MALTA 2024–2025: Assignment

The points marked for each exercise give an indication of their relative importance.

Provide your solution before the start of next lecture on October 18th. If you are unable to attend next lecture, then send your solution by email to giovanni.neglia@inria.fr.

Motivate your answers.

Ex. 1 — (2 points) Show that a finite class has finite VC-dimension.

Ex. 2 — (2 points) 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. 3 — (3 points) Consider the class H_k of binary functions over $\mathbb{R}$ which assume value 1 exactly on k points, i.e.,

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

What is the VC-dimension of H?

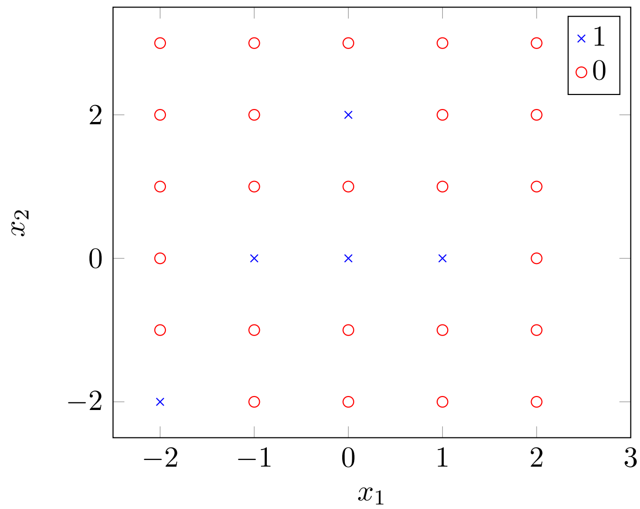
Ex. 4 — (4 points) Consider the 0–1 loss and the class of equilateral triangles in the plane with a side parallel to the first coordinate and sides' length equal to s and:

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

where

$$h_{a,b,r}(x_1, x_2) = \begin{cases} 1, & \text{if } x_2 \geq b \text{ and } x_2 \leq b + \sqrt{3}(x_1 - a) \text{ and } x_2 \leq b + \sqrt{3}s - \sqrt{3}(x_1 - a), \\ 0, & \text{otherwise.} \end{cases}$$

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. Consider the dataset of 30 points in the figure below (all samples have integer coordinates). Find an ERM predictor $h_{a,b,s}$ (specify its parameters). What is its empirical loss? How could you bound its expected loss?



Ex. 5 — (5 points) Consider a binary classification problem over $\mathbb{R}$ and the 0–1 loss function. Let $\mathcal{P}_n = \{\sum_{i=0}^{n} a_i x^i, a_1, a_2, \ldots, a_n \in \mathbb{R}\}$ denote the class of all polynomials over $\mathbb{R}$ with degree at most n. We consider the following hypothesis class:

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

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

1. Is the class H PAC learnable?
2. Is the class H efficiently PAC-learnable?
3. Is the class H efficiently agnostic PAC-learnable? (a discussion about the problem is sufficient)

Ex. 6 — (2 points) You have a dataset with m samples. Samples contain real input variables x_1, x_2, x_3 and a real output variable y. You would like to learn a predictor of the form $y = \theta_1 x_1 + \theta_2 \sin(x_2) + \theta_3 \cos(x_3)$ with θ_1, θ_2, and θ_3 real variables. You adopt as loss the squared loss.

1. How would you learn the predictor?