ELL880: Computational Learning Theory and the Mind

Minor Test II, Maximum marks: 20

Section 1.

1. In proving the no-free-lunch theorem in class, we took a subset $C \subset \mathbb{X}$, $|C| = 2N$, and showed that for any learner A which sees a training set S with only N samples from C, there exists a data-generating distribution $\mathcal{D}$ such that

$$\mathop{\mathbb{E}}_{S\sim\mathcal{D}^N}\left[L_{\mathcal{D}}(A(S))\right] \geq \frac{1}{4}.$$

(a) Complete the proof by showing formally (with full working) that the above implies lack of PAC learnability; specifically, that **[4]**

$$\mathop{p}_{S\sim\mathcal{D}^N}\left(L_{\mathcal{D}}(A(S)) \geq \frac{1}{8}\right) \geq \frac{1}{7}.$$

(Hint: Make use of Markov's measure concentration inequality, which says that for a non-negative random variable X, $\forall a \geq 0$, $p(X \geq a) \leq \mathbb{E}[X]/a$.)

(b) PAC learnability allows us to specify a minimal sample complexity; *i.e.*, it allows us to demand as much data as we need, so long as it is a finite quantity. The above appears to be showing lack of PAC learnability only when the training set is of a particular size N. Why can we not get around the above argument by saying that if our learner sees more data (say, all $2N$ samples from C), then we can in fact get PAC learnability? **[3]**

2. Consider the hypothesis class $\mathcal{H} = \{h_{a,b,c,d} : a, b, c, d \in \mathbb{R}, a \leq b \leq c \leq d\}$, where $\forall x \in \mathbb{R}$:

$$h_{a,b,c,d}(x) = \begin{cases} 1 & \text{if } a \leq x \leq b, \text{ or } c \leq x \leq d; \\ 0 & \text{otherwise.} \end{cases}$$

Obtain the VC-dimension of $\mathcal{H}$, showing a full proof. **[6]**

3. In class we said that nonuniform learnability is a relaxation of PAC learnability, and that it is possible for a hypothesis class to be nonuniformly learnable but not PAC learnable. But now, Prof. Kinpav[1] comes along and gives the following counter-argument.

> If $\mathcal{H}$ is 'nonuniformly learnable', as you say, then there exists a learner which can (ϵ, δ)-compete with every hypothesis h in $\mathcal{H}$, given a training set of size at least $N_{\mathcal{H}}^{NUL}(\epsilon, \delta, h)$. Right? Now, for all possible values of ϵ and δ, let me define
>
> $$N_{\mathcal{H}}(\epsilon, \delta) = \max_{h\in\mathcal{H}} N_{\mathcal{H}}^{NUL}(\epsilon, \delta, h).$$
>
> If I have $N \geq N_{\mathcal{H}}(\epsilon, \delta)$, then my learner is (ϵ, δ)-competitive with all hypotheses in $\mathcal{H}$, and hence in particular with the one with lowest true loss. Hence, with sample complexity $N_{\mathcal{H}}(\epsilon, \delta)$, my learner is also a PAC learner for $\mathcal{H}$. Thus, your 'nonuniform learnability' is useless, PAC learnability is all you need. *Spasiba, da svidanya*![2]

(a) Explain, as precisely as you can, why Prof. Kinpav is wrong. **[4]**

(b) Why does Prof. Kinpav's argument appear intuitively correct? Are there some conditions under which it would indeed be valid (*i.e.*, under which nonuniform learnability would indeed imply PAC learnability)? If so, what might such conditions be? **[3]**

[1] Any resemblance to persons living or dead is purely coincidental.
[2] Russian for *Thank you, bye*!