Math 490 Solutions

Fall 2018

1. **[Triple-S 3.6: Agnostic generalizes]**

Suppose that $\mathcal{H}$ is agnostic PAC learnable with the agnostic PAC learner $\mathcal{A}$. Show that $\mathcal{H}$ is PAC learnable with the canonical learning algorithm derived by $\mathcal{A}$.

Solution:

Suppose that $\mathcal{H}$ is agnostic PAC learnable with the agnostic PAC learner $\mathcal{A}$. Fix $\epsilon, \delta \in (0,1)$ then there exists a function $m_{\mathcal{H}} : (0,1)^2 \to \mathbb{N}$ so that for any distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$, if $m \geq m_{\mathcal{H}}(\epsilon, \delta)$ and there are m samples $S^m = ((X_1, Y_1), ..., (X_m, Y_m))$ drawn from $\mathcal{D}$ then

$$\mathbb{P}\left\{ L_{\tilde{\mathcal{D}}}(h) \leq \min_{h' \in \mathcal{H}} L_{\tilde{\mathcal{D}}}(h') + \epsilon \right\} > 1 - \delta,$$

where $h := \mathcal{A}(S^m)$.

Fix distribution $\mathcal{D}$ over $\mathcal{X}$, and suppose we are given the labeling function $f : \mathcal{X} \to \{0,1\}$, with the realizable assumption on $\mathcal{H}, \mathcal{D}, f$. Suppose that $m \geq m_{\mathcal{H}}(\epsilon, \delta)$. Consider the m data drawn from $\mathcal{D}$, $S^m = (X_1, ..., X_m)$. Using the given labeling function f, the restriction on our distribution is of small consequence. By the realizability hypothesis, $\min_{h' \in \mathcal{H}} L_{(\mathcal{D},f)}(h') = 0$.

Define $\hat{\mathcal{A}}(S^m) := \mathcal{A}(S^m, f(S^m))$, where $f(S^m) := (f(X_1), ..., f(X_m))$. Finally, let $\tilde{\mathcal{D}}$ be the distribution over $\mathcal{X} \times f(\mathcal{X})$. Then by definition $L_{\tilde{\mathcal{D}}}(\mathcal{A}(S^m, f(S^m))) = \mathbb{P}\{h(x) \neq f(x)\} = L_{(\mathcal{D},f)}(\hat{\mathcal{A}}(S^m))$. Moreover $0 \leq \min_{h' \in \mathcal{H}} L_{\tilde{\mathcal{D}}}(h') \leq \min_{h' \in \mathcal{H}} L_{(\mathcal{D},f)}(h') = 0$. Since $f(\mathcal{X}) \subset \mathcal{Y}$, it follows that $\mathcal{A}$ PAC learns the data $(S^m, f(S^m))$. That is, if $h := \hat{\mathcal{A}}(S^m) = \mathcal{A}(S^m, f(S^m))$ then

$$\mathbb{P}\left\{ L_{(\mathcal{D},f)}(h) \leq \epsilon \right\} = \mathbb{P}\left\{ L_{\tilde{\mathcal{D}}}(h) \leq 0 + \epsilon \right\} = \mathbb{P}\left\{ L_{\tilde{\mathcal{D}}}(h) \leq \min_{h' \in \mathcal{H}} L_{\tilde{\mathcal{D}}}(h') + \epsilon \right\} > 1 - \delta. \qquad \square$$

NOTE: The purpose of this problem is to change the algorithm, just an iota, to get it to learn on the realizable case. The issue here is that in the realizable case your not given distributions over $\mathcal{X} \times \mathcal{Y}$ but distributions over $\mathcal{X}$ instead. The reason this is not trivial, is that you have to account for this in your algorithm and your loss function using the overpowered label function given f after the realization that $\mathcal{X} \times f(\mathcal{X})$ is the correct space of distributions for $\mathcal{A}$. The minutia of this problem may be nontrivial to account for but hopefully the concept will become clear after seeing the solution.

A common mistake was to assume that $y_i's$ where given instead of produced using f.

2. **[Triple-S 4.2: Bounded loss functions]**

Generalize Corollary 4.6 in Triple-S for loss functions with range $[a, b]$.

Solution:

Let $\overline{L_{\mathcal{D}}} = \frac{L_{\mathcal{D}}-a}{b-a}$ and $\overline{L_S} = \frac{L_S-a}{b-a}$. Then by the proof of Corollary 4.6 in Triple-S, we have that for $m \geq \frac{\log(2|\mathcal{H}|/\delta)}{2\left(\frac{\epsilon}{b-a}\right)^2}$ then

$$\begin{aligned}
\mathcal{D}\left(\{S : \exists h \in \mathcal{H}, |L_{\mathcal{D}} - L_S| > \epsilon\}\right) &= \mathcal{D}\left(\left\{S : \exists h \in \mathcal{H}, \left|\frac{L_{\mathcal{D}} - a}{b-a} - \frac{L_S - a}{b-a}\right| > \frac{\epsilon}{b-a}\right\}\right) \\
&= \mathcal{D}\left(\left\{S : \exists h \in \mathcal{H}, \left|\overline{L_{\mathcal{D}}} - \overline{L_S}\right| > \frac{\epsilon}{b-a}\right\}\right) \leq \delta.
\end{aligned}$$

By Corollary 4.4 in Triple-S the result follows.

3. **[Triple-S 5.1: Proof of N.F.L. using equation (5.2)]**

Using equation (5.2) and Lemma B.1 in Triple-S; finish the proof of the No Free Lunch theorem by showing that $\mathbb{P}(L_{\mathcal{D}}(\mathcal{A}(S)) \geq 1/8) \geq \frac{1}{7}$.

Solution:

Then by equation (5.2) $\mathbb{E}[L_{\mathcal{D}}(\mathcal{A}(S))] \geq 1/4$. By Lemma B.1 in Triple-S

$$\mathbb{P}\left\{L_{\mathcal{D}}(\mathcal{A}(S)) \geq 1/8\right\} = \mathbb{P}\left\{L_{\mathcal{D}}(\mathcal{A}(S)) \geq 1 - 7/8\right\} \geq \frac{\mathbb{E}[L_{\mathcal{D}}(\mathcal{A}(S))] - (1 - 7/8)}{7/8} = \frac{1/4 - 1/8}{7/8} = \frac{1}{7}.$$