Seminar class 6

Exercise 1 (exercise 10.1 in the book)

10.1 **Boosting the Confidence:** Let A be an algorithm that guarantees the following: There exist some constant $\delta_0 \in (0,1)$ and a function $m_{\mathcal{H}} : (0,1) \to \mathbb{N}$ such that for every $\epsilon \in (0,1)$, if $m \geq m_{\mathcal{H}}(\epsilon)$ then for every distribution $\mathcal{D}$ it holds that with probability of at least $1 - \delta_0$, $L_{\mathcal{D}}(A(S)) \leq \min_{h \in \mathcal{H}} L_{\mathcal{D}}(h) + \epsilon$.

Suggest a procedure that relies on A and learns $\mathcal{H}$ in the usual agnostic PAC learning model and has a sample complexity of

$$m_{\mathcal{H}}(\epsilon, \delta) \leq k\, m_{\mathcal{H}}(\epsilon) + \left\lceil \frac{2\log(4k/\delta)}{\epsilon^2} \right\rceil,$$

where

$$k = \lceil \log(\delta)/\log(\delta_0) \rceil.$$

Hint: Divide the data into $k+1$ chunks, where each of the first k chunks is of size $m_{\mathcal{H}}(\epsilon)$ examples. Train the first k chunks using A. Argue that the probability that for all of these chunks we have $L_{\mathcal{D}}(A(S)) > \min_{h \in \mathcal{H}} L_{\mathcal{D}}(h) + \epsilon$ is at most $\delta_0^k \leq \delta/2$. Finally, use the last chunk to choose from the k hypotheses that A generated from the k chunks (by relying on Corollary 4.6).

Corollary 4.6. *Let $\mathcal{H}$ be a finite hypothesis class, let Z be a domain, and let $\ell : \mathcal{H} \times Z \to [0,1]$ be a loss function. Then, $\mathcal{H}$ enjoys the uniform convergence property with sample complexity*

$$m_{\mathcal{H}}^{UC}(\epsilon, \delta) \leq \left\lceil \frac{\log(2|\mathcal{H}|/\delta)}{2\epsilon^2} \right\rceil.$$

Furthermore, the class is agnostically PAC learnable using the ERM algorithm with sample complexity

$$m_{\mathcal{H}}(\epsilon, \delta) \leq m_{\mathcal{H}}^{UC}(\epsilon/2, \delta) \leq \left\lceil \frac{2\log(2|\mathcal{H}|/\delta)}{\epsilon^2} \right\rceil.$$

Exercise 2 (exercise 7.6 in the book "Foundations of Machine Learning", 2nd edition)

7.6 Fix $\epsilon \in (0, 1/2)$. Let the training sample be defined by m points in the plane with $\frac{m}{4}$ negative points all at coordinate $(1,1)$, another $\frac{m}{4}$ negative points all at coordinate $(-1,-1)$, $\frac{m(1-\epsilon)}{4}$ positive points all at coordinate $(1,-1)$, and $\frac{m(1+\epsilon)}{4}$ positive points all at coordinate $(-1,+1)$. Describe the behavior of AdaBoost when run on this sample using boosting stumps. What solution does the algorithm return after T rounds?

Exercise 3 (exercise 10.3 in the book)

10.4 In this exercise we discuss the VC-dimension of classes of the form $L(B,T)$. We proved an upper bound of $O(dT\log(dT))$, where $d = \text{VCdim}(B)$. Here we wish to prove an almost matching lower bound. However, that will not be the case for all classes B.

1. Note that for every class B and every number $T \geq 1$, $\text{VCdim}(B) \leq \text{VCdim}(L(B,T))$. Find a class B for which $\text{VCdim}(B) = \text{VCdim}(L(B,T))$ for every $T \geq 1$.
   Hint: Take $\mathcal{X}$ to be a finite set.
2. Let B_d be the class of decision stumps over $\mathbb{R}^d$. Prove that $\log(d) \leq \text{VCdim}(B_d) \leq 5 + 2\log(d)$.
   Hints:
   - For the upper bound, rely on Exercise 10.11.
   - For the lower bound, assume $d = 2^k$. Let A be a $k \times d$ matrix whose columns are all the d binary vectors in $\{\pm 1\}^k$. The rows of A form a set of k vectors in R^d. Show that this set is shattered by decision stumps over $\mathbb{R}^d$.
3. Let $T \geq 1$ be any integer. Prove that $\text{VCdim}(L(B_d,T)) \geq 0.5T\log(d)$.
   Hint: Construct a set of $\frac{T}{2}k$ instances by taking the rows of the matrix A from the previous question, and the rows of the matrices $2A, 3A, 4A, \ldots, \frac{T}{2}A$. Show that the resulting set is shattered by $L(B_d,T)$.

6.11 **VC of union:** Let $\mathcal{H}_1, \ldots, \mathcal{H}_r$ be hypothesis classes over some fixed domain set $\mathcal{X}$. Let $d = \max_i \text{VCdim}(\mathcal{H}_i)$ and assume for simplicity that $d \geq 3$.

1. Prove that

$$\text{VCdim}\left(\cup_{i=1}^{r}\mathcal{H}_i\right) \leq 4d\log(2d) + 2\log(r).$$