

AML Assignment 1

Suto Robert-Lucian (411)

1 Exercise 1.

1.1 a) a finite hypothesis class H with $VCdim(H) = 2023$.

According to the third Seminar, page 3 to 6, the class of Boolean conjunctions over the variables $x_1, x_2, \dots, x_d, d \geq 2$

$$\mathcal{H}_{\text{con}}^d = \left\{ h : \{0, 1\}^d \rightarrow \{0, 1\}, h(x_1, x_2, \dots, x_d) = \bigwedge_{i=1, d} l(x_i) \right\}$$
$$l(x_i) = \text{literal of variable } x_i$$
$$l(x_i) \in \left\{ x_i, \overline{x_i}, \underset{\text{missing}}{1} \right\}$$

Has the VC dimension equal to d . This is proven in exercises 3.3 where we showed that

$$h_{\mathcal{J}} = \bigwedge_{j \notin \mathcal{J}} \overline{x_j} = \bigwedge_{j \in \{1, \dots, d\} \setminus \mathcal{J}} \overline{x_j}$$

Can shatter $\mathcal{J} = \{1, 2, 4\}$, define $h_{\mathcal{J}} = \overline{x_3} \wedge \overline{x_5} \wedge \overline{x_6}$ ($d = 6$).

$$\begin{aligned} h_{\mathcal{J}}(e_1) &= h_{\mathcal{J}}(1, 0, 0, 0, 0, 0) = \overline{0} \wedge \overline{0} \wedge \overline{0} = 1 \\ h_{\mathcal{J}}(e_2) &= h_{\mathcal{J}}(0, 1, 0, 0, 0, 0) = \overline{0} \wedge \overline{0} \wedge \overline{0} = 1 \\ h_{\mathcal{J}}(e_4) &= h_{\mathcal{J}}(0, 0, 0, 1, 0, 0) = \overline{0} \wedge \overline{0} \wedge \overline{0} = 1 \\ h_{\mathcal{J}}(e_3) &= h_{\mathcal{J}}(0, 0, 1, 0, 0, 0) = \overline{1} \wedge \overline{0} \wedge \overline{0} = 0 \\ h_{\mathcal{J}}(e_5) &= h_{\mathcal{J}}(0, 0, 0, 0, 1, 0) = \overline{0} \wedge \overline{1} \wedge \overline{0} = 0 \\ h_{\mathcal{J}}(e_6) &= h_{\mathcal{J}}(0, 0, 0, 0, 0, 1) = \overline{0} \wedge \overline{0} \wedge \overline{1} = 0 \end{aligned}$$

So, $h_{\mathcal{J}}(e_j) = 1$ if $j \in \mathcal{J}$ and $h_{\mathcal{J}}(e_j) = 0$ if $j \notin \mathcal{J}$. This proves that $\mathcal{H}_{\text{con}}^d$ shatters $C \Rightarrow VC \dim(\mathcal{H}_{\text{con}}^d) \geq d$. The other implication (That $VC \dim(\mathcal{H}_{\text{con}}^d) \leq d$), is proven in 3.4 by contradiction.

This class is also finite because it includes all possible boolean conjunctions over d variables, with the total number of possible conjunctions being finite. If d is known and equal to 2023, we can conclude that the VC dimension of $\mathcal{H}_{\text{con}}^{2023} = 2023$.

1.2 b) an infinite hypothesis class \mathbf{H} with $VCdim(\mathbf{H}) = 2023$.

According to Lecture 7, page 21 to 28, the linear classifiers that go through origin, noted as $(\mathcal{H}S_0^n)$

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

has the VCdim of n and is also infinite, w being able to take any real value. For this particular case, we can consider that

$$VCdim(\mathcal{H}S_0^{2023}) = 2023$$

1.3 c) an infinite hypothesis class \mathbf{H} with $VCdim(\mathbf{H}) = \infty$.

