Assignment 1

Badea Adrian Catalin

Group 407

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

Solution
Let H^d_{rec} be the class of axis aligned rectangles in $\mathbb{R}^d$.

In *Lecture 6* we proved that $VCdim(H^2_{rec}) = 4$.
In *Seminar 3* we proved generally that $VCdim(H^d_{rec}) = 2 * d$.

Then, we can choose $d = 1011$, for which $VCdim(H^{1011}_{rec}) = 2022$ as the class of aligned rectangles in $\mathbb{R}^{1011}$.

Proof
In order to show that $VCdim(H^d_{rec}) = 2 * d$, we need to show that
1) there exists a set C of $2d$ points that is shattered by (H^d_{rec}).

2) every set C of $2d + 1$ points is not shattered by (H^d_{rec}).

$$\mathcal{H}^d_{rec} = \{h_{(a_1,b_1,a_2,b_2,\ldots,a_d,b_d)} \mid a_i \leq b_i, i = \overline{1,d}\}$$

$$h_{(a_1,b_1,a_2,b_2,\ldots,a_d,b_d)}(\underset{\underline{x}=(x^1,x^2,\ldots,x^d)}{\underline{x}}) = \begin{cases} 1, & a_i \leq x^i \leq b_i \quad \forall i = \overline{1,d} \\ 0, & \text{otherwise} \end{cases}$$

In order to show that $VC\dim(\mathcal{H}^d_{rec}) = 2d$, we need to show that:
1) There exists a set C of $2d$ points that is shattered by $\mathcal{H}^d_{rec}$ (this will mean that $VC\dim(\mathcal{H}^d_{rec}) \geq 2d$)
2) Every set C of $2d + 1$ points is not shattered by $\mathcal{H}^d_{rec}$ (this will mean that $VC\dim(\mathcal{H}^d_{rec}) < 2d + 1$)

Proof of 1) Consider $C = \{c_1, c_2, c_3, \ldots, c_{2d-1}, c_{2d}\}$ where

$$\begin{array}{rll l}
c_1 & = (1,0,0,\ldots,0) & = e_1 & \\
c_2 & = (0,1,0,\ldots,0) & = e_2 & \\
 & \vdots & & \\
c_d & = (0,0,0,\ldots,1) & = e_d & c_i = e_i = -c_{i+d} \\
c_{d+1} & = (-1,0,0,\ldots,0) & = -e_1 & \quad \forall i = \overline{1,d} \\
c_{d+2} & = (0,-1,0,\ldots,0) & = -e_2 & \\
 & \vdots & & \\
c_{2d} & = (0,0,0,\ldots,-1) & = -e_d &
\end{array}$$

For $d = 2$, we will have in 2 dimensions:

$$c_1 = (1,0) \qquad c_2 = (0,1) \qquad c_3 = (-1,0) \qquad c_4 = (0,-1)$$

We want to show that, for each labeling $(y_1, y_2, \ldots, y_{2d})$ of the points $(c_1, c_2, \ldots, c_{2d})$ (there are 2^{2d} possible labelings), there exists a function h in $\mathcal{H}^d_{rec}$ such that $h(c_i) = y_i \; \forall i = \overline{1,2d}$.

Consider a labeling $(y_1, y_2, \ldots, y_{2d}) \in \{0,1\}^{2d}$.

Each point c_i has all components $= 0$, apart from component i if $i \in \{1, \ldots, d\}$ or $i - d$ if $i \in \{d+1, \ldots, 2d\}$.

$$\begin{array}{rl|rl|c|rl}
c_1 & = (\quad 1,0,0,\ldots,0) & c_2 & = (0, \quad 1,0,\ldots,0) & & c_d & = (0,0,0,\ldots, \quad 1) \\
c_{d+1} & = (\ -1,0,0,\ldots,0) & c_{d+2} & = (0, \ -1,0,\ldots,0) & \cdots & c_{2d} & = (0,0,0,\ldots, \ -1)
\end{array}$$

We want to find $h = h_{(a_1,b_1,a_2,b_2,\ldots,a_d,b_d)}$ such that $h_{(a_1,b_1,a_2,b_2,\ldots,a_d,b_d)}(c_i) = y_i$.

The choice of the interval $[a_i, b_i]$ is influenced by the labels y_i and y_{i+d} of the points c_i and c_{i+d}. As all other points $c_1, c_2, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{i+d-1}, c_{i+d+1}, \ldots, c_{2d}$ have 0 on the i-th component, we have that $[a_i, b_i]$ should contain 0, otherwise each point will be labelled with 0.

