Mathematics of Machine Learning
Homework 13

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

June 4, 2021

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

Problem 1

In the lecture notes, we proved that, in the realizable case, the Halving algorithm makes at most $O(\log_2 |\mathcal{F}|)$ mistakes where $\mathcal{F}$ is the class of classifiers. Modify Halving algorithm for the non-realizable case when there exits a classifier f^* that makes only m mistakes.

Problem 2

Suppose that we have a learning algorithm with regret R_T bounded by $\frac{1}{\eta} + \eta T$. If the time horizon was known, then we would choose $\eta = \sqrt{\frac{1}{T}}$ to minimize the regret bound as we did at the end of the proof of Theorem 21.2 in the lecture notes. The doubling trick is a technique that allows the learner to have a similar regret bound without knowing the time horizon. It goes as follows: Divide the time into intervals, i.e, for every $k \in \mathbb{N}$, the k-th interval is $2^k, \dots, 2^{k+1} - 1$. We run the algorithm in the beginning of each interval by setting $\eta_k := 2^{-k/2}$. Prove that the learning algorithm, implemented with the doubling trick, cannot have a regret R_2^T worse than $\frac{\sqrt{2}}{\sqrt{2}-1} R_T$.

Problem 3

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a differentiable function. Recall that the directional derivative of f in the direction of the vector v is defined as

$$\nabla_v f(x) := \lim_{t \longrightarrow 0} \frac{f(x + tv) - f(x)}{t}.$$

(a) Prove that the directional derivative of f in the direction of the vector v can be written in terms of the inner product between v and the gradient of f.

(b) Use letter "a" to justify the sentence: " The gradient of a function points out to the direction in which the function increases".

(c) Prove that if f is convex and if the gradient at the point x_o is zero, then x_0 is a global minimum, i.e, $f(x_0) \le f(x)$ for all points x in the domain of f.