = \{1, if e_i$ belongs to B, or 1, if e_i belongs to B for all subsets B of A.

We will show that $VCdim(H) = n$ and choose $n = 2022$ for this particular case.

We first show that $VCdim(H) >= n$ by finding a set A of n points in Rn that is shattered by H. As in in example given in the lectures for the halfspaces we take $A = e1, e2, ..., en$ to be the orthonormal basis of R^n. Using the alternative definition of shattering, we need to show that for every subset B of A there is a function hB that labels +1 all elements in B and 1 all elements in A minus B.

By construction, our H contains exactly one corresponding hypothesis for each subset B in A:

$hB = hw, 0 = sign(\Sigma w_i x_i)$ where $w_i = \{1, if e[i]$ belongs to B 1, if e_i does not belong to B .

We have $hB = sign(w, e_i) = w_i$, so it will assign +1 for all elements in B and 1 for all elements in B minus A, therefore proving that $VCdim(H) >= n$.

Now we will show that $VCdim(H) < n + 1$. We use a property shown in the lectures: for a finite hypothesis class H we have the following upper bound: $VCdim(H) <= log2(|H|)$

For our choice of H we have exactly one hypothesis for each subset of A, resulting in $|H| = 2|A| = 2n$. Therefore: $VCdim(H) <= [log2(2^n])] = n$

Using both parts results in $VCdim(H) = n$ with n being 2022.

3 Exercise 2

2. **(0.5 points)** What is the maximum value of the natural even number n, $n = 2m$, such that there exists a hypothesis class $\mathcal{H}$ with n elements that

shatters a set C of $m = \frac{n}{2}$ points? Give an example of such an $\mathcal{H}$ and C. Justify your answer.

4 Solution 2

During the lectures the following property was given: $VCdim(H) <= log2(|H|)$

Another definition of "Shattering" from the lectures states that: A hypothesis class H shatters a finite set C of X, if the restriction of H to C is the set of all functions from C to 0, 1. That is $|H(C)| = 2^|C|$.

If H shatters a set of n divided by 2 points, the its VCdim(h) must be greater than or equal to n divided by 2.

But because of the previosuly mentioned property, we also know VCdim(H) is smaller than or equal to $log2(n)$.

Because of the two previously mentioned arguments we will obtain the following $n/2 <= log2(n)$.

But we know that n = 2 * m. So $m <= log2(2 * m)$ which is equivalent to $m <= 1 + log2(m)$, by moving everything in the left hand side we will obtain $m - 1 - log2(m) <= 0$. Because we are required to find the maximal value for n (or m in this case, we thus need to maximize the value fro m).

By plotting in an online plot setting it can be seen that this function has values smaller than or equal to zero just between 1 and 2 (hitting zero for m=1 and for m=2), thus 2 ¿ 1, then m = 2 is the maximal value so then n = 2 * 2 = 4 is the maximal value for which the condition defined in the hypothesis.

5 Exercise 3

3. **(0.75 points)** Let $\mathcal{X} = R^2$ and consider $\mathcal{H}$ the set of axis aligned rectangles with the center in origin O(0, 0). Compute the $VCdim(\mathcal{H})$.

6 Solution 3

Because H is defined as the set of all axis aligned rectangles with the center in the origin O(0,0), thus, it can be considered as a sort of particular case of the H^2_{rec} with the exception that the rectangle must have its center (defined as the intersection of its diagonals in the origin O(0, 0)).

We know that Hrec2 is the set of axis aligned rectangles in the R2. $Hrec2 = [a, b][c, d]|a <= b, c < +d, a, b, c, dR$

In order to compute the VCdim(H) we will eventually proof that H can shatter 2 points, but cannot shatter 3 points and thus the VCdim(H) will be 2.

As it can be seen from the figures, it can be shown that for 2 points and the 4 possible cases $2\hat{2} = 4$ the H shatters all of them, then $VCdim(H) >= 2$.

As it can be seen from the figures, it can be shown that for 3 points and the 16 possible cases ($2^3 = 8$) possible cases, 16 because we are treating both the cases when the 3 points are collinear and when the 3 points can form a triangle)

