Exercises
Advanced Machine Learning
Fall 2019

Series 8, 16 Dec 2019 (PAC learning)

Machine Learning Laboratory
Dept. of Computer Science, ETH Zürich
Prof. Joachim M. Buhmann
Web https://ml2.inf.ethz.ch/courses/aml/

Teaching assistant: **Viktor Gal**
viktor.gal@inf.ethz.ch

Solution 1 (PAC learnability):

By the first inequality, if

$$\mathcal{R}\left(\hat{c}_n\right) - \inf_{c\in\mathcal{C}} \mathcal{R}\left(c\right) > \epsilon,$$

then

$$\sup_{c\in\mathcal{C}} \left|\hat{\mathcal{R}}_n\left(c\right) - \mathcal{R}\left(c\right)\right| > \frac{\epsilon}{2}.$$

Therefore,

$$\begin{aligned}\mathbf{P}\left(\mathcal{R}\left(\hat{c}_n\right) - \inf_{c\in\mathcal{C}} \mathcal{R}\left(c\right) > \epsilon\right) &\le \mathbf{P}\left(\sup_{c\in\mathcal{C}} \left|\hat{\mathcal{R}}_n\left(c\right) - \mathcal{R}\left(c\right)\right| > \frac{\epsilon}{2}\right)\\ &\le 2\left|\mathcal{C}\right| \exp\left(-\frac{2n\epsilon^2}{4}\right).\end{aligned}$$

Hence, if $n \to 0$, then $\mathbf{P}\left(\mathcal{R}\left(\hat{c}_n\right) - \inf_{c\in\mathcal{C}} \mathcal{R}\left(c\right) > \epsilon\right) \to 0$.

Solution 2 (PAC learnability of intervals):

1. Note that $\ell_\epsilon^+ \le X_{\min}^n$ iff $X_i \notin [\ell^*, \ell_\epsilon^+)$, for any $i \le n$. Therefore,
$$\begin{aligned}\mathbf{P}\left(\ell_\epsilon^+ \le X_{\min}^n\right) &= \left(1 - \mathbf{P}\left(\ell^* \le X_i < \ell_\epsilon^+\right)\right)^n\\ &= \left(1-\epsilon\right)^n.\end{aligned}$$

2. By the hint and the previous task, $\mathbf{P}\left(\ell_\epsilon^+ \le X_{\min}^n\right) = \left(1-\epsilon\right)^n \le \exp\left(-n\epsilon\right)$. Now, $n > \frac{1}{\epsilon}\log\left(\frac{1}{\delta}\right)$ implies that $\exp\left(-n\epsilon\right) < \delta$, which yields the result.

3. PAC learnability: Let $\mathcal{A}$ be the algorithm that computes $X_{\min}^n$ and outputs $\hat{c}_n$. If $n = \lceil\frac{1}{\epsilon}\log\left(\frac{1}{\delta}\right)\rceil$, then the output satisfies
$$\mathbf{P}\left(\mathcal{R}\left(\hat{c}_n\right) > \epsilon\right) < \delta.$$
This inequality follows from the hint: $\mathbf{P}\left(\mathcal{R}\left(\hat{c}_n\right) > \epsilon\right) = \mathbf{P}\left(\ell_\epsilon^+ \le X_{\min}^n\right)$, which is proven as follows:
$$\begin{aligned}\mathcal{R}\left(\hat{c}_n\right) > \epsilon &\Leftrightarrow \mathbf{P}\left(\hat{c}_n\left(X_0\right) \neq Y_0\right) > \epsilon\\ &\Leftrightarrow \mathbf{P}\left(\mathbb{I}_{\left[X_{\min}^n,\infty\right)}\left(X_0\right) \neq \mathbb{I}_{\left[\ell^*,\infty\right)}\left(X_0\right)\right) > \epsilon\\ &\Leftrightarrow \mathbf{P}\left(\ell^* \le X_0 \le X_{\min}^n\right) > \epsilon\\ &\Leftrightarrow \ell_\epsilon^+ \le X_{\min}^n.\end{aligned}$$

Efficient PAC learnability: Observe that the algorithm $\mathcal{A}$ runs in $O(n)$-time and when $n \geq \lceil \frac{1}{\epsilon} \log\left(\frac{1}{\delta}\right) \rceil$, we have that $\mathbf{P}\left(\mathcal{R}\left(\hat{c}_n\right) \leq \epsilon\right) \geq 1 - \delta$. As $\lceil \frac{1}{\epsilon} \log\left(\frac{1}{\delta}\right) \rceil$ is polynomial in $1/\epsilon$ and $1/\delta$, we conclude that $\mathcal{C}$ is efficiently PAC learnable.