Advanced Machine Learning - Assignment 1

Bouruc Petru-Liviu

26/04/2023

1 Exercise 1

a) An example of finite hypothesis class $\mathcal{H}$ that has VCdim($\mathcal{H}$) = 2023 could be the set of monotone boolean conjunctions over $\{0,1\}^{2023}$

$$\mathcal{H}_{mcon}^{2023} = \{h : \{0,1\}^{2023} \to \{0,1\}, h(x_1, x_2, ..., x_{2023}) = \bigwedge_{i=\overline{1,2023}} l(x_i)\} \cup \{h^-\},\ l \in \{x_i, 1\}$$

This is class is a particular case for $\mathcal{H}_{con}^d$ (which we know is finite), where we don't take into consideration the negations, and augment with all negative and positive hypotheses, so our class is still *finite*. It was proven in seminar 3, exercise 3.5 that VCdim($\mathcal{H}_{mcom}^d$) = d, so, in this particular case, we can choose d = 2023.

b) An example of infinite hypothesis class $\mathcal{H}$ that has VCdim($\mathcal{H}$) = 2023 could be the set of linear classifiers $\mathcal{H}S_0^{2023}$

$$\mathcal{H}S_0^{2023} = \{h_{w,0} : \mathbb{R}^{2023} \to \{-1,1\}, h_{w,0}(\mathbf{x}) = sign\left(\sum_{i=1}^{2023} w_i x_i\right) \mid w \in \mathbb{R}^{2023}\}$$

From it's base structure, we can observe that $|\mathcal{H}S_0^{2023}| = \infty$ because it is defined over $\mathbb{R}^{2023}$. It was proved in *Lecture* 7 that VCdim($\mathcal{H}S_0^n$) $= n$, so, in this particular case, we can choose $n = 2023$.

1. There exists a set A of size 2023 that is shattered by $\mathcal{H}$ (VCdim($\mathcal{H}$) $\geq$ 2023)

Let A be a set of 2023 points, $A = \{e_0, e_1, ..., e_{2023}\} \in \mathbb{R}^{2023}$. For every B $\subseteq$ A, there is a function $h_b \in \mathcal{H}S_0^{2023}$ such that h_B gives label +1 to all elements in B and -1 to all elements in A-B. We can choose w = $(w_1, w_2, w_3, ..., w_{2023})$ as:

$$w_i = \begin{cases} 1, & e_i \in B \\ -1, & e_i \notin B \end{cases}$$

Then, $h_B = \text{sign}(< w, e_i >) = w_i$ will generate label +1 for elements in B, and -1 for elements not in B. This means that there exists

2. Every set A of size 2024 is not shattered by $\mathcal{H}$ (VCdim($\mathcal{H}$) < 2024)

From 1 and 2, we can conclude that VCdim($\mathcal{H}S_0^{2023}$) = 2023.

c) An example of infinite hypothesis class $\mathcal{H}$ that has VCdim($\mathcal{H}$) $= \infty$ could be the set of sin functions $\mathcal{H}_{sin}$

$$\mathcal{H}_{sin} = \{h_\theta : \mathbb{R} \to \{0,1\} \mid h_\theta(\mathbf{x}) = \lceil sin(\theta x) \rceil, \theta \in \mathbb{R}\}, \lceil -1 \rceil = 0$$

From it's structure, we can observe that $|\mathcal{H}_{sin}| = \infty$ because it is defined over $\mathbb{R}$. It was proved in *Lecture* 7 that VCdim($\mathcal{H}_{sin}$) $= \infty$.

2 Exercise 2

a) At a fisrt look, we can see that our class is the class of spheres of radius a. Let's consider the realizability assumption: there exists a labeling function $f \in \mathcal{H}, f = h_{a^*}$ with $L(h_{a^*}) = 0$; let $a^* \in \mathbb{R}$ such that there exists $h_{a^*}(x) = 1_{\|x\|_2 \leq a^*}$.