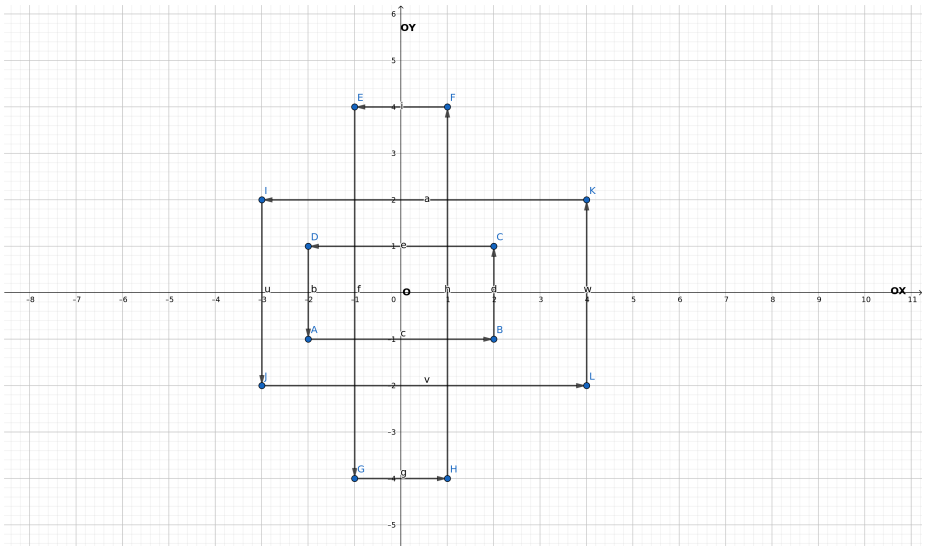

Figure 1: Figure for exercise 3

the H shatters does not of them, then VCdim(H) cannot be greater than or equal to 3, thus the VCdim(H) will be 2.

7 Exercise 4

4. **(1 point)** Let $\mathcal{X} = R^2$ and consider $\mathcal{H}_\alpha$ the set of concepts defined by the area inside a right triangle ABC with two catheti AB and AC parallel to the axes (Ox and Oy), and with the ratio AB/AC = α (fixed constant > 0). Consider the realizability assumption. Show that the class $\mathcal{H}_\alpha$ is (ϵ, δ)-PAC learnable by giving an algorithm A and determining an upper bound on the sample complexity $m_H(\epsilon, \delta)$ such that the definition of PAC-learnability is satisfied.

8 Solution 4

9 Exercise 5

5. **(1.25 points)** Consider $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3$, where:

$$\mathcal{H}_1 = \{h_{\theta_1} : R \to \{0,1\} \mid h_{\theta_1}(x) =_{[x \geq \theta_1]} (x) =_{[\theta_1, +\infty)} (x), \theta_1 \in R\},$$

$$\mathcal{H}_2 = \{h_{\theta_2} : R \to \{0,1\} \mid h_{\theta_2}(x) =_{[x < \theta_2]} (x) =_{(-\infty, \theta_2)} (x), \theta_2 \in R\},$$

$$\mathcal{H}_3 = \{h_{\theta_1, \theta_2} : R \to \{0,1\} \mid h_{\theta_1, \theta_2}(x) =_{[\theta_1 \leq x \leq \theta_2]} (x) =_{[\theta_1, \theta_2]} (x), \theta_1, \theta_2 \in R\}.$$

