COMS 4995 (Randomized Algorithms): Exercise Set #8

For the week of October 28–November 1, 2019

Instructions:

(1) *Do not turn anything in.*

(2) The course staff is happy to discuss the solutions of these exercises with you in office hours or in the course discussion forum.

(3) While these exercises are certainly not trivial, you should be able to complete them on your own (perhaps after consulting with the course staff or a friend for hints).

Exercise 36

In the *realizable* special case of PAC learning, the ground truth function $f : \mathbb{R}^d \to \{0, 1\}$ belongs to the hypothesis class $\mathcal{H}$. In this case, there will always be a hypothesis in $\mathcal{H}$ with zero training error (if nothing else, f itself), and so the ERM algorithm will return an (arbitrary) hypothesis with zero training error. The worry is that the algorithm will be tricked into outputting a hypothesis with large generalization error that just happened to make no errors on the training data.

Suppose $\mathcal{H}$ is finite. Prove that once

$$m \geq \frac{c}{\epsilon}\left(\ln\frac{1}{\delta} + \ln|\mathcal{H}|\right)$$

for a sufficiently large constant $c > 0$, with probability at least $1 - \delta$ (over a size-m sample from the unknown distribution D), the ERM algorithm outputs a hypothesis with generalization error at most ϵ.[1]

Exercise 37

Recall from Lecture #15 the growth function $G_{\mathcal{H}}(\cdot)$ of a hypothesis class $\mathcal{H}$, with

$$G_{\mathcal{H}}(m) := \max_{S\,:\,|S|=m} |\{h|_S \,:\, h \in \mathcal{H}\}|\,.$$

In other words, $G_{\mathcal{H}}(m)$ is the maximum number of distinct restrictions that the functions of $\mathcal{H}$ induce on any size-m sample.

Recall also from Lecture #15 that we proved a uniform convergence bound parameterized by the growth function of $\mathcal{H}$: as long as

$$m \geq \frac{c_1}{\epsilon^2}\left(\ln\frac{1}{\delta} + \ln|G_{\mathcal{H}}(2m)|\right) \tag{1}$$

for a sufficiently large constant c_1, with high probability over a random sample S of size m from the unknown distribution D, for every $h \in \mathcal{H}$, the training error of h on S is within $\pm\epsilon$ of the generalization error of h with respect to D.

Prove that if $|G_{\mathcal{H}}(m)| = O(m^d)$, then the inequality (1) holds once

$$m \geq c_2 \cdot \frac{d}{\epsilon^2}\ln\left(\frac{d}{\epsilon\delta}\right)$$

for a sufficiently large constant c_2.

[1]Note the key point: in the realizable case, the dependence of the sample complexity of PAC learning on $\frac{1}{\epsilon}$ improves from quadratic to linear.

Exercise 38

Let $\mathcal{H}$ be the set of linear classifiers in the plane. (I.e., an $h \in \mathcal{H}$ is specified by $a, b, c \in \mathbb{R}$, and labels points $(x, y) \in \mathbb{R}^2$ either 1 or 0 according to whether $ax + by + c \geq 0$ or not.) Prove that $\mathcal{G}_{\mathcal{H}}(m) = O(m^2)$.

Exercise 39

Extend your argument in the previous exercise to obtain a bound on the growth rate function of linear classifiers in $\mathbb{R}^d$ for arbitrary $d \in \mathbb{N}$.