MLT Homework 5

Ana Borovac
Jonas Haslbeck
Bas Haver

October 15, 2018

Question 1

Subquestion 1.1

Consider a hypothesis class $\mathcal{H} = \cup_{n=1}^{\infty} \mathcal{H}_n$*, where for every* $n \in \mathbb{N}$*,* $\mathcal{H}_n$ *is finite. Find a weighting function* $w : \mathcal{H} \to [0, 1]$ *such that* $\sum_{h \in \mathcal{H}} w(h) \leq 1$ *and so that for all* $h \in \mathcal{H}$*,* $w(h)$ *is determined by* $|\mathcal{H}_{n(h)}|$*.*

Solution

Since we have a countably infinite union of finite sets, we know that the number of elements is countably infinite. Therefore, we can number them as:

$$h_1, h_2, \ldots$$

If we pick weights as:

$$w(h_i) = \left(\frac{1}{2^{|H_{n(h_i)}|}} \right)^i ; \quad i = 1, 2, \ldots$$

the sum of the weights in the worst case would be, when $|H_{n(h_i)}| = 1$ $(\forall i)$ and we have infinity number of different hypothesis:

$$\sum_{i=1}^{\infty} w(h_i) = \sum_{i=1}^{\infty} \left(\frac{1}{2} \right)^i = 1$$

Subquestion 1.2

Define such a function w *when for all* n $\mathcal{H}_n$ *is countable (possibly infinite).*

Solution

Countably infinite union of countable sets is again a countable set, so we can choose the same weighted function as before.

Question 2

In this question we wish to show a No-Free-Lunch result for nonuniform learnability.

Subquestion 2.1

Let A be a nonuniform learner for class $\mathcal{H}$. For each $n \in \mathbb{N}$ define $\mathcal{H}_n^A = \{h \in \mathcal{H} : m^{NUL}(0.1, 0.1, h) \leq n\}$. Prove that each such class $\mathcal{H}_n$ has a finite VC-dimension.

Solution

Because $\mathcal{H}$ is nonuniform learnable, it is a union of agnostic PAC learnable hypothesis classes (Theorem 7.2). That is, $\mathcal{H} = \bigcup_{n\in\mathbb{N}} \mathcal{H}_n^A$, and $\mathcal{H}_n^A$ is agnostic PAC learnable for all $n \in \mathbb{N}$. Because each $\mathcal{H}_n^A$ is agnostic PAC learnable, it also has a finite VC-dimension, by the Fundamental Theorem of Statistical Learning (Theorem 6.7).

Subquestion 2.2

Prove that if class $\mathcal{H}$ in nonuniformly learnable then there are classes $\mathcal{H}_n$ so that $\mathcal{H} = \cup_{n\in\mathbb{N}}\mathcal{H}_n$ and, for every $n \in \mathbb{N}$, VCdim$(\mathcal{H}_n)$ is finite.

Solution

This question was solved with the help of [1].

Let define:

$$\mathcal{H}_n = \{h \in \mathcal{H} : m^{\mathrm{NUL}}(0.1, 0.1, h) \leq n\}$$

It is obvious that $\mathcal{H} = \cup_{n\in\mathbb{N}}\mathcal{H}_n$. From previous point we also know that VCdim$(\mathcal{H}_n)$ is finite.

Subquestion 2.3

Let $\mathcal{H}$ be a class that shatters some infinite set. Then for every sequence of classes $(\mathcal{H}_n : n \in \mathbb{N})$ such that $\mathcal{H} = \cup_{n\in\mathbb{N}}\mathcal{H}_n$, there exists some n for which VCdim$(\mathcal{H}_n) = \infty$.

Solution

By assumption the hypothesis class $\mathcal{H}$ shatters some infinite set K. Let $(\mathcal{H}_n : n \in \mathbb{N})$ be a set of hypothesis classes, each having a finite VC-dimension. We define subsets $K_n \subseteq K$ such that, for all n, $|K_n| > \text{VCdim}(H_n)$, and all subsets are nonoverlapping, so $K_n \cap K_m = \emptyset$.

Now, for each K_n we pick a function $f_n : K_n \to \{0, 1\}$ so that no $h \in H_n$ agrees with f_n on the domain K_n. Such a function exists for all K_n, because H_n does not shatter K_n, since $|K_n| > \text{VCdim}(H_n)$.

Next, we define the function $f : K \to \{0, 1\}$ by combining all f_n's

$$f(k) = \begin{cases} f_1(k) & \text{if } k \in K_1 \\ \vdots \\ f_n(k) & \text{if } k \in K_n \end{cases}$$

where $k \in K$ are singletons in K.

We define $H_0 = \cup_{n \in \mathbb{N}} \mathcal{H}_n$. By the construction of all f_n's it follows that f is not contained in H_0. However, $f \in H$, because H shatters K by assumption. Put differently, $f \in (\mathcal{H} \setminus H_0)$.

This implies that $(\mathcal{H} \setminus H_0) \neq \emptyset$. And since we did not put any restrictions on H_n's except $\text{VCdim}(H_n) < \infty$, $(\mathcal{H} \setminus H_0)$ must contain at least one hypothesis class that has infinite VC-dimension, wich concludes the proof.

Subquestion 2.4

Construct a class $\mathcal{H}_1$ of functions from the unit interval $[0, 1]$ to $\{0, 1\}$ that is nonuniformly learnable but not PAC learnable.

Solution

This question was solved with the help of [1].

Let denote:

$$\mathcal{H}_n = \{h_{a_1,\ldots,a_n,b_1,\ldots,b_n}; 0 \leq a_1 < b_1 \leq 1, \ldots, 0 \leq a_n < b_n \leq 1\}$$

where:

$$h_{a_1,\ldots,a_n,b_1,\ldots,b_n}(x) = \begin{cases} 1; & a_1 \leq x \leq b_1 \text{ or } \cdots \text{ or } a_n \leq x \leq b_n \\ 0; & \text{otherwise} \end{cases}$$

We claim that $\text{VCdim}(\mathcal{H}_n) = 2n$.