According to Lecture 7, page 10 to 20, the set of sin functions

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

has the VCdim equal to ∞ and is also infinite.

2 Exercise 2.

Consider \mathcal{H} to be the following hypothesis class:

$$\mathcal{H} = \left\{ h_a : \mathbb{R}^3 \rightarrow \{0, 1\} \mid h_a(\mathbf{x}) = \mathbf{1}_{[\|\mathbf{x}\|_2 \leq a]}(\mathbf{x}), \mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3, \|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2} \right\}$$

Consider the realizability assumption.

2.1 a) show that \mathcal{H} can be (ϵ, δ) -PAC learned by giving an algorithm A and determining the sample complexity $m_{\mathcal{H}}(\epsilon, \delta)$ such that the definition of PAC learnability is satisfied.

We can rewrite all h_a in \mathcal{H} as follows:

$$h_a(\mathbf{x}) = \begin{cases} 1 & \text{if } \|\mathbf{x}\|_2 \leq a \\ 0 & \text{if } \|\mathbf{x}\|_2 > a \end{cases}$$

Since $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}$ denotes a type of Euclidean norm the condition $\|x\|_2 \leq a$ is equivalent to $|\mathbf{x}| \leq a$ and that the condition $\|x\|_2 > a$ is equivalent to $|\mathbf{x}| > a$. Therefore, we can rewrite h_a as:

$$h_a(\mathbf{x}) = \begin{cases} 1 & \text{if } |\mathbf{x}| \leq a \\ 0 & \text{if } |\mathbf{x}| > a \end{cases}$$

This is a standard form of a threshold function: $h_a(\mathbf{x})$ is 1 if the input $|\mathbf{x}|$ is less than or equal to the threshold a , and 0 otherwise. From knowing the realizability assumption, let $a^* \in \mathbf{R}$ such that h_{a^*} labels all points correctly. Considering that \mathcal{H} is a family of threshold functions we can refer to the ERM algorithm presented in Lecture 5.

- Given a training set $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ of m labeled examples, where $x_i \in \mathbb{R}^3$ and $y_i \in \{0, 1\}$ for all $i \in [m]$. (Also apply the norm over all of the examples)
- take $b = \max\{x_i \mid (x_i, 1)\}$ (if no positive example appears in S , take $b = -\infty$)
- output $A(S) = h_b$

Knowing that \mathcal{H} belongs to the $\mathcal{H}_{thresholds}$ family, we can assume that \mathcal{H} can be (ϵ, δ) -PAC learned using the ERM learning rule with a sample complexity of $m = \lceil \frac{1}{\epsilon} \log \frac{1}{\delta} \rceil$ (Lecture 5 page 30)

2.2 b) compute $VCdim(\mathcal{H})$.

Considering that \mathcal{H} belongs to the family of $\mathcal{H}_{thresholds}$, we can use the proof from Lecture 6 to support that $VCdim(\mathcal{H}) \leq \mathcal{H}_{lines} = 1$. To show that the $VCdim(\mathcal{H}) \geq 1$, We can simply take a set consisted of one point in \mathbb{R}^3 for example

$p = (0, 1, 1)$. To obtain label 0, we can chose $a = 0$, and for label 1, we can chose $a = 2$. Thus $\text{VCdim}(\mathcal{H}) \geq 1$, making it equal to 1.

3 Exercise 3.

Compute $\text{VCdim}(\mathcal{H})$ where \mathcal{H} is the following hypothesis class:

$$\mathcal{H} = \{h_{\theta_1, \theta_2} : \mathbb{R}^2 \rightarrow \{0, 1\} \mid h_{\theta_1, \theta_2}(\mathbf{x}) = h_{\theta_1, \theta_2}(x_1, x_2) = \mathbf{1}_{[\theta_1 + x_1 \cdot \sin(\theta_2) + x_2 \cdot \cos(\theta_2) > 0]}, \theta_1, \theta_2 \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^2\}$$

Comparing h_{θ_1, θ_2} with the function from \mathcal{H}_{lines} (Lecture 6, page 28-29)

$$\mathcal{H}_{lines} = \{h_{a,b,c} : \mathbb{R}^2 \rightarrow \{0, 1\} \mid h_{a,b,c}(\mathbf{x}) = \mathbf{1}_{[ax+by+c>0]}, a, b, c \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^2\}$$

we can see that h_{θ_1, θ_2} can be represented in the form of $h_{a,b,c}$.

