MATH/COMP 562 ASSIGNMENT 1
DUE: FEB 2ND (THURSDAY)
VERSION: January 31, 2023

ADAM M. OBERMAN

Instructions. Submit your solutions on MyCourses course page. I prefer proofs to be handwritten, but Latex typesetting is allowed. You can get help from other students, but you should do the write up yourself. Code solutions can be a PDF of a workbook, you can use any programming language, but NumPy is recommended.

Changes:

(1) This version includes a link to sample code.
(2) Updated deadline

PAC learning and Limits in Machine Learning.

Exercise 1.1 (PAC Learning of higher dimensional rectangles). Do Mohri Exercise 2.2.

Exercise 1.2 (PAC learning definition restated as expected loss). Do Shalev-Shwartz Exercise 4.1. Hint: assume L takes values in $[0, 1]$. If $\mathbb{P}(L > \epsilon) < \delta)$ what does this say about $\mathbb{E}[L]$ in terms of ϵ and δ?

Exercise 1.3 (Hoeffding's Inequality and Bounds on Generalization Gap). Do Shalev-Shwartz Exercise 4.2. Hint: you are going through the proof of the result, applying Hoeffding's inequality with $(b - a)$ instead of 1.

Exercise 1.4. Prove Mohri Theorem 2.13. (Hint: you can follow the proof in the book, but make sure you can do this kind of argument).

Exercise 1.5 (Non-uniform convergence). Suppose $a \in [0, .5)$ and consider a parameterized family of hypotheses

$$\mathcal{H} = \{h_w : \mathcal{X} \to \mathcal{Y}, y = h_w(x) \mid w \in (a, 1)\},$$

Assume the following bound holds. For all $\delta > 0$,

$$|\widehat{L}_{S^m}(h_w) - L(h_w)| \leq \frac{1}{\sqrt{m}} \left(\log(1/\delta) + \frac{1}{w} \right), \quad \text{with probability} \geq 1 - \delta$$

(a) Use the assumption above to show that for each h_w, the empirical loss converges to the expected loss as $m \to \infty$.
(b) Show that, if $a > 0$, the convergence is uniform in w (in other words, there is an $m = m(\epsilon, \delta)$ which does not depend on w).
(c) What happens if $a = 0$? Is the convergence uniform? Just explain why

Exercise 1.6 (Rewriting the probability bound in terms of the generalization gap). Let $p \geq 1$ and $C > 0$ be constants. Starting from the estimate

$$\mathbb{P}[|\widehat{L}_S(h) - L(h)| > \epsilon] \leq C \exp(-2m\epsilon^p)$$

Find the expression, RHS, expressed as a function of m, C, p, δ, which makes the following expression equivalent. For any $\delta \geq 0$,

$$|\widehat{L}_S(h) - L(h)| \leq RHS, \quad \text{with probability} \geq 1 - \delta$$

Coding exercises on PAC learning bounds. Sample Code for exercises (from Viet). Instructions: clone a copy into you own Google Drive using "File – Save a Copy in Drive" `https://colab.research.google.com/drive/1EKH0sNtSB9rFFl7yNPcbXp2P0iPCLn usp=sharing`

Exercise 1.7. Consider the PAC learning bounds for the axis-aligned rectangles problem (Mohri Example 2.4). Note the learning bound is of the form

$$R(h_S) \leq \frac{4}{m} \log(4/\delta), \quad \text{with probability} \geq 1 - \delta$$

Here m is the number of samples in the dataset S, and $R(h_S)$ is the (true) expected error. Perform a simulation of the problem, and measure the error. Do this for a number of values of m, and run 100 experiments for each m. Recall the expected (not empirical) error is given in terms of the area between the true and empirical error. Plot the expected error (which is random) as a function of m. Show the error bands, and compare to the estimate. (It's okay if the estimate is not tight).

Exercise 1.8. Redo the previous problem, but for a higher dimensional rectangle, as in Mohri Exercise 2.2. In addition to the previous curve, plot curves in dimension 3, 4, and 5. Discuss the dependance on dimension.