CS 6316 Homework 2

James Harbour, gtr8rh@virginia.edu

February 18, 2024

1 Problem 1

Let $\mathcal{X}$ be a discrete domain, and let $\mathcal{H}_{\text{Singleton}} = \{h_z : z \in \mathcal{X}\} \cup \{h^-\}$, where for each $z \in \mathcal{X}$, h_z is the function defined by

$$h_z(x) := \begin{cases} 1 & \text{if } x = z \\ 0 & \text{if } x \neq z \end{cases}$$

and h^- is the null hypothesis, namely $h^-(x) = 0$ for all $x \in \mathcal{X}$. The realizability assumption here implies that the true hypothesis f labels negatively all examples in the domain, perhaps except one.

1. Describe an algorithm that implements the ERM rule for learning $\mathcal{H}_{\text{Singleton}}$ in the realizable setup.

Solution. For brevity, we write $\mathcal{H} := \mathcal{H}_{\text{Singleton}}$. Sample finite $S = \{x_1, \dots, x_m\}$ according to $\mathcal{D}$. We wish to cosntruct a learning algorithm A such that the hypothesis $A(S)$ satisfies

$$h_S := A(S) \in \operatorname*{argmin}_{h \in \mathcal{H}} L_S(h)$$

As $\mathcal{X}$ is discrete, a short computation reveals that realizability implies $f = 0$ or $f = \mathbb{1}_{z^*}$ for some $z^* \in \mathcal{X}$. Fix $z \in \mathcal{X}$.

If $f(x) = 0$ for all $x \in \mathcal{S}$, then

$$L_S(h_z) = \frac{1}{|S|}|\{x \in S : h_z(x) \neq f(x)\}| = \frac{1}{|S|}|\{x \in S : \mathbb{1}_z(x) \neq 0\}| = \frac{1}{|S|}\mathbb{1}_z(S).$$

In the other case, suppose that there is some z^* with $f(z^*) = 1$, in short $|f^{-1}(\{1\})| \in \{0, 1\}$. Clearly, if $z = z^*$ then $L_S(h_z) = 0$. On the other hand, if $z \neq z^*$, then

$$\begin{aligned} L_S(h_z) &= \frac{1}{|S|}|\{x \in S : h_z(x) \neq f(x)\}| = \frac{1}{|S|}|\{x \in S : \mathbb{1}_z(x) \neq \mathbb{1}_{z^*}(x)\}| \\ &= \frac{1}{|S|}\left|\{x \in S \setminus \{z^*\} : \mathbb{1}_z(x) \neq 0\} \sqcup \{x \in S \cap \{z^*\} : \mathbb{1}_z(x) \neq 1\}\right| = \frac{1}{|S|}\left(\mathbb{1}_z(S) + \mathbb{1}_{z^*}(S)\right) \end{aligned}$$

In both cases, we have the identity

$$L_S(h_z) = \frac{1}{|S|}(\mathbb{1}_z(S) + \mathbb{1}_{f^{-1}(\{1\})}(S)) = \frac{1}{|S|}\left(\mathbb{1}_z(S) + \mathbb{1}_1(f(S))\right) \tag{1}$$

We also compute

$$L_S(h^-) = \frac{1}{|S|}|\{x \in S : f(x) \neq 0\}| = \frac{1}{|S|}\mathbb{1}_1(f(S)).$$

If $f^{-1}(\{1\}) \cap S \neq \emptyset$, let $z^* \in S$ be such that $f(z^*) = 1$. Then note that $L_S(h_{z^*}) = 0$, so we have $h_{z^*} \in \operatorname{argmin}_{h \in \mathcal{H}} L_S(h)$. If $f^{-1}(\{1\}) \cap S = \emptyset$, then $L_S(h^-) = 0$ whence $h^- \in \operatorname{argmin}_{h \in \mathcal{H}} L_S(h)$. Hence, defining

$$A(S) := \begin{cases} h_{z^*} & \text{if } f^{-1}(\{1\}) \cap S = \{z^*\} \\ h^- & \text{if } f^{-1}(\{1\}) \cap S = \emptyset, \end{cases}$$

gives us a learning algorithm which implements the ERM rule for learning $\mathcal{H} = \mathcal{H}_{\text{Singleton}}$.

□

2 Problem 2

Let $\mathcal{X} = \mathbb{R}^2$, $\mathcal{Y} = \{0, 1\}$, and let $\mathcal{H}$ be the class of concentric circles in the plane, that is, $\mathcal{H} = \{h_r : r \in \mathbb{R}_+\}$, where $h_r(x) = \mathbb{1}_{\{\|x\| \leq r\}}$. Design an ERM algorithm to learn $\mathcal{H}$ and explain why it is ERM.

Proof. By realizability, there is some $r^* > 0$ such that

$$0 = L_{(\mathcal{D}, f)}(h_{r^*}) = \mathcal{D}(\{x \in \mathbb{R}^2 : f(x) \neq \mathbb{1}_{B_{r^*}(0)}(x)\})$$

whence

$$1 = \mathcal{D}(\{x \in \mathbb{R}^2 : f(x) = \mathbb{1}_{B_{r^*}(0)}(x)\}).$$

Since this occurs with probability 1, we may as well assume $f = \mathbb{1}_{Br^*(0)}$.

