Assignment 2

Deadline: Year 1 - Sunday, 20th of June, 23:59, Year 2 - Sunday, 13th of June, 23:59

Upload your solutions as a zip archive at: https://tinyurl.com/AML-2021-ASSIGNMENT2

1. **(2 points)** Consider $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$, where
   $\mathcal{H}_1 = \{h_{\theta_1}: \mathbb{R} \to \{0,1\}, h_{\theta_1}(x) = \mathbf{1}_{[x \geq \theta_1]}(x) = \mathbf{1}_{[\theta_1,+\infty)}(x), \theta_1 \in \mathbb{R}\}$ and
   $\mathcal{H}_2 = \{h_{\theta_2}: \mathbb{R} \to \{0,1\}, h_{\theta_2}(x) = \mathbf{1}_{[x < \theta_2]}(x) = \mathbf{1}_{(-\infty,\theta_2)}(x), \theta_2 \in \mathbb{R}\}$
   - **a.** Give an efficient ERM algorithm for learning $\mathcal{H}$ and compute its complexity for the realizable case. **(0.5 points)**
   - **b.** Give an efficient ERM algorithm for learning $\mathcal{H}$ and compute its complexity for the agnostic case. **(0.5 points)**
   - **c.** Compute the shattering coefficient $\tau_H(m)$ of the growth function for m $\geq$ 0 for hypothesis class $\mathcal{H}$. **(0.5 points)**
   - **d.** Compare your result with the general upper bound for the growth functions and show that $\tau_H(m)$ obtained at previous point c is not equal to the upper bound. **(0.25 points)**
   - **e.** Does there exist a hypothesis class $\mathcal{H}$ for which $\tau_H(m)$ is equal to the general upper bound (over $\mathbb{R}$ or another domain $\mathcal{X}$)? If your answer is yes please provide an example, if your answer is no please provide a justification. **(0.25 points)**

2. **(1.5 points)** Consider a modified version of the AdaBoost algorithm that runs for exactly three 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 ε_1, at round 2 we obtain distribution $\mathbf{D}^{(2)}$, weak classifier h_2 with error ε_2).
   - in the third round we compute for each i = 1, 2, …, m:
   $$D^{(3)}(i) = \begin{cases} \dfrac{D^{(1)}(i)}{Z}, if\ h_1(x_i) \neq h_2(x_i) \\ 0,\ \ otherwise \end{cases}$$
   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 $h_{final}(x) = \text{sign}(h_1(x) + h_2(x) + h_3(x))$.

   Assume that at each round *t = 1, 2, 3* the weak learner returns a weak classifier h_t for which the error ε_t satisfies $\varepsilon_t \leq 1/2 - \gamma_t$, $\gamma_t > 0$.
   - a. What is the probability that the classifier h_1 (selected at round 1) will be selected again at round 2? Justify your answer. **(0.50 points)**
   - b. Consider $\gamma = \min\{\gamma_1, \gamma_2, \gamma_3\}$. Show that the training error of the final classifier h_{final} is at most $\frac{1}{2} - \frac{3}{2}\gamma + 2\gamma^3$ and show that this is strictly smaller than $\frac{1}{2} - \gamma$. **(1 point)**

3. **(1.5 points)** Let Σ be a finite alphabet and let $\mathcal{X} = \Sigma^m$ be a sample space of all strings of length m over Σ. Let $\mathcal{H}$ be a hypothesis space over $\mathcal{X}$, where $\mathcal{H} = \{h_w : \Sigma^m \to \{0,1\}, w \in \Sigma^*, 0 < |w| \leq m, s.t. h_w(x) = 1\ if\ w\ is\ a\ substring\ of\ x\}$.
   a. Give an upper bound (any upper bound that you can come up) of the VC-dimension of $\mathcal{H}$ in terms of $|\Sigma|$ and m. **(0.25 points)**
   b. Give an efficient algorithm for finding a hypothesis h_w consistent with a training set in the realizable case. What is the complexity of your algorithm? **(1.25 points)**

Example: let $\Sigma = \{\mathrm{a, b, c}\}, \mathrm{m} = 4$ and the training set S = $\{$(aabc, 1), (baca, 0), (bcac, 0), (abba, 1)$\}$. The output of the algorithm should be h_{ab}.

Ex-officio: 0.5 points