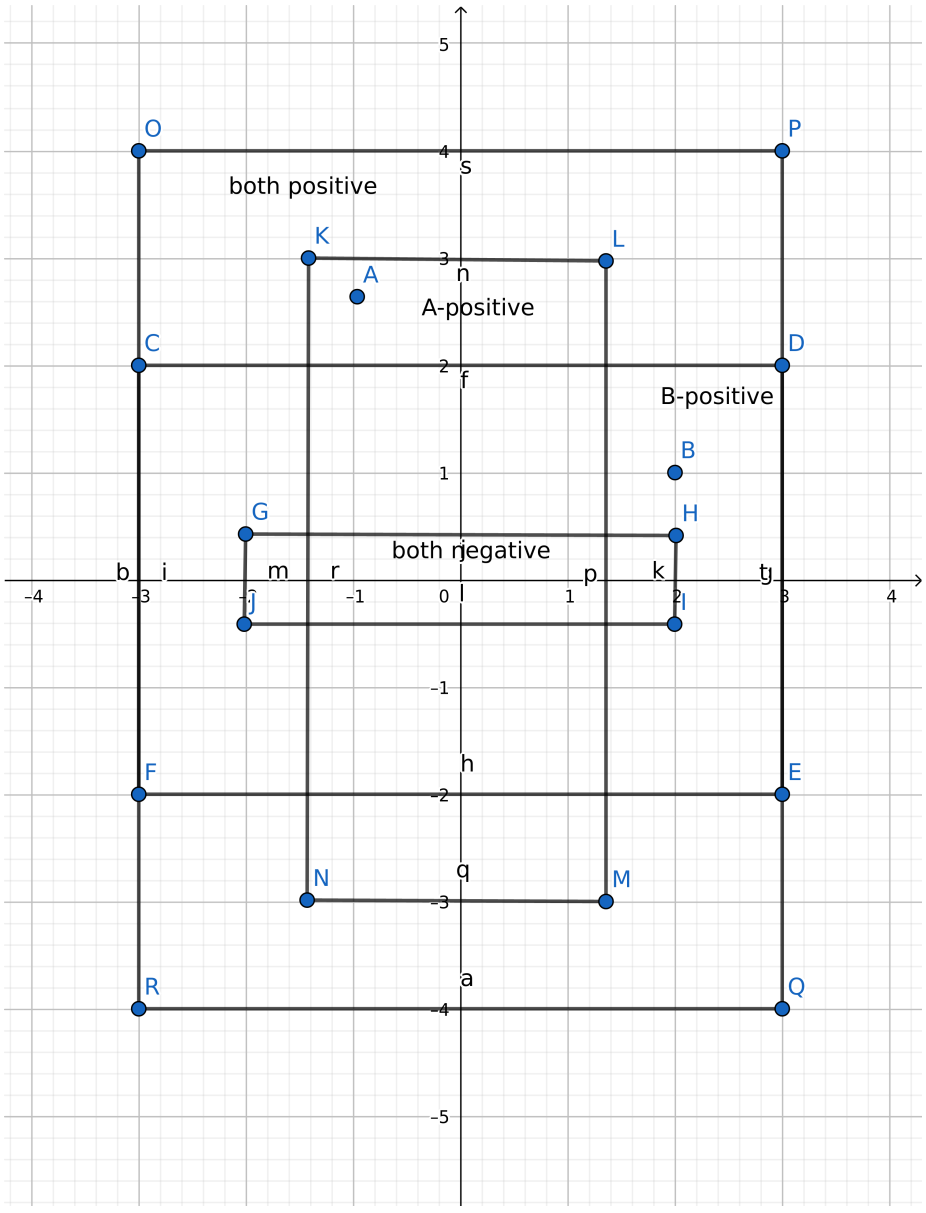

Figure 2: 2 points case

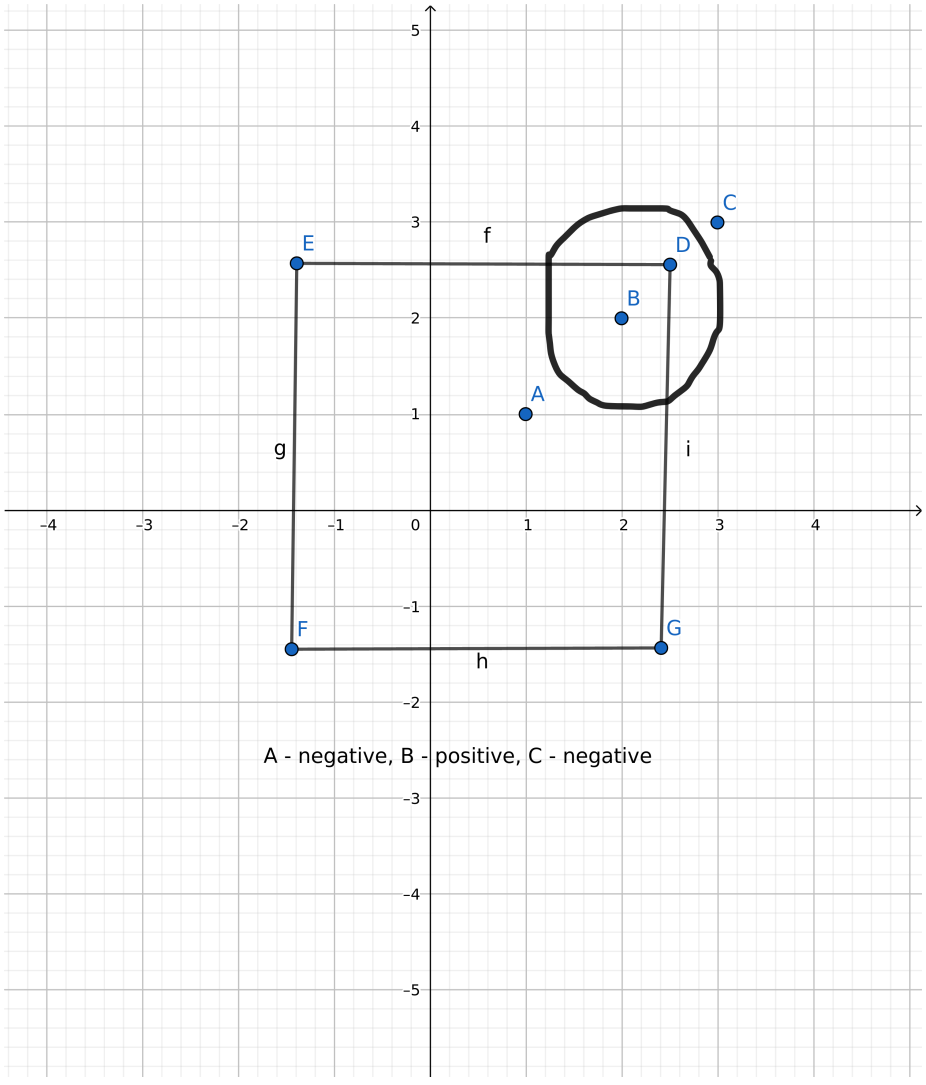

Figure 3: 3 points colinear case

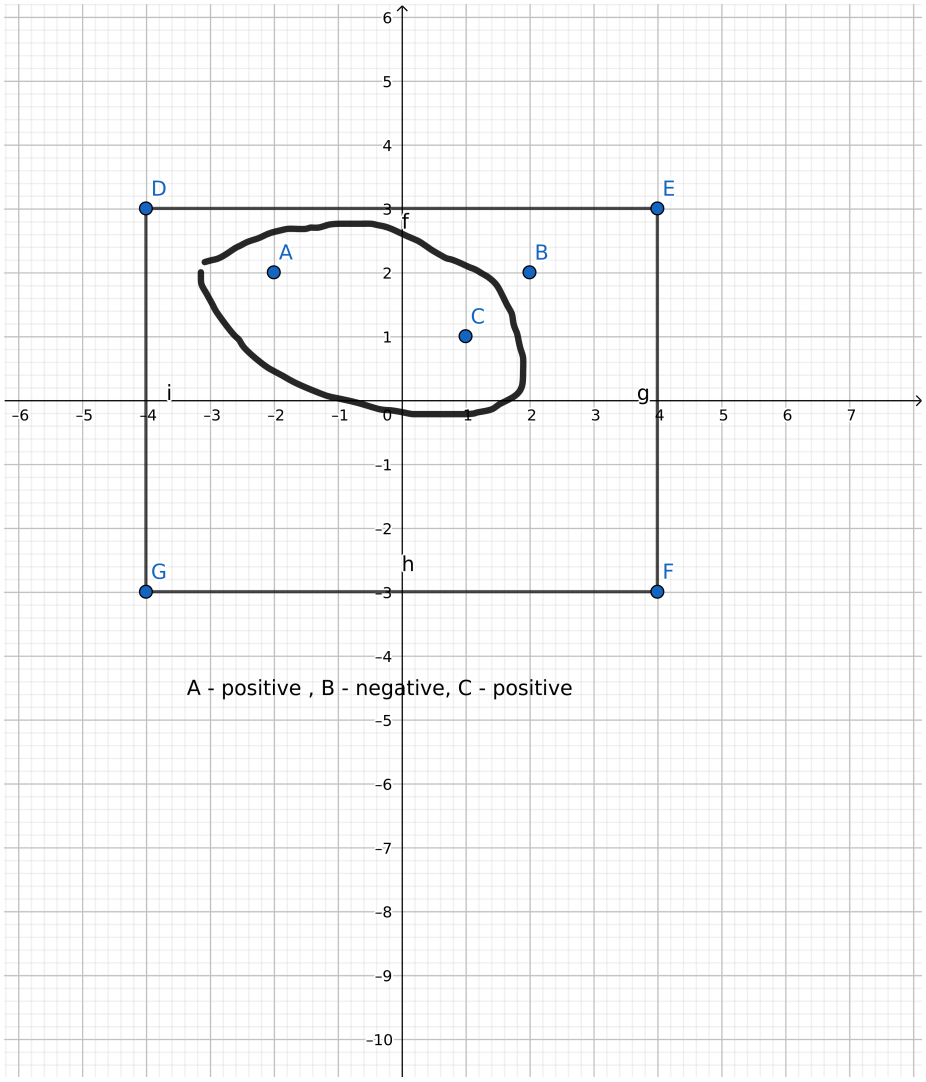

Figure 4: 3 points triangle case

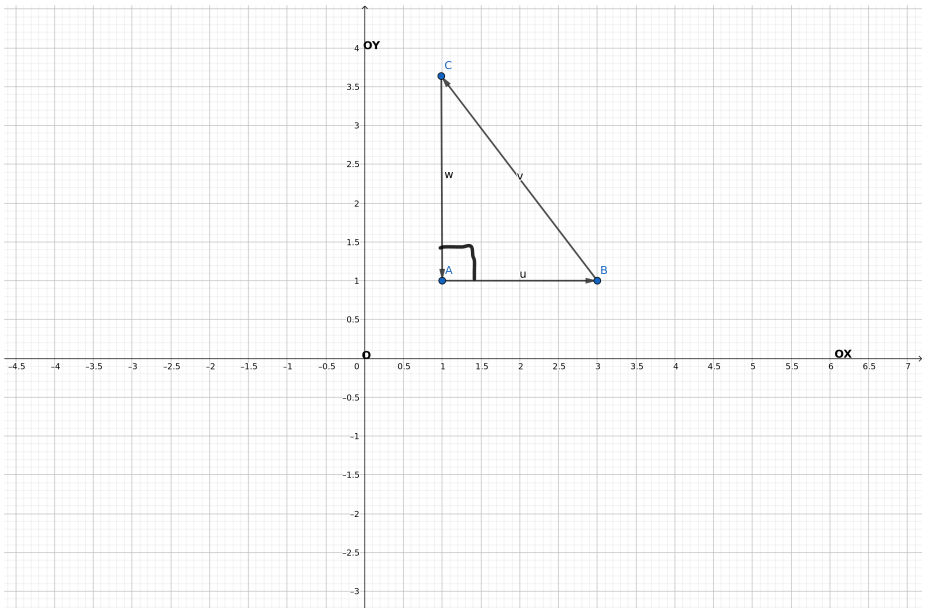

Figure 5: Figure for exercise 4

Figure for exercise 4

Consider the realizability assumption.

a) Compute VCdim($\mathcal{H}$).

b) Show that $\mathcal{H}$ is PAC-learnable.

c) Give an algorithm A and determine an upper bound on the sample complexity $m_{\mathcal{H}}(\epsilon, \delta)$ such that the definition of PAC-learnability is satisfied.

10 Solution 5

a) We know that the VCdim(H intervals) = 2 and VCdim(H thresholds) = 1. So, in our case VCdim(H1) = VCdim(H2) = 1 and VCdim(H3) = 2

H is defined as the reunion of H1, H2 and H3. By applying the VCdim definition, we want to prove that VCdim(H) = 2. Thus, we want to show that: 1. There C included in R, where $|C| = 2$, that is shattered by H. ($VCdim(H) >= 2$) 2. For any C included R, where $|C| = 3$, C is not shattered by H ($VCdim(H) < 3$)

Let's consider the case with 2 points, however we might choose them there will be a theta1 or a theta2 such that there will be a function to separate each case (explanation: they are both 1 and we choose H2 function with a large enough theta2; they are both 0 so we choose a H2 function with a small enough theta2; the smaller one is 1 and the bigger one is 0 then we choose a function

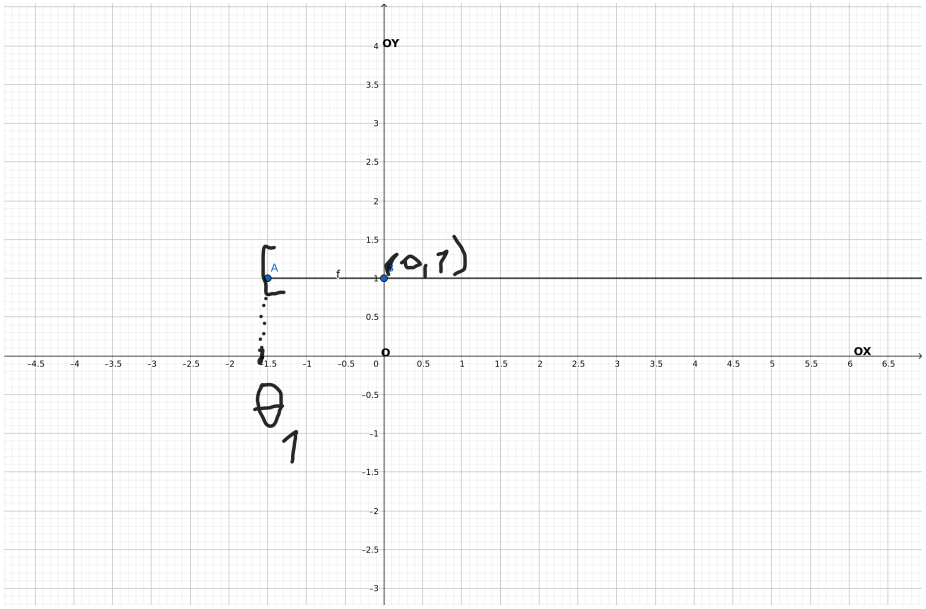

Figure 6: Figure for exercise 5 (H1)

Figure for exercise 5 (H1)

