Mathematics of Machine Learning
Homework 12

Instructors: Afonso S. Bandeira & Nikita Zhivotovskiy
Course Coordinator: Pedro Abdalla Teixeira

May 21, 2021

Try to solve the questions before looking to the answers. Every item must be proved rigorously. Starred problems are harder.

Problem 1

Let $L^* := \inf_{f:X\to\{0,1\}} \mathbb{P}(f(X) \neq Y)$ denotes the Bayes error and let f^* be the optimal Bayes classifier.

(a) Write L^* in terms of $\eta(x) = \mathbb{P}(Y = 1|X = x)$.

(b) Let $\tilde{\eta}(x)$ be a nonnegative function that ε-approximates $\eta(x)$ in the $L1$-sense, precisely $\mathbb{E}|\eta(X) - \tilde{\eta}(X)| \leq \varepsilon$. Prove that the classifier f defined as

$$f(x) := \begin{cases} 1, \tilde{\eta}(x) \geq \frac{1}{2} \\ 0, \text{otherwise} \end{cases}$$

satisfies the following error bound

$$\mathbb{P}(f(X) \neq Y) - L^* \leq 2\varepsilon.$$

Problem 2

In the lecture, we proved that the Halving algorithm does not make more than $\log_2 |\mathcal{F}|$ mistakes, where $|\mathcal{F}|$ denotes the size of the class. Prove that this bound is tight in the following sense: for any integer n there is a class of size 2^n such that halving algorithms makes at least n mistakes when run on a particular sequence.

Problem 3

Prove that the Halving algorithm determines a sample compression scheme of finite size for any finite class. Give an upper bound on its sample complexity in the PAC learning setup. Does the output classifier belong to $\mathcal{F}$?