So $[a_i, b_i]$ depends on y_i and y_{i+d}, and $[a_i, b_i]$ decides basically the label of points c_i and c_{i+d}:

$$c_i = (0, \ldots, 0, 1, 0, \ldots, 0) \qquad c_{i+d} = (0, \ldots, 0, -1, 0, \ldots, 0)$$

Possible cases:

1) $y_i = 0, y_{i+d} = 0$, then $[a_i, b_i] \cap \{-1, 1\} = \emptyset$
$[a_i, b_i]$ should not contain points -1 and 1.
In this case, take $a_i = -0.5, b_i = 0.5$ (many other choices are possible)

2) $y_i = 0, y_{i+d} = 1$, then $[a_i, b_i] \cap \{-1, 1\} = \{-1\}$
$[a_i, b_i]$ should contain only point -1 such that c_{i+d} will get label 1.
In this case, take $a_i = -2, b_i = 0.5$ (many other choices are possible)

3) $y_i = 1, y_{i+d} = 0$, then $[a_i, b_i] \cap \{-1, 1\} = \{1\}$
$[a_i, b_i]$ should contain only point $+1$ such that c_i will get label 1.
In this case, take $a_i = -0.5, b_i = 2$ (many other choices are possible)

4) $y_i = 1, y_{i+d} = 1$, then $[a_i, b_i] \cap \{-1, 1\} = \{-1, 1\}$
$[a_i, b_i]$ should contain both points $\{-1, 1\}$ such that c_i and c_{i+d} will get label 1.
In this case, take $a_i = -2, b_i = 2$ (many other choices are possible)

In all cases, we have that $h_{(a_1,b_1,a_2,b_2,\ldots,a_d,b_d)}(c_i) = y_i, \forall i = \overline{1, 2d}$, where each interval $[a_i, b_i]$ was determined based on y_i and y_{i+d}, $i = \overline{1, d}$.

So, $VC\dim(\mathcal{H}^d_{rec}) \geq 2d$.

$\square$

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.

Solution

If class $\mathcal{H}$ with n elements shatters a set C of $m = \frac{n}{2}$ points, then $VCdim(\mathcal{H}) >= \frac{n}{2}$. (1)

Also, because $\mathcal{H}$ is a class with n elements, then $\mathcal{H}$ is a finite hypothesis class, which means that $VCdim(\mathcal{H}) <= \log_2 n$. (2)

From (1) and (2), we can say that $\frac{n}{2} <= \log_2 n$, where $n \in N$, which means that $\frac{n}{2} - \log_2 n <= 0$.

If we take into consideration that $n = 2 * m$, where $m \in N$, the equation is transformed into $m - \log_2 2 * m <= 0, => m - \log_2 m - 1 <= 0$.

If $m = 0, n = 2 * m = 0$, the hypothesis class (H) is $\emptyset$.

Let $f(m) = m - \log_2 m - 1$, $m \in m$. We have to find the maximum m where $f(m) <= 0, m >= 1, m \in \mathcal{N}$.

For $m = 1, f(m) = 1 - 0 - 1 = 0 <= 0$

For $m = 2, f(m) = 2 - 1 - 1 = 0 <= 0$

For $m = 3, f(m) = 3 - \log_2^3 - 1 = 2 - \log_2^3 > 0$.

Also, for any $m >= 3$,we can show that $f'(m) >= 0$, because $f'(m) = 1 - \frac{1}{m * \ln m}$ and $1 - \frac{1}{m * \ln m} < 1$. As both m and $\ln(m) >= 1$ $(m >= 3)$, then $m * \ln m > 1$.

So we have to find an example for $m = 2, n = 4$, so a hypothesis class with n elements, which shatters a set C of 2 points.

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

Solution

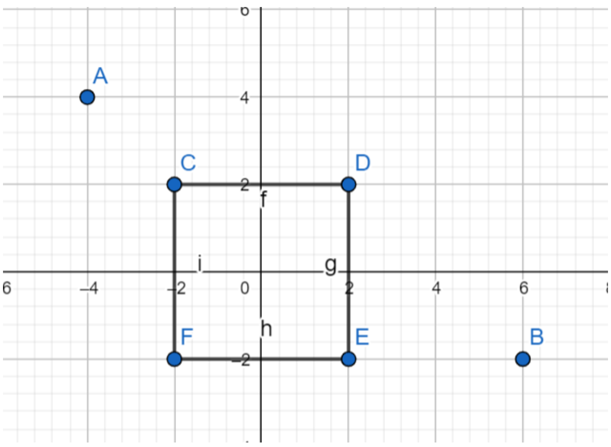
$VCdim(\mathcal{H}) = 2$

Proof

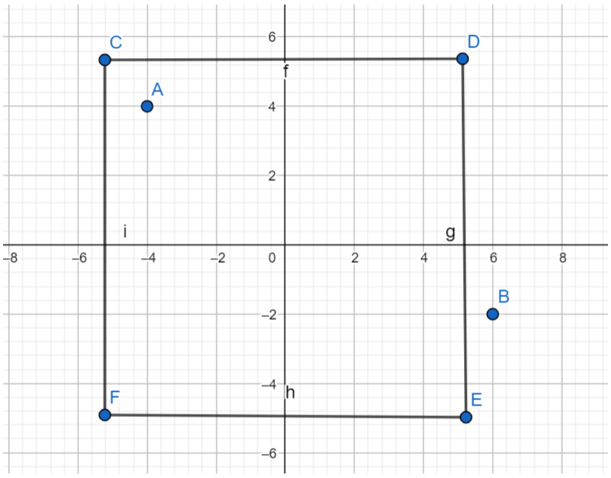
We will note $(H^2_{rec_{co}})$ as the class of axis aligned rectangles with the center in origin $O(0, 0)$.

1) There exists a set C of 2 points that is shattered by $H^2_{rec_{co}}$ (this will mean that $H^2_{rec_{co}} \geq 2d$).

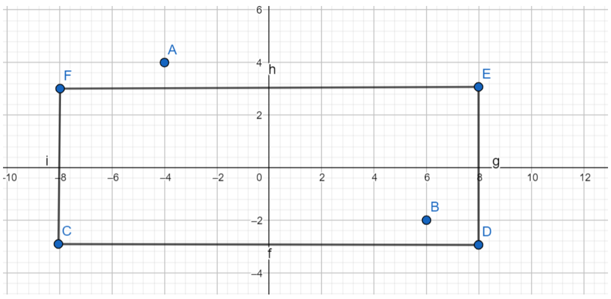
We will choose A(-4, 4) and B(-2, 6) from R^2 We will show that any labeling of these two points is valid by an axis aligned rectangle with the center in O(0, 0).



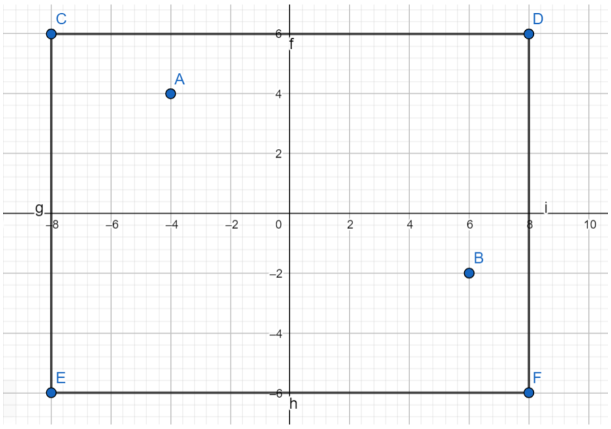
Label (A-0, B-0)



Label (A-1, B-0)



Label (A-0, B-1)



Label (A-1, B-1)

So, the given set of points is labelled by the class of axis aligned rectangles with the center in the origin.

2) Every set C of 3 points is not shattered by $\mathcal{H}_{rec}^{d}$ (this will mean that $VC\dim(\mathcal{H}_{rec}^{d}) < 2d+1$).

A rectangle with the center in the origin $O(0,0)$ has the four points : $A1(-x,y), A2(-x,-y), A3(x,-y), A4(x,y)$. So, it marks with label 1 any point which has coordinates (a,b), where $|a| <= |x|$ and $|b| <= |y|$

Let $A = (x1,y1), B = (x2,y2), C = (x3,y3)$, where $A,B,C \in \mathcal{X} = R^2$, considering $|x1| <= |x2| <= |x3|$.

Let's considering all the possibilities for the ordering of $y1, y2, y3$.

i) $|y1| <= |y2| <= |y3|$, in which the label (0, 1, 0) can't be assigned by any rectangle (A2 can't be labelled as 1 without marking the A1 as well, as they have smaller x and y coordinates).

ii) $|y1| <= |y3| <= |y2|$, in which the label (0, 1, 0) can't be assigned by any rectangle (A2 can't be labelled as 1 without marking A1 as well).

iii) $|y2| <= |y1| <= |y3|$, in which the label (0, 0, 1) can't be assigned by any rectangle (A3 can't be labelled as 1 without marking A1 and A2).

iv) $|y2| <= |y3| <= |y1|$, in which the label (0, 0, 1) can't be assigned by any rectangle (A3 can't be labelled as 1 without marking A2 as well).

v) $|y3| <= |y1| <= |y2|$, in which the label (0, 1, 0) can't be assigned by any rectangle (A2 can't be labelled as 1 without marking A1 as well).

vi) $|y3| <= |y2| <= |y1|$, in which the label (1, 0, 1) can't be assigned by any rectangle (in order to assign A1 and A3, the width and height of the rectangle must be bigger than $|x3|$ and $|y1|$ and then it labels A2 as well).