Consider the training set $S = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m) \mid x_i \in \mathbb{R}^3, y_i = h_{a^*}(x_i)\}$. We have to construct an algorithm to this training set which has the loss equal to 0, in order to be an ERM and to demonstrate that our hypothesis class is (ϵ, δ)-PAC learnable using ERM rule.

Applying an algorithm similar to that from *lecture* 5, we can see that we want to find the smallest sphere that encloses all the positive labels.

Let's consider the algorithm A:

- construct $S_n = \{(x_1^n, y1), ..., (x_m^n, y_m) \mid x_i^n = \|x_i\|_2, (x_i, y_i) \in S\}$
- therefore, after transforming our initial set in S_n, now we can apply the algorithm A from $\mathcal{H}_{thresholds}$ (lecture 5)
- A(S_n) $= h_b$, where $h_b = max\{x_i \mid (x_i, 1) \in S_n\}$ (if there are no positive examples in S_n we take $b = -\infty$)

According to the lemma from *Lecture* 5, $\mathcal{H}_{thresholds}$ is (ϵ, δ)-PAC learnable, with the sample complexity of $\lceil \frac{1}{\epsilon} log \frac{1}{\delta} \rceil$. Norm function is a surjective function, so we can conclude that $\mathcal{H}$ is a particular, but similar case for $\mathcal{H}_{thresholds}$, so it is (ϵ, δ)-PAC learnable, with the sample complexity of $m_{\mathcal{H}}(\epsilon, \delta) = \lceil \frac{1}{\epsilon} log \frac{1}{\delta} \rceil$.

b) In *Lecture* 6, it was proven that VCdim($\mathcal{H}_{thresholds}$) $= 1$, so we can conclude that our hypothesis class has $VCdim(\mathcal{H}) \leq 1$. In order to show that $VCdim(\mathcal{H}) \geq 1$, we can take for example one point A(1,1,1) that gives the combinations of labels (0 and 1) for different parameter a (for example 0 and 100). In conclusion, $VCdim(\mathcal{H}) = 1$.

3 Exercise 3

For this exercise, we can see that our hypothesis class is a particular case for the hypothesis class $\mathcal{H}_{lines}$, in which we set $a = sin(\theta_2), b = sin(\theta_2), c = \theta_1, a, b \in [-1, 1]$. So, the first assumption is that $\mathcal{H} \subseteq \mathcal{H}_{lines}$, thus $VCdim(\mathcal{H}) \leq VCdim(\mathcal{H}_{lines})$. In *lecture* 6 it was shown that VCdim($\mathcal{H}_{lines}$) $= 3$. (1)

Now we want to prove that $VCdim(\mathcal{H}) \geq 3$. In order to do this, as indicated in *lecture* 7, we will take a set A of 3 non-colinnear points in $\mathbb{R}^3$ and try to demonstrate that $\mathcal{H}$ shatters that set A. As an example, we can take $\mathcal{C} = \{(0, -4), (2, 1), (6, -2)\}$

- for label (0,0,0): $\theta_1 = -4$ and $\theta_2 = -3$
- for label (1,0,0): $\theta_1 = -3.5$ and $\theta_2 = -3$
- for label (0,1,0): $\theta_1 = -1.5$ and $\theta_2 = 0.5$
- for label (0,0,1): $\theta_1 = -4$ and $\theta_2 = 1.5$
- for label (1,1,0): $\theta_1 = 2$ and $\theta_2 = -1.5$
- for label (1,0,1): $\theta_1 = -3$ and $\theta_2 = 2.5$
- for label (0,1,1): $\theta_1 = -2$ and $\theta_2 = 1$
- for label (1,1,1): $\theta_1 = -1$ and $\theta_2 = 2$

So, we can see that that we can obtain every combination of labels, thus $\mathcal{C}$ shatters $\mathcal{H}$ and $VCdim(\mathcal{H}) \geq 3$. (2)

From (1) and (2), we can conclude that VCdim($\mathcal{H}$) = 3.