Choose the minimal $s \geq 0$ such that $S \cap f^{-1}(\{1\}) \subseteq B_s(0)$ and set $A(S) := h_s$. Note that, as $S \cap f^{-1}(\{1\}) \subseteq B_{r^*}(0)$, by minimality we have $s \leq r^*$. Using that all nonzero labelled examples lie inside $B_s(0)$, we compute

$$\begin{aligned} L_S(h_s) &= \frac{1}{|S|} |\{x \in S : h_s(x) \neq f(x)\}| = \frac{1}{|S|} |\{x \in S : \mathbb{1}_{B_s(0)}(x) \neq \mathbb{1}_{B_{r^*}(0)}(x)\}| \\ &= \frac{1}{|S|} |S \cap (B_{r^*}(0) \,\Delta\, B_s(0))| = \frac{1}{|S|} |S \cap (B_{r^*}(0) \setminus B_s(0))| = 0. \end{aligned}$$

Hence, the algorithm $A(S)$ is ERM.

□

3 Problem 3

Let $\mathcal{H}$ be a hypothesis class of binary classifiers. Show that if $\mathcal{H}$ is agnostic PAC learnable, then $\mathcal{H}$ is PAC learnable as well. Furthermore, if A is a successful agnostic PAC learner for $\mathcal{H}$, then A is also a successful PAC learner for $\mathcal{H}$.

Proof. Suppose $\mathcal{H}$ is agnostic PAC learnable and let $A_{\text{ag}}(S)$ and $m_{\mathcal{H}} : (0, 1)^2 \to \mathbb{N}$ be the corresponding agnostic algorithm and agnostic sample complexity. Fix $\varepsilon, \delta \in (0, 1)^2$, a distribution $\mathcal{D}$ on $\mathcal{X}$, and a labeling function $f : \mathcal{X} \to \{0, 1\}$ such that $(\mathcal{H}, \mathcal{D}, f)$ is realizable. Define a probability measure (distribution) $\mathcal{F}$ on $X \times \{0, 1\}$ by

$$d\mathcal{F}(x, y) = d\delta_{f(x)}(y)\, d\mathcal{D}(x)\,,$$

equivalently

$$\int_{X \times Y} g(x, y)\, d\mathcal{F} = \int_X \int_{\{0,1\}} g(x, y)\, d\delta_{f(x)}(y)\, d\mathcal{D}(x) \quad \text{for all Borel } g : X \times \{0, 1\} \to \mathbb{C}.$$

To begin, we compute for any hypothesis $h \in \mathcal{H}$ that

$$\begin{aligned}
\mathop{\mathbb{P}}_{(x,y)\sim\mathcal{F}}[h(x) \neq y] &= \int_{X\times Y} \mathbb{1}_{\{h(x)\neq y\}}(x,y)\, d\mathcal{F}(x,y) \\
&= \int_X \int_{\{0,1\}} \mathbb{1}_{\{h(x)\neq y\}}(x,y)\, d\delta_{f(x)}(y)\, d\mathcal{D}(x) \\
&= \int_X \mathbb{1}_{\{(\tilde{x},\tilde{y}):h(\tilde{x})\neq\tilde{y}\}}(x, f(x))\, d\mathcal{D}(x) \\
&= \mathcal{D}(\{x \in \mathcal{X} : h(x) \neq f(x)\}) = L_{(\mathcal{D},f)}(h).
\end{aligned}$$

Now, since $(\mathcal{H}, \mathcal{D}, f)$ is realizable, it follows that

$$\inf_{h\in\mathcal{H}} \mathop{\mathbb{P}}_{(x,y)\sim\mathcal{F}}[h(x) \neq y] = \inf_{h\in\mathcal{H}} L_{(\mathcal{D},f)}(h) = 0.$$

Then, by agnostic PAC learnability, running A_{ag} on $m \geq m_{\mathcal{H}}(\varepsilon, \delta)$ i.i.d. examples generated by $\mathcal{F}$ returns a hypothesis $A_{\text{ag}}(S)$ such that

$$\begin{aligned}
1 - \delta &\leq \mathop{\mathbb{P}}_{S\sim\mathcal{F}^m}[L_{\mathcal{F}}(A_{\text{ag}}(S)) \leq \inf_{h\in\mathcal{H}} L_{\mathcal{F}}(h) + \varepsilon] \\
&= \mathop{\mathbb{P}}_{S\sim\mathcal{F}^m}\left[\mathop{\mathbb{P}}_{(x,y)\sim\mathcal{F}}[A_{\text{ag}}(S)(x) \neq y] \leq \inf_{h\in\mathcal{H}} \mathop{\mathbb{P}}_{(x,y)\sim\mathcal{F}}[h(x) \neq y] + \varepsilon\right] \\
&= \mathop{\mathbb{P}}_{S\sim\mathcal{F}^m}\left[\mathop{\mathbb{P}}_{(x,y)\sim\mathcal{F}}[A_{\text{ag}}(S)(x) \neq y] \leq \varepsilon\right] \\
&= \mathop{\mathbb{P}}_{S\sim\mathcal{F}^m}\left[L_{(\mathcal{D},f)}(A_{\text{ag}}(S)) \leq \varepsilon\right] = \mathop{\mathbb{P}}_{S\sim\mathcal{D}^m}[L_{(\mathcal{D},f)}(A_{\text{ag}}(S)) \leq \varepsilon]
\end{aligned}$$

This shows that $\mathcal{H}$ is PAC learnable and also that $A = A_{\text{ag}}$ is a successful PAC learner for $\mathcal{H}$.

$\square$