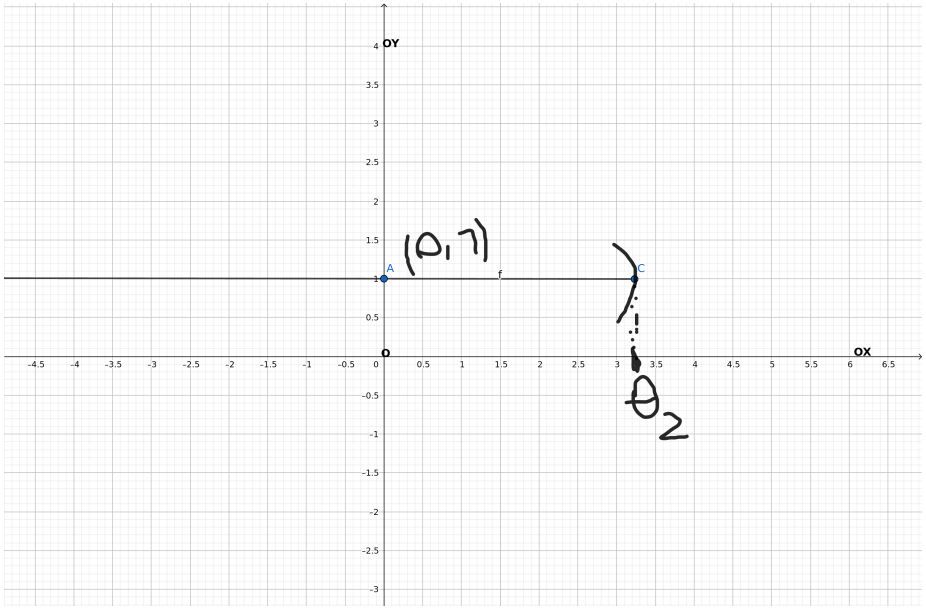

Figure 7: Figure for exercise 5 (H2)

Figure for exercise 5 (H2)

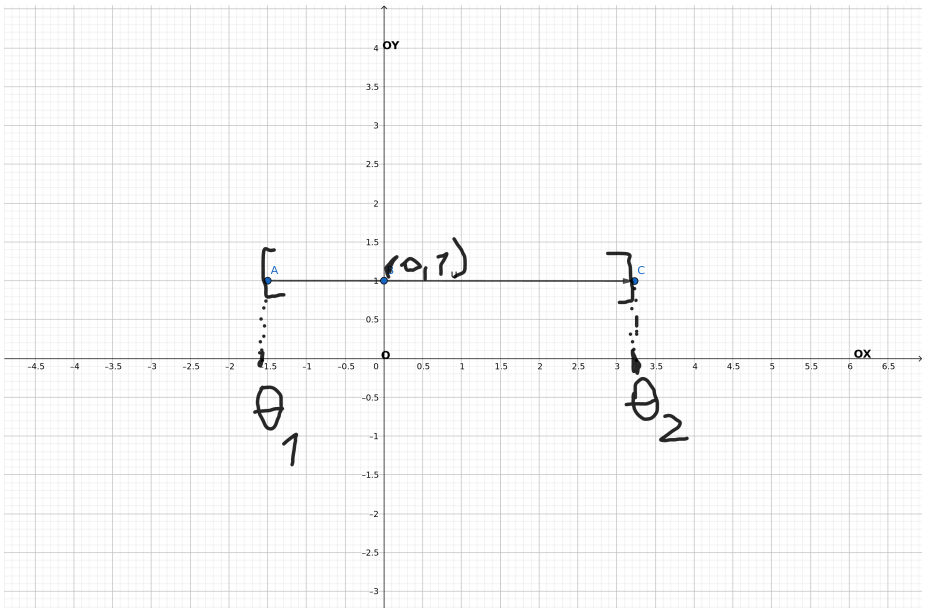

Figure 8: Figure for exercise 5 (H2)
Figure for exercise 5 (H3)

from H2 with theta2 between them; the smaller one is 0 and the higher one is 1 then we choose a function from H1 with theta1 between them). So, VCdim(H) ≥ 2.

Now we will come to prove that VCdim cannot be 3. Let's say we choose 3 points A,B,C and we consider a table of possible values for each of them:

- 0 0 0 - function from H1 or H2
- 0 0 1 - function from H1
- 0 1 0 - function from H3
- 0 1 1 - function from H1
- 1 0 0 - function from H2
- 1 1 0 - function from H2
- 1 1 1 - function from H1 or H2
- 1 0 1 - no possible function!

For the combination 1,0,1 (meaning that A belongs to 1 class, B belongs to 0 class, C belongs to 1 class) and A ¡ B ¡ C it is no possible to find any function that shatters these 3 points because a function from H1 cannot have a 1 to the

left of a 0, a function from H2 cannot have a 1 to the right of a zero, and a function from H3 just admits "ones" "surrounded" by "zeros" to the left and right, but not "zeros" which are "surrounded" by "ones". The demonstration itself is trivial.

b) Since VCdim(H) is finite, then H is also PAC-learnable because of the Fundamental Theorem of Statistical Learning where 6 statements are given, each being equivalent to each other. In this case, statement 6 from the Theorem is equivalent to statement 4.