- $\text{VCdim}(\mathcal{H}_n) \geq 2n$: We would like to show that $\mathcal{H}_n$ shatters a set of $2n$ points. Assume that $0 \leq x_1 < \cdots < x_{2n} \leq 1$. Labeling for which we need the largest number of disjoint intervals is when x-es with even indexes are labeled 0 and others are labeled 1 (or vice versa). This is exactly n intervals that we need, so $\mathcal{H}_n$ shatters $\{x_1, \ldots, x_{2n}\}$.

- VCdim($\mathcal{H}_n$) $\leq 2n$: Now, we are going to take a set of $2n+1$ elements; $0 \leq x_1 < \cdots < x_{2n+1} \leq 1$. If we want that $\mathcal{H}_n$ shatters our set, there must exist a labeling function which can label the situation where x-es with odd indexes are labeled 1 and others are labeled 0. In order to do that we would need at least $n+1$ intervals, which we do not have. So, $\mathcal{H}_n$ does not shatter a set of $2n+1$ elements.

Denote $\mathcal{H} = \bigcup_{i=1}^{\infty} \mathcal{H}_n$. From previous points we can conclude that $\mathcal{H}$ is nonuniformly learnable. It is not PAC learnable due to VCdim($\mathcal{H}$) $= \infty$.

Subquestion 2.5

Construct a class $\mathcal{H}_2$ of functions from the unit interval $[0,1]$ to $\{0,1\}$ that is not nonuniformly learnable.

Solution

This question was solved with the help of [1].

Define $\mathcal{H}$ as a set of all intervals over $[0,1]$ (hypothesis labels an element with 1 if it is inside of at least one interval, otherwise it labels it with 0). If is obvious that $\mathcal{H}$ shatters a set $S = \{\frac{1}{n};\ n \in \mathbb{N}\}$. We also know that S is infinite. Therefore $\mathcal{H}$ is not nonuniformly learnable (from previous points).

Question 3

Prove the Symmetrization Lemma (= double sample trick):
For any $\epsilon > 0$ such that $n\epsilon^2 \geq 2$,

$$\sup_{f \in \mathcal{F}} P_n \left[(R(f) - R_n(f)) \geq \epsilon \right] \leq P_{2n} \left[\sup_{f \in \mathcal{F}} (R'_n(f) - R_n(f)) \geq \frac{\epsilon}{2} \right]$$

where R is the risk, R_n empirical risk for the sample $Z_1, \ldots, Z_n$ and R'_n empirical risk for ghost sample $Z'_1, \ldots, Z'_n$.

Solution

This question was solved with the help of [2].

Denote f^* a function that maximize $(R(f) - R_n(f))$. First, we would like to prove: If $(R(f^*) - R_n(f^*) \geq \epsilon)$ and $(R(f^*) - R'_n(f^*) \leq \frac{\epsilon}{2})$ then $(R'_n(f^*) - R_n(f^*) \geq \frac{\epsilon}{2})$.

$$\begin{aligned}
\epsilon &< R(f^*) - R_n(f^*) \\
&= R(f^*) - R'_n(f^*) + R'_n(f^*) - R_n(f^*) \\
&\leq R'_n(f^*) - R_n(f^*) + \frac{\epsilon}{2}
\end{aligned}$$

So, $(R'_n(f^*) - R_n(f^*)) \geq \frac{\epsilon}{2}$. If we write that with indicators:

$$I\left[R(f^*) - R_n(f^*) > \epsilon, R(f^*) - R'_n(f^*) \leq \frac{\epsilon}{2}\right] \leq I\left[R'_n(f^*) - R_n(f^*) \geq \frac{\epsilon}{2}\right]$$

Now, we are going to compute expected value over $Z'_1, \ldots, Z'_n$:

$$I\left[R(f^*) - R_n(f^*) > \epsilon\right] P_n\left[R(f^*) - R'_n(f^*) \leq \frac{\epsilon}{2}\right] \leq P_n\left[R'_n(f^*) - R_n(f^*) \geq \frac{\epsilon}{2}\right]$$

With Chebishyev inequality we get and $Var(f^*) \leq \frac{1}{4}$ (due to $f^* \in [0,1]$) :

$$P_n\left[R(f^*) - R'_n(f^*) \leq \frac{\epsilon}{2}\right] \geq 1 - \frac{4Var(f^*)}{n\epsilon^2} \geq 1 - \frac{1}{n\epsilon^2} \geq \frac{1}{2}$$

Therefore:

$$\sup_{f \in \mathcal{F}} I\left[R(f) - R_n(f) \geq \epsilon\right] \leq 2P_n\left[\sup_{f \in \mathcal{F}}(R'_n(f) - R_n(f)) \geq \frac{\epsilon}{2}\right]$$

Last, we again compute expectation but this time over $Z_1, \ldots, Z_n$:

$$\sup_{f \in \mathcal{F}} P_n\left[R(f) - R_n(f) \geq \epsilon\right] \leq 2P_{2n}\left[\sup_{f \in \mathcal{F}}(R'_n(f) - R_n(f)) \geq \frac{\epsilon}{2}\right]$$

References

[1] Gyora M. Benedek and Alon Itai. Nonuniform learnability. *J. Comput. Syst. Sci.*, 48(2):311–323, 1994.

[2] Han Liu John Lafferty and Larry Wasserman. Concentration of measure. `www.stat.cmu.edu/~larry/=sml/Concentration.pdf`. Online; accessed 14 October 2018.