Knowing that both sine and cosine are both bounded to $[-1, 1]$, we can rewrite $\cos(\theta_2)$ as a , $\sin(\theta_2)$ as b , and θ_1 as c . Now, we can rewrite \mathcal{H} as follows:

$$\mathcal{H} = \{h_{a,b,c} : \mathbb{R}^2 \rightarrow \{0, 1\} \mid h_{a,b,c}(\mathbf{x}) = \mathbf{1}_{[ax+by+c>0]}, a, b, c \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^2\}$$

where $a, b \in [-1, 1]$ and $c \in \mathbb{R}$.

Let's have a look at the initial equation of \mathcal{H}_{lines} . To achieve similar bounding of $a, b \in [-1, 1]$ that is present in \mathcal{H} , we could normalize the entire equation by dividing the entire equation with euclidean norm of a and b , thus, achieving the same values present in \mathcal{H} . ($a, b \in [-1, 1]$ and $c \in \mathbb{R}$). Because of the normalization, the direction of the line would not be affected, meaning that we can still recreate every line in \mathbb{R}^2 . From this rendition, we can observe that \mathcal{H} is included in the family of \mathcal{H}_{lines} , implying that $\text{VCdim}(\mathcal{H}) \leq \text{VCdim}(\mathcal{H}_{lines})$, which is proven in Lecture 6 to be 3.

We can now verify that h_{θ_1, θ_2} can shatter a set of three points in order to determine that the VCdim is ≥ 3 . We can chose $C = \{(-2, 1), (1, -4), (2, 3)\}$

- | | |
|---|---|
| 1. • If $\theta_1 = -2, \theta_2 = -2 \rightarrow (0, 0, 0)$ | • If $\theta_1 = -1.6, \theta_2 = -0.5 \rightarrow (1, 0, 1)$ |
| • If $\theta_1 = -2, \theta_2 = -1.5 \rightarrow (1, 0, 0)$ | • If $\theta_1 = -1.5, \theta_2 = 1.7 \rightarrow (0, 1, 1)$ |
| • If $\theta_1 = -2, \theta_2 = -0.3 \rightarrow (0, 0, 1)$ | • If $\theta_1 = -0.7, \theta_2 = -2 \rightarrow (1, 1, 0)$ |
| • If $\theta_1 = -2, \theta_2 = 1.9 \rightarrow (0, 1, 0)$ | • If $\theta_1 = 1.4, \theta_2 = 1 \rightarrow (1, 1, 1)$ |

4 Exercise 4.

Consider \mathcal{H}_α to be the class of axis aligned rectangles with fixed aspect-ratio α , where $\alpha \in \mathbb{R}$ and $\alpha > 0$:

$$\mathcal{H}_\alpha = \left\{ h_{a,b,c,d} : \mathbb{R}^2 \rightarrow \{0, 1\} \mid h_{a,b,c,d}(\mathbf{x}) = \mathbf{1}_{[a,b] \times [c,d]}(\mathbf{x}), a, b, c, d \in \mathbb{R}, a < b, c < d, \frac{d-c}{b-a} = \alpha \right\}$$

Consider the realizability assumption.

4.1 a) give a learning algorithm A that returns a hypothesis h_S from \mathcal{H}_α , $h_S = A(S)$ consistent with the training set S (h_S has empirical risk 0 on S).