11 Exercise 6

6. **(1 point)** A decision list may be thought of as an ordered sequence of if-then-else statements. The sequence of conditions in the decision list is tested in order, and the answer associated with the first satisfied condition is output.

More formally, a *k-decision list* over the boolean variables $x_1, x_2, \ldots, x_n$ is an ordered sequence $L = \{(c_1, b_1), (c_2, b_2), \ldots, (c_l, b_l)\}$ and a bit b, in which each c_i is a conjunction of at most k literals over $x_1, x_2, \ldots, x_n$ and each $b_i \in \{0, 1\}$. For any input $a \in \{0, 1\}^n$, the value $L(a)$ is defined to be b_j where j is the smallest index satisfying $c_j(a) = 1$; if no such index exists, then $L(a) = b$. Thus, b is the "default" value in case a falls off the end of the list. We call b_i the bit associated with the condition c_i.

The next figure shows an example of a *2-decision list* along with its evaluation on a particular input.

Show that the VC dimension of 1-decision lists over $\{0, 1\}^n$ is lower and upper bounded by linear functions, by showing that there exists $\alpha, \beta, \gamma, \delta \in R$ such that:

$$\alpha \cdot n + \beta \leq VCdim(\mathcal{H}_{1-decision\ list}) \leq \gamma \cdot n + \delta$$

Hint: Show that 1-decision lists over $\{0, 1\}^n$ compute linearly separable functions (halfspaces).

Ex-officio: 0.5 points

12 Solution 6

Before showing that the VC dimensions of 1-decision lists over 0,1n is bounded by linear functions, we will first show that 1-decision lists over 0,1n̂ computes linearly separable functions (halfspaces).

In the hypothesis it is mentioned that a k-decision list over the boolean variables x1,x2,...,xn is an ordered sequence L = (c1,b1), (c2,b2),..., (cl,bl) and a bit b, in which each ci is a conjunction of at most k literals over x1,x2,...,xn and each bi 0,1. In the lectures, it is also mentioned that these boolean variables x_i with i from 1 to n are also named features, so we will refer to them in this

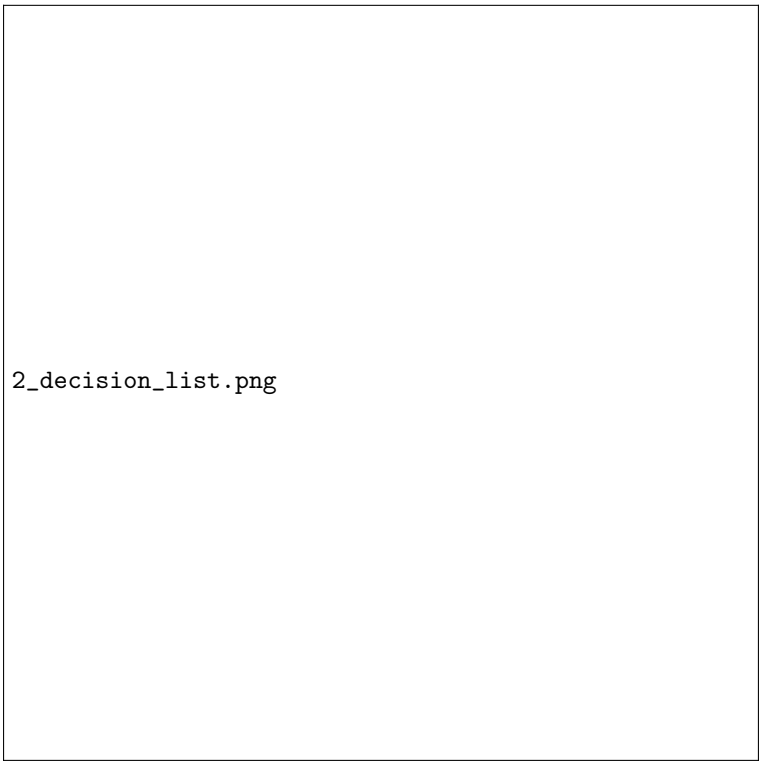

Figure 9: *A 2-decision list and the path followed by an input. Evaluation starts at the leftmost item and continues to the right until the first condition is satisfied, at which point the binary value below becomes the final result of the evaluation.*

way as well)in a way, this is comparable to the actual scenarion in a machine learning context where there are n possible features a sequence of length can have at each step, with each one of them corresponding to an output, or a b_i in our case, you can also kind of see them as a sort of a decision tree where you have the possibility to stop and produce and output at every single step or to keep going down the tree and in the end you will be faced with two leaves b_l if the result of the last conjuction is True or the default b values if the last result was False, as same the other previous l-1 results). Since we are dealing with 1-decision lists, this means that we will be having no operation between the features(boolean variables) other than their negation and each conjuction will have at most 1 literal.

