AML Assignment 2

Suto Robert-Lucian (411)

1 Exercise 1.

Consider $\mathcal{H}$ the following hypothesis class :

$$\mathcal{H} = \left\{ h_a : \mathbb{R} \to \{0,1\} \mid a > 0, a \in \mathbb{R}, \text{ where } h_a(x) = \mathbf{1}_{[-a,a]}(x) = \left\{ \begin{array}{ll} 1, & x \in [-a,a] \\ 0, & x \notin [-a,a] \end{array} \right\} \right.$$

1.1 a. Compute the growth function $\tau_H(m)$ (also known as shatter coefficient) for $m \geq 0$ for hypothesis class $\mathcal{H}$

From a first look, we can see that our hypothesis class, has a similar form to $\mathcal{H}_{thresholds}$, thus we can have a look at Lecture 8.

Consider $\mathrm{C} = \{\mathrm{c_1}, \mathrm{c_2}, \ldots, \mathrm{c_m}\}$ a set of m points, with $\mathrm{c_i} < \mathrm{c_j}$.
$\mathcal{H}_C$ can have at most $\mathrm{m}+1$ different functions: take $\mathrm{a_1} < \mathrm{c_1} < \mathrm{a_2} < \mathrm{c_2} < \ldots < \mathrm{a_m}$ $< \mathrm{c_m} < \mathrm{a_{m+1}}$, than we will have $|\mathcal{H}_C| = \left\{\mathrm{h_{a\_1}},\ \mathrm{h_{a_2 2}}, \ldots, \mathrm{h_{a\_m+1}}\right\} = \mathrm{m}+1$ as
$\mathrm{h_{a\_1}}$ labels points $\mathrm{c_1}, \mathrm{c_2}, \ldots, \mathrm{c_m}$ with labels $(0,0,0,\ldots,0,0)$
$\mathrm{h_{a\_2}}$ labels points $\mathrm{c_1}, \mathrm{c_2}, \ldots, \mathrm{c_m}$ with labels $(1,0,0,\ldots,0,0)$
. . . .
$\mathrm{h_{a\_3}}$ label points $\mathrm{c_1}, \mathrm{c_2}, \ldots, \mathrm{c_m}$ with labels $(1,1,0,\ldots,0,0)$
$\mathrm{h_{a\_m}}$ label points $\mathrm{c_1}, \mathrm{c_2}, \ldots, \mathrm{c_m}$ with labels $(1,1,1,\ldots,1,0)$
$\mathrm{h_{a\_m+1}}$ label points $\mathrm{c_1}, \mathrm{c_2}, \ldots, \mathrm{c_m}$ with labels $(1,1,1,\ldots,1,1)$

Which concludes that:

$$\tau_H(m) = \max_{C \subseteq X : C| = m} |H_C| = m+1$$

1.2 b. Compare your result from the previous point with the general upper bound given by the Sauer lemma. Are they equal or different?

In order to compute the upper bound, given by the Sauer lemma, we need to calculate the VCdim($\mathcal{H}$). We can simply prove that the VCdim($\mathcal{H}$) $= 1$.

Consider x_1, x_2, two points. While placing them inside our hypothesis class, we can observe that we have three possible cases:

- One, where $|x_1| < |x_2|$, where the label (0,1) cannot be achieved;
- Second one, where $|x_1| = |x_2|$. The problem is the labels (1,0) and (0,1) cannot be achieved;
- And the last one, $|x_1| > |x_2|$, where the label (1,0) cannot be achieved;

Because two points cannot be shattered, we can safely say that VCdim($\mathcal{H}$) $= 1$.

According to Sauer's lemma, the upper bound is given by the following inequality:

$$\tau_{\mathcal{H}}(m) \leq \sum_{i=0}^{d} C_m^d$$

where d is given by VCdim($\mathcal{H}$), and is known to be 1. Thus, the upper bound, given by the lemma is:

$$\sum_{i=0}^{d} C_m^i = C_m^0 + C_m^1 = 1 + m$$

which is equal to what we computed at subpoint a.

2 Exercise 2.

Consider de concept class C_a formed by the union of two closed intervals $[a, a+1] \cup [a+2, a+4]$, where $a \in \mathbb{R}$. Give an efficient ERM algorithm for learning the concept class C_a and compute its complexity for each of the following cases:

2.1 a. realizable case.

2.2 b. agnostic case.

3 Exercise 3.

Consider a modified version of the AdaBoost algorithm that runs for exactlythree rounds as follows:

- the first two rounds run exactly as in AdaBoost (at round 1 we obtain distribution $\mathbf{D}^{(1)}$, weak classifier h_1 with error c_1; at round 2 we obtain distribution $\mathbf{D}^{(2)}$, weak classifier h_2 with error c_2)
- in the third round we compute for each $i = 1, 2, \ldots, m$:

$$\mathbf{D}^{(3)}(i) = \begin{cases} \frac{D^{(1)}(i)}{Z}, & \text{if } h_1(x_i) \neq h_2(x_i) \\ 0, & \text{otherwise} \end{cases}$$

  where Z is a normalization factor such that $\mathbf{D}^{(3)}$ is a probability distribution.
- obtain weak classifier h_3 with error ϵ_3.
- output the final classifier $h_{final}(x) = \text{sign}\left(h_1(x) + h_2(x) + h_3(x)\right)$.

Assume that at each round $t = 1, 2, 3$ the weak learner returns a weak classifier h_t for which the error c_t satisfies $c_t \leq \frac{1}{2} - \gamma_t, \gamma_t > 0$.

3.1 **a) What is the probability that the classifier h_1 will be selected again at round 2 ? Justify your answer.**

3.2 **b) Consider $\gamma = \min\{\gamma_1, \gamma_2, \gamma_3\}$. Show that the training error of the final classifier $h_{\textbf{final}}$ is at most $\frac{1}{2} - \frac{3}{2}\gamma + 2\gamma^3$ and show that this is strictly smaller than $\frac{1}{2} - \gamma$.**

$$
\begin{aligned}
&\frac{1}{2} - \frac{3}{2}\gamma_{\min} + 2\gamma_{\min}^3 < \frac{1}{2} - \gamma_{\min} \implies \\
&-\frac{1}{2} + \gamma_{\min} + \frac{1}{2} - \frac{3}{2}\gamma_{\min} + 2\gamma_{\min}^3 < 0 \implies \\
&-\frac{3}{2}\gamma_{\min} + 2\gamma_{\min}^3 + \gamma_{\min} < 0 \implies \\
&\frac{3}{2}\gamma_{\min} - 2\gamma_{\min}^3 - \gamma_{\min} > 0 \implies \\
&\frac{1}{2}\gamma_{\min} - 2\gamma_{\min}^3 > 0 \implies \\
&\gamma_{\min}\left(\frac{1}{2} - 2\gamma_{\min}^2\right) > 0 \implies \\
&\frac{1}{2} - 2\gamma_{\min}^2 > 0 \implies \\
&2\gamma_{\min}^2 < \frac{1}{2} \implies \\
&\gamma_{\min}^2 < \frac{1}{4} \implies \\
&\gamma_{\min} < \frac{1}{2}
\end{aligned}
$$

From the assumption made in the request (the weak learner returns a weak classifier h_t for which the error c_t satisfies $c_t \leq \frac{1}{2} - \gamma_t, \gamma_t > 0$), we can affirm that $\gamma_{\min} < \frac{1}{2}$ is true, hence, is correct.