For completing this request, we can have a look in the first seminar and think of a similar approach, by choosing the tightest and the biggest rectangle as following: Consider the following algorithm A , that takes as input the training set S and outputs $h_S = h_{(a_S, b_S, c_S, d_S)}$, where

$$\begin{aligned} a_S &= \min_{\substack{i=1, \dots, m \\ y_i=1}} x_{i1} & c_S &= \min_{\substack{i=1, \dots, m \\ y_i=1}} x_{i2} \\ b_S &= \max_{\substack{i=1, \dots, m \\ y_i=1}} x_{i1} & d_S &= \max_{\substack{i=1, \dots, m \\ y_i=1}} x_{i2} \end{aligned}$$

- If all $y_i = 0$, then all points x_i have label 0, so there is no positive example. In this case, choose $z = (z_1, z_2)$ a point that is not in the training set S and take $a_S = b_S = z_1, c_S = d_S = z_2$. (Seminar 1)
- We can now choose, $h_S = h_{(a_S, b_S, c_S, d_S)} = \mathbf{1}_{[a_S, b_S] \times [c_S, d_S]}$ as the indicator function of the tightest rectangle $R_S = [a_S, b_S] \times [c_S, d_S]$ enclosing all positive examples.

For the biggest rectangle, we inverse the points mentioned above as following:

$$\begin{aligned} a'_S &= \max_{\substack{i=1, \dots, m \\ y_i=1}} x_{i1} & c'_S &= \max_{\substack{i=1, \dots, m \\ y_i=1}} x_{i2} \\ b'_S &= \min_{\substack{i=1, \dots, m \\ y_i=1}} x_{i1} & d'_S &= \min_{\substack{i=1, \dots, m \\ y_i=1}} x_{i2} \end{aligned}$$

If no valid point x_i labeled 0 that satisfies the condition of the boundary is found, then we simply set $a_S = a_m$. The same thing goes for the rest of the points, resulting in a new indicator function $h'_S = h_{(a'_S, b'_S, c'_S, d'_S)} = \mathbb{1}_{[a'_S, b'_S] \times [c'_S, d'_S]}$

4.2 b) find the sample complexity $m_{\mathcal{H}_\alpha}(\epsilon, \delta)$ in order to show that \mathcal{H}_α is PAC - learnable.

5 Exercise 5.

Compute $\text{VCdim}(\mathcal{H})$ where \mathcal{H} is the following hypothesis class:

$$\mathcal{H} = \{h_\theta : \mathbb{R} \rightarrow \{0, 1\} \mid h_\theta(x) = \mathbf{1}_{[\theta, \theta+1] \cup [\theta+2, \theta+4] \cup [\theta+6, \theta+9]}(x), \theta \in \mathbb{R}\}$$

From the beginning, we can conclude that the maximum dimension that \mathcal{H} can have is 6. Taking a set of 7 elements would make it impossible for the label (1,0,1,0,1,0,1) to appear. Now we can look at the smaller dimensions. Also, considering that we have 3 intervals, it's safe to assume that the VC dimension of \mathcal{H} is greater or equal to 2.

For the dimension of 3, we can continue with another simple set $\mathcal{C} = \{1, 2, 3\}$, alongside the theta values of:

- $\theta = 1000$ to achieve (0,0,0)
- $\theta = 3$ to achieve (0,0,1)
- $\theta = 1.5$ to achieve (0,1,0)
- $\theta = -8$ to achieve (1,0,0)
- $\theta = -7$ to achieve (1,1,0)
- $\theta = 2$ to achieve (0,1,1)
- $\theta = -3$ to achieve (1,0,1)
- $\theta = -5$ to achieve (1,1,1)

Thus, \mathcal{C} can be shattered, resulting in $\text{VCdim}(\mathcal{H}) \geq 3$. We can also find a set of 4 points that can be shattered by h_θ . Consider $\mathcal{C} = \{1, 2, 6.5, 8.5\}$.