So an example of 1-decision list could be the following:

- l - the size of L = 4
- b = 0 - default value
- L = (x1, 0), (not(x2), 1), (x3, 1), (not(x4), 0)

- with x1,...,x6 being the literals, and an input a=011011
- x1=0
- x2=1
- x3=1
- x4=0 5=1
- x6=1

The output for L(a) is 1, since x1 is 0, not(x) will also be 0, but x3 will be 1 and the bit corresponding to x3 which is b3 (which is also 1).

We also know that the VC dim of HS^n (the set of all halfspaces - linear classifiers in R^n) is equal to n+1, as we know that this is mentioned in the lectures in the part where it is demonstrated that $VCdim(HS_0^n)$ is equal to n.

For the case where n is equal to 1, x1 can take either the value 0, either the value 1 and similarly it can produce 0, 1 or both. This case is linearly seperable, because we will have the points (0,0), (1,0), (0,1) and (1, 1); the first two points can be separated from the last two points by an infinity of halfspaces (such as one determined by a linear function like $f : R- > R$ with $f(x) = 0.5$ for any x belonging to R). For the case where n is equal to 2, we know that the posibilities for $x1, x2$ are $(0, 0), (0, 1), (1, 0), (1, 1)$, each of these can produce 0, 1 or both. This case is also linearly separable as this will produce a 3D cube (with size 1) determined by the points (0,0,0), (0,1,0), (1,0,0), (1,1,0), (0,0,1), (0,1,1), (1,0,1) and (1,1,1) which can be a separated by an infinity of halfspaces (such as one determined by a linear function like $f : R * R- > R$ with $f(x, y) = 0.5$ for any x, y belonging to R) Similarly for the cases where n is 1 or 2, in the general case it can be proven that for any n, there will be 2n̂ possible combinations for x1, x2, ..., xn which will further result in 2 (to the power of n) + 1 combinations of coordinates in a n + 1 dimensional space and each of these n+1-dimensional space we will have an infinity of n-dimensional spaces (halfspaces as per the context of the n+1-dimensional spaces,)

So, we know that an n-variables 1-decision list is determining an n+1-space that can be linearly separated by halfspaces of dimension.. Now that we know that 1-decision lists compute linearly separable functions. We must show that their VC dimensions is lower and upper bounded by linear functions (in essence, to prove that their VC dimension is a linear function in itself).

Also, from the lectures, we know that for H_{conj}^d (the class of conjunctions of at most d boolean literals x1, x2,..., xd) we have the following relationship: $|H_{conj}^d| = 3^d + 1$.

Setting the 1-decision lists problem aside for a moment, we will take a look at the H_{conj}^1. So, $|H_{conj}^1| = 3^1+1 = 4$ ¡ infinite, so we will have finite VC dimension (less than $log2(3^1+1) = log2(4) = 2$), also from the lectures we know that H_{conj}^d is equal to d then the VC dim for H_{conj}^1 is 1. Furthermore as it was shown in the lectures as well the sample complexity mH(epsilon, delta) is bounded by the following expression: $C1*(1+log(1/delta))/(epsilon) <= mH(epsilon, delta) <=$

$C2 * (log(1/epsilon) + log(1/delta))/epsilon$, so $mH(epsilon, delta)$ is polynomial in: $1/epsilon$, $1/delta$ and d which is 1 in our case, so then just polynomial in $1/epsilon$ and $1/delta$

Method 1: We also know the following property from lectures $VCdim(H)log2|H|$, in this case $|H| = 2^n$ because there are 2^n possible combinations of inputs to $x1, x2, ..., xn$, so then $VCdim(H) <= log2(2^n <= VCdim(H) <= n$. So, VCdim(H) can be at most linear in n (thus proving the upper bounded part), now what's left is to prove that VCdim(H) is also lower bounded.

In the paper (https://arxiv.org/pdf/2010.07374.pdf) it was proven by Yildiz in 2015 (2nd page, related works, 3rd paragraph) that a lower bound for this class of 1-decision lists (also known as decision stumps) is $[log2(l + 1)] + 1$ where "l" is the number of features, in our case this is n, thus making the expression equivalent to proving that $[log2(n + 1)] + 1$ is a lower bound. Whilst Gey proved in 2018 (2nd page, related works, 4th paragraph) that the VC dimension for 1-decision list is given by the largest integer d which satisfies $2 * n >= d!/(d - d/2)! * (d/2)!$.

Method 2: In our case we the 1-decision lists can be seen as a product of n distinct H_conj^1 as we are in fact having n different conjuctions and $VCdim(H^1_{conj})$ is 1, then our 1-decisions lists must have a VCdim belonging to the class of functions linear in n, thus there exists ,,, R such that: $n+ <= VCdim(H1decisionlist) <= n+$.