So, there can't be any set of 3 points shattered by $H_{rec_{co}}^{2}$, so $VCdim(\mathcal{H}) < 3$.

Form 1) and 2), $VCdim(\mathcal{H}) = 2$

4. **(1 point)** Let $\mathcal{X} = \mathbb{R}^2$ 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 $= \alpha$ (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.

Solution

Informally, H is the set of similar triangles obtained from the base triangle with the points $A = (0, 0), B = (\alpha, 0), C = (0, 1)$ by doing the following operations: rescaling and translation.

Algorithm A: For each point $P \in S$ that is labelled as 1, generate the parallel lines through it to the sides of triangle ABC. Iterating through all the points, we obtain 3 sets of lines:

LAB = lines for which $\exists P \in S$ with label(P) = 1, $P \in d, d||AB$

LAC = lines for which $\exists P \in S$ with label(P) = 1, $P \in d, d||AC$

LBC = lines for which $\exists P \in S$ with label(P) = 1, $P \in d, d||BC$

From LAB choose $d1$ that has smallest x. From LAC choose $d2$ that has the smallest y. From LAC choose $d3$ that has the biggest x.

Our algorithm will return the triangle that is at the intersection of the $d1, d2, d3$, which is similar to the ΔXYZ (because $d1$, $d2$, $d3$ are parallel with the lines of the ΔXYZ, it follows that this triangle is indeed in $\mathcal{H}$).

If there are no points labelled as 1, return any triangle that does not contain any points from S. Obviously, $L_S(h_S) = 0$.

We need to prove now that $\mathcal{H}$ is PAC-learnable, doing the following:

For the following steps, we fix the $\epsilon, \delta > 0$.

In our case, all the positive points must be inside ΔABC, while the others outside..,

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

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

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

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

Consider the realizability assumption.

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

Solution

Initial consideration. If we pick $\theta_1 \leq \theta_2$, then every element is labelled as 1.

We first prove that $VCdim(\mathcal{H}) \geq 2$.

Let $\{x_1, x_2\} \in R^2$, where $x_1 < x_2$

$H(x1) = H(x2) = 0$, pick $\theta_1 = x_2 + 1, \theta_2 = x_1 - 1$.

$H(x1) = H(x2) = 1$, pick $\theta_1 = \theta_2 = x_1$.

$H(x1) = 0, H(x2) = 1$, pick $\theta_1 = x_2, \theta_2 = x_1 - 1$.

$H(x1) = 1, H(x2) = 0$, pick $\theta_1 = x_2 + 1, \theta_2 = x_1$
$\mathcal{H}$ shatters any set $\{x_1, x_2\} \in R^2$. So, $VCdim(\mathcal{H}) \geq 2$.

Now, we will prove that $VCdim(\mathcal{H}) < 3$

Let's consider $x_1, x_2, x_3 \in \mathcal{R})$, where $x_1 < x_2 < x_3$ and the following labelling $\mathcal{H}(x_1) = 0$, $\mathcal{H}(x_2) = 1$, $\mathcal{H}(x_3) = 0$

if $x_2 \in \mathcal{H}_1$ and $x_3 > x_2$, then x_3 will be labelled by $\mathcal{H}_1$ as 1.

if $x_2 \in \mathcal{H}_2$ and $x_1 < x_2$, then x_1 will be labelled by $\mathcal{H}_2$ as 1.

if $x_2 \in \mathcal{H}_3$, then $= \theta_1 \leq \theta_2$, then any $x \in \mathcal{R}$ will be labelled as 1, so both x_1 and x_3 will be mislabelled.

This proves that $VCdim(\mathcal{H}) < 3$. Having proved previously that $VCdim(\mathcal{H}) >= 2$
Then, prove that $VCdim(\mathcal{H}) = 2$.

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

Solution

As is it shown in *Lecture 9*, any function which has a finite $VCdim$ is also *PAC-learnable*, within The Fundamental Theorem of Statistical Learning Theory.

[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.

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.

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 $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ 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