

MALTA 2021–2022: Assignment

Provide your solution before the start of next lecture on October 12th.
If you are unable to attend next lecture, then send your solution by email
to giovanni.neglia@inria.fr.

Motivate your answers.

Ex. 1 — Show that a finite class has finite VC-dimension.

Ex. 2 — Prove that in the definition of agnostic PAC learnability, when the loss is the 0-1 loss function and the learning algorithm is required to output a hypothesis in H ($A(S) \in H$), it is possible to replace the probabilistic condition

$$\forall \epsilon, \delta \in (0, 1), \exists m_H(\epsilon, \delta) \in \mathbb{N}, \text{ such that } \forall m \geq m_H(\epsilon, \delta),$$

$$L_D(A(S)) - \min_{h \in H} L_D(h) \leq \epsilon, \text{ with prob. larger than } 1 - \delta.$$

with the following condition on the expected value

$$\lim_{m \rightarrow \infty} \mathbb{E}_{S \in D^m} \left[L_D(A(S)) - \min_{h \in H} L_D(h) \right] = 0$$

Hints: We have already mapped during the lecture a probabilistic condition to a condition on expected values. The Markov inequality may help you.

Ex. 3 — If you show that H cannot shatter any set of size n , do you need to check if it can shatter a set of size $n' > n$? Why?

Ex. 4 — Consider the class H_k of binary functions over \mathbb{R} which assume value 1 exactly on k points, i.e.,

$$H_k = \left\{ h : \mathbb{R} \rightarrow \{0, 1\}, \text{ such that } \exists k \text{ distinct values } x_1, x_2, \dots, x_k \in \mathbb{R}, \right. \\ \left. \text{such that } h(x) = 1 \text{ if } x \in \{x_1, x_2, \dots, x_k\} \text{ and } h(x) = 0 \text{ otherwise} \right\}.$$

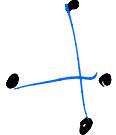
What is the VC-dimension of H ?

Ex. 5 — Consider the class of squares in the plane

$$H = \{h_{a,b,r} : \mathbb{R}^2 \rightarrow \{0, 1\}, \text{ for some } a, b \in \mathbb{R}, \text{ and } r \in \mathbb{R}^+\},$$

where

$$h_{a,b,r}(x_1, x_2) = \begin{cases} 1, & \text{if } x_1 \in [a, a+r] \text{ and } x_2 \in [b, b+r], \\ 0, & \text{otherwise.} \end{cases}$$



1. What is the VC-dimension of H ?

Hint: The following geometric result may help: given any set A of 4 points in the plane, it is always possible to split A in two disjoint sets B and C ($B \cup C = A$, $B \cap C = \emptyset$) such that $\mathcal{CH}(B) \cap \mathcal{CH}(C) \neq \emptyset$, where $\mathcal{CH}(S)$ denotes the convex hull of the set S .

2. What if r is fixed to 1?

Ex. 6 — Consider a binary classification problem over \mathbb{R} and the 0-1 loss function. Let \mathcal{P} denote the class of all polynomials over \mathbb{R} . We consider the following hypothesis class:

$$H = \{h(x) = g(p(x)), \text{ for some } p \in \mathcal{P}\},$$

where $g : \mathbb{R} \rightarrow \{0, 1\}$ and $g(x) = 1$ if and only if $x > 0$.

Is the class H PAC learnable?

(EX.1) H s.t. $|H|$ is finite $\Rightarrow H$ is PAC LEARNABLE $\Rightarrow \text{VCdim}(H) < +\infty$

(EX.2) Give, from APAC definition:

$$L_D(A(s)) - \min_{h \in H} L_D(h) \leq \varepsilon \text{ w.p. } \geq 1 - \delta$$

for the ε -REPRESENTATIVENESS property:

$$|L_S(h) - L_D(h)| \leq \varepsilon, \forall h \in H$$

we can so fix h^* s.t. we have $\min_{h \in H} L_D(h)$

we can write it as:

$$\left| \frac{1}{m} \sum_{i=1}^m I(h^*, (x_i, y_i)) - \mathbb{E}_{(x,y) \sim D} [I(h^*, (x, y))] \right| \leq \varepsilon$$

because we have:

- an arbitrary variable $I(h^*, (x, y))$
- their expected value $L_D(h^*) = \mathbb{E}_{(x,y) \sim D} [I(h^*, (x, y))]$
- the estimation of the medium value $L_S(h) = \frac{1}{m} \sum_{i=1}^m I(h^*, (x_i, y_i))$, then
- and it's variance from the expectation ε

from the "WEAK LAW OF GREAT NUMBER" for m that goes to ∞

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\left| \frac{1}{m} \sum_{i=1}^m I(h^*, (x_i, y_i)) - \mathbb{E}_{(x,y) \sim D} [I(h^*, (x, y))] \right| \leq \varepsilon \right) = 1 \Rightarrow$$

$$\Rightarrow \lim_{m \rightarrow \infty} \mathbb{P} (|L_S(h^*) - L_D(h^*)| \leq \varepsilon) = 1$$

- So if with m that goes to ∞ S is always ε -representative, we can expect that the difference between the emp

