CSC 588: Homework 2

Chicheng Zhang

February 4, 2022

Please complete the following exercises and read the following instructions carefully.

- Your solutions to these problems will be graded based on both correctness and clarity. Your arguments should be clear: there should be no room for interpretation about what you are writing. Otherwise, I will assume that they are wrong, and grade accordingly.
- If you feel unable to make progress on any of the questions, you can post your questions on Piazza. Try posing your questions to be as general as possible, so that it can promote discussion among the class.
- You are encouraged to discuss the homework questions with your classmