Assignment 1
Advanced Machine Learning
Issa Bilal (AI - 407)
2023

Exercise 1: Give examples for:

a) a finite hypothesis class H with VCdim(H) = 2023. Justify your choice.

Solution: we define an H hypothesis class as follows:

H = $\{h_{a_1,b_1,\dots,a_{1011},b_{1011},s}$(x): $\mathbb{R} \longrightarrow \{-1,1\} \mid a_i < b_i \leq 2022,\ a_i, b_i \in \mathbb{N}$, i = $\overline{1,\ 1011}$, s $\in \{-1,1\}$, $h_{a_1,b_1,\dots,a_{1011},b_{1011},s}$(x) = {s, x$\in[a_i,b_i]$ | -s, otherwise}

Where we define a and b to take values from the set of natural numbers $\mathbb{N}$ and give the labeling 1 if a point falls within an interval and 0 if it is outside. The S parameter acts as a label switching. Therefore the algorithm A shatters any given 2022 points plus the S parameter which according to the solution from seminar 3, exercise 4 we can prove that VCdim(H) = 2023.

b) an infinite hypothesis class H with VCdim(H) = 2023. Justify your choice.

H = $\{h_{a_1,b_1,\dots,a_{1011},b_{1011},s}$(x): $\mathbb{R} \longrightarrow \{-1,1\} \mid a_i < b_i,\ a_i, b_i \in \mathbb{N}$, i = $\overline{1,\ 1011}$, s $\in \{-1,1\}$, $h_{a_1,b_1,\dots,a_{1011},b_{1011},s}$(x) = {s, if x$\in[a_i,b_i]$ | -s, otherwise}}

Solution: the solution for this one is very similar to the previous point with the difference being that we define a hypothesis class H_{rec} where a and b take values in $\mathbb{R}$ instead of $\mathbb{N}$ (as shown in seminar 3 exercise 2). Which turns the hypothesis class itself to infinite.

c) an infinite hypothesis class H with VCdim(H) = ∞. Justify your choice.

Solution: one example of an infinite hypothesis class is H_{sin} which is proven in lecture 7.

$$H_{sin} = \{h_\theta : \mathbb{R} \longrightarrow \{0,1\} \mid h_\theta = [\sin(\theta x)],\ \theta \in \mathbb{R}\}$$

Exercise 2:

a) Solution: we take the $H_{thresholds}$ example from lecture 5 where x is the norm of a 3D point and $H_{thresholds}$ is the indicator function for H:

Therefore we get $|H_{thresholds}| = \infty$.

Let a* $\in$ [0,∞] such that h_{a*}(x) = $1_{[||x||_2 \leq a*]}$(x) achieves L(h_{a*}) = 0 (realizability assumption in the PAC learning scenario)

Consider a distribution D over $\mathbb{R}^3$ and take a_0 < a* $\in$ [0,∞] such that:

$P_{x\sim D}$(x $\in$ [a_0,a*]) = ε

If D_x(-∞,a*) ≤ ε take a_0 = -∞

Consider the training set S = {(x_1,y_1), (x_2,y_2), ..., (x_m,y_m)} and the following algorithm A:

- Take b = max{x_i|(x_i, 1) $\in$ S} (if ∀y_i = 0, b = -∞)
- Output A(S) = h_s = h_b

Then $L_D(h_s)$ > ε ⇒ b < a_0. We have that:

$P_{S\sim D^m}$ ($L_D(h_s)$ > ε) = P(b < a_0) = $(1 - \varepsilon)^m \leq e^{-\varepsilon m}$

Take m = [$\frac{1}{\varepsilon}\log\frac{1}{\delta}$]

So, we have that: $P_{S\sim D^m}$ ($L_D(h_s)$ > ε) < δ

Which shows that our hypothesis class is PAC learnable.

Exercise 3:

Solution: to get the VCdim(H) we have to prove that the given hypothesis class is included in H_{lines}.

Because we know that both the sin and cos take values in [-1,1], and θ takes values in ℝ. therefore we have a linear inequality which can be written as $x_1a + x_2b$ + c > 0, where a,b replace the sin and cos, and c replaces the θ_1.

Which in turn means that our hypothesis class H ⊆ H_{lines}.

Given the demonstration in Lecture 6, it results that our hypothesis class has a VCdim(H) ≤ 3.

To prove that the VCdim(H) is strictly equal to 3 we can apply the norm on the entire inequality in $\mathbb{R}^2$ to obtain the following equation:

$$\frac{a}{\sqrt{a^2+b^2}}x_1 + \frac{b}{\sqrt{a^2+b^2}}x_2 + \frac{c}{\sqrt{a^2+b^2}} = 0$$

After the nomalization of the vector the direction will be preserved but $\frac{c}{\sqrt{a^2+b^2}}$ will still take values in $\mathbb{R}$. which means that our hypothesis class H is actually H_{lines}. Which in turn results that the VCdim(H) = 3. The rest of the proof can be found in lecture 6.

Exercise 4:

a) Similarly to the seminar 1, exercise 2, we first need to find an algorithm A that returns a hypothesis h $\in H_\alpha$ consistent with the training set

S = $\{(x_1,y_1,.....,x_m,y_m)\} \in \mathbb{R}$.

Similar to the problem in the first seminar we build a classifying rectangle for the positive labels (y_i = 1):

$a_{1s} = min_{i=\overline{1,m}}\, x_{i1}$

$a_{2s} = min_{i=\overline{1,m}}\, x_{i2}$

$b_{1s} = max_{i=\overline{1,m}}\, x_{i1}$

$b_{2s} = max_{i=\overline{1,m}}\, x_{i2}$

In the edge case where there are no positive samples there are two approaches:
1- we pick a point outside of the training set that maintains the aspect ratio of the rectangle or simply choose a random point outside of the data set that has enough space to maintain the aspect ratio.

Now we fix two of the coordinates of an axis be it (a_{1s},a_{2s}) or (b_{1s},b_{2s}), and expand the other two coordinate points to obtain the desired aspect ratio α, then we centralize the obtained rectangle inside of R* such that it doesn't exceed it's borders. Therefore By construction, A is an ERM,
meaning that the loss $L_{h*,D}(h_s)$ = 0.