1.
 - If $\theta = -5.0 \rightarrow (1, 1, 0, 0)$
 - If $\theta = -4.9 \rightarrow (0, 1, 0, 0)$
 - If $\theta = -3.9 \rightarrow (0, 0, 0, 0)$
 - If $\theta = -3.0 \rightarrow (1, 0, 0, 0)$
 - If $\theta = -2.5 \rightarrow (1, 0, 1, 0)$
 - If $\theta = -2.0 \rightarrow (1, 1, 1, 0)$
 - If $\theta = -0.9 \rightarrow (0, 1, 1, 0)$
 - If $\theta = -0.5 \rightarrow (0, 1, 1, 1)$
 - If $\theta = -0.0 \rightarrow (1, 1, 1, 1)$
 - If $\theta = 0.1 \rightarrow (1, 0, 1, 1)$
 - If $\theta = 0.6 \rightarrow (1, 0, 0, 1)$
 - If $\theta = 1.0 \rightarrow (1, 1, 0, 1)$
 - If $\theta = 1.1 \rightarrow (0, 1, 0, 1)$
 - If $\theta = 2.1 \rightarrow (0, 0, 0, 1)$
 - If $\theta = 2.5 \rightarrow (0, 0, 1, 1)$
 - If $\theta = 2.6 \rightarrow (0, 0, 1, 0)$

Thus, we can consider that \mathcal{C} is shatterable, resulting in $\text{VCdim}(\mathcal{H}) \geq 4$.

Now, finding 32 values for θ and 5 different values for x that would fit this general case, seems a little bit impossible. Let's see if we can prove that the VCdim of \mathcal{H} is lower than 5.

Consider the general case $a < b < c < d < e$, and the subset 1 0 1 0 1. We want to place a in the first interval, b in none, c in the second interval, d in none, and e in the third interval. This ordering is based on the order of real numbers.

If we take a to be in $[\theta, \theta + 1]$, b not in this interval, and c in $[\theta + 2, \theta + 4]$, then the distance (a, b) must be strictly greater than 1. This is because the distance of 1 is between $\theta + 1$ and $\theta + 2$. The same argument applies to the intervals $[\theta + 4, \theta + 6]$ for c and e. Therefore, the distance from c to e is greater than 2, which is the minimum distance needed to have a gap between $\theta + 4$ and $\theta + 6$. Thus, the distance from a to e is strictly greater than 3.

We also note that the intervals have lengths 1, 2, and 3. Therefore, it is not possible to obtain all numbers in the same interval. This leads to a contradiction, which implies that the subset 1 1 1 1 1 cannot be obtained.

For the subset 1 0 1 0 0, we should have:

- a in $[\theta, \theta + 1]$, b in none, c in $[\theta + 2, \theta + 4]$, d and e in none
- a in $[\theta + 2, \theta + 4]$, b in none, c in $[\theta + 6, \theta + 9]$

To achieve the first situation, the distance from a to c must be at most 4, which is the extreme case where a is the left endpoint of the first interval and c is the

right endpoint of the second interval. To achieve the second situation, the distance from a to c must be at most $7 ((\theta + 9) - (\theta + 2))$. Thus, taking the less restrictive situation, we obtain that $\text{dist}(a, c) \leq 7$.

The idea is that at this rate, we should either arrive at a contradiction or a "recipe" that tells us exactly how to construct the set that is a good example.

We can also consider the symmetric case $0\ 0\ 1\ 0\ 1$, which gives:

- c in $[\theta, \theta + 1]$, d in none, e in $[\theta + 2, \theta + 4]$, a and b in none
- c in $[\theta + 2, \theta + 4]$, d in none, e in $[\theta + 6, \theta + 9]$, a and b in none

From here, we can obtain $\text{dist}(c, e)$ is at most 4 and at most 7, respectively, which implies that the total distance is at most 7.