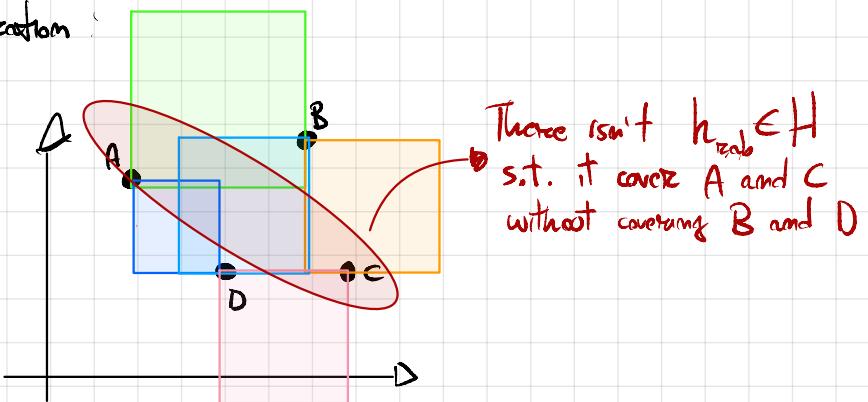
$$\lim_{m \rightarrow \infty} \mathbb{E}_{S \sim D^m} \left[L_D(A(s)) - \min_{h \in H} L_D(h) \right] = 0$$

(EX.3) Given $C \subset X$ of size m , if H doesn't shatter C , H cannot shatter C' such that $C \subset C'$ of size $m' > m$ because in order to shatter a set, H must shatter all subset of it.

(EX.4) The VC dimension of H_k is k because for a set of $k+1$ elements you can't find a function $h \in H_k$ that set all elements to 1.

(EX.5) The $\text{VCdim}(H) = 3$ because with regular squares and 4 points H can't shatter all the 2 points configuration:

$$\begin{aligned}\{A, B\} &\rightarrow h \\ \{A, D\} &\rightarrow h \\ \{B, C\} &\rightarrow h \\ \{B, D\} &\rightarrow h \\ \{C, D\} &\rightarrow h \\ \{A, C\} &\rightarrow \cancel{\exists h \in H}\end{aligned}$$

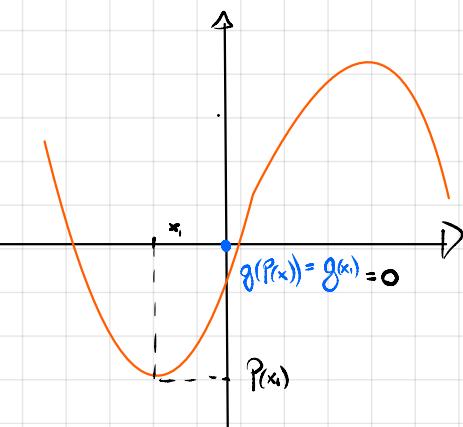
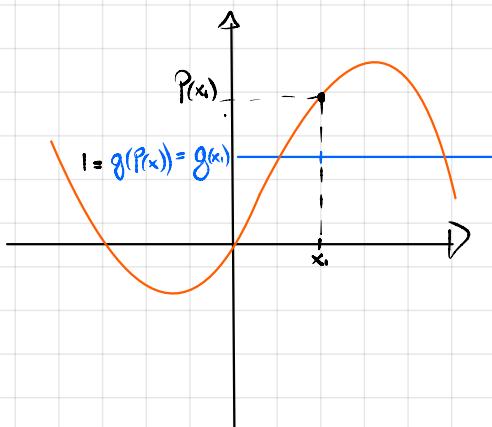


(EX.6) The class H is NOT PAC learnable.

- Let P the class of all polynomials over \mathbb{R}
- Let $g: \mathbb{R} \rightarrow \{0, 1\} \mid g(x) = 1 \Leftrightarrow x \geq 0$
- Let $H = \{h(x) = g(p(x)), p \in P\}$
- For every set of points in \mathbb{R} H can shatter them
- For a single point $x_i \in \mathbb{R}$ we have 2 cases:

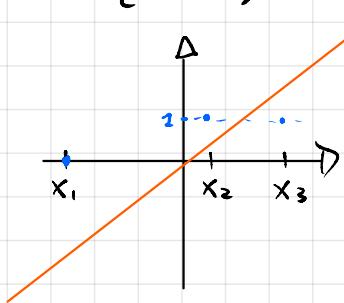
$\forall x_i > 0$ we can choose $p \in P \mid p(x) > 0$ so $h(x) = g(p(x)) = g(x) = 1$

$\forall x_i < 0$ we can choose $p \in P \mid p(x) < 0$ so $h(x) = g(p(x)) = g(x) = 0$

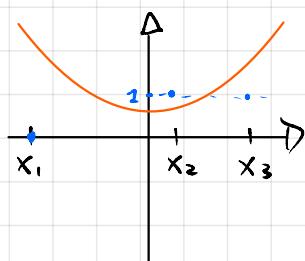


- given any set of point $X = \{x_1, \dots, x_m\}$ we can easily find $\text{P}(\mathcal{P})$ such that $\exists h$ matching any given $\{0,1\}^m$ combination associated with X .
- for example with 3 points

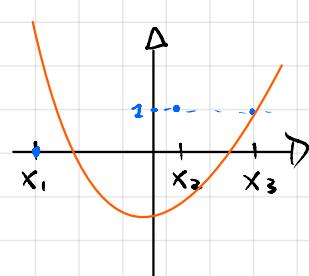
$$\{1,1,1\}$$



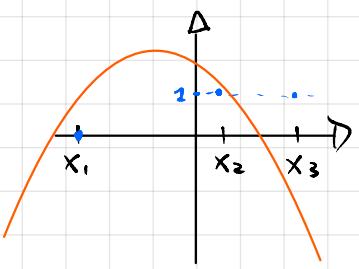
$$\{0,1,1\}$$



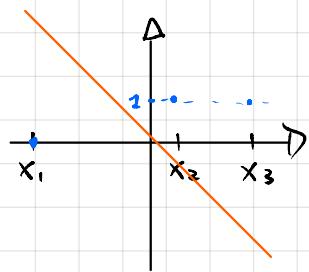
$$\{0,0,1\}$$



$$\{0,1,0\}$$



$$\{0,0,0\}$$



- In conclusion $\text{VCdim}(h) = +\infty$