Mehryar Mohri
Foundations of Machine Learning 2024
Courant Institute of Mathematical Sciences
Homework assignment 1
Sep 17, 2024
Due: October 01, 2024

A Probability tools

1. Let $f\colon (0, +\infty) \to \mathbb{R}_+$ be a function that admits an inverse f^{-1}, and let X be a random variable. Suppose that for all $t > 0$, the probability that X exceeds t is bounded by $f(t)$, i.e., $\mathbb{P}[X > t] \leq f(t)$. Prove that for any $\delta > 0$, with probability at least $1 - \delta$, the random variable X satisfies $X \leq f^{-1}(\delta)$.

2. Let Z be a discrete random variable that takes non-negative integer values. Prove that $\mathbb{E}[Z] = \sum_{n \geq 1} \mathbb{P}[Z \geq n]$. *Hint: express* $\mathbb{P}[Z = n]$ *as* $\mathbb{P}[Z \geq n] - \mathbb{P}[Z \geq n+1]$.

B Label bias

1. Let $\mathcal{D}$ be a distribution over $\mathcal{X}$, and let $f\colon \mathcal{X} \to \{-1, +1\}$ be a labeling function. Our goal is to approximate the label bias of the distribution $\mathcal{D}$, denoted by p_+, which is defined as:

$$p_+ = \underset{x \sim \mathcal{D}}{\mathbb{P}}[f(x) = +1].$$

Let S be a labeled sample of size m, drawn i.i.d. from $\mathcal{D}$. Using S, derive an estimate $\hat{p}_+$ of p_+. Show that for any $\delta > 0$, with probability at least $1 - \delta$, the following inequality holds:

$$|p_+ - \hat{p}_+| \leq \sqrt{\frac{\log(2/\delta)}{2m}}.$$

Justify each step of the proof carefully.

C Learning in the presence of noise

1. In Lecture 2, we showed that the concept class of axis-aligned rectangles is PAC-learnable. Consider now the case where the training points received by the learner are subject to the following noise: points negatively labeled are unaffected by noise but the label of a positive training point is randomly flipped to negative with probability $\eta \in (0, \frac{1}{2})$. The exact value of the noise rate η is not known to the learner but an upper bound η' is supplied to him with $\eta \leq \eta' < 1/2$. Show that the algorithm described in class returning the tightest rectangle containing positive points can still PAC-learn axis-aligned rectangles in the presence of this noise. To do so, you can proceed using the following steps:

   (a) Using the notation of the lecture slides, assume that $\mathbb{P}[\mathrm{R}] > \epsilon$. Suppose that $R(\mathrm{R}') > \epsilon$. Give an upper bound on the probability that R' misses a region r_j, $j \in [1, 4]$ in terms of ϵ and η'?

   (b) Use that to give an upper bound on $\mathbb{P}[R(\mathrm{R}') > \epsilon]$ in terms of ϵ and η' and conclude by giving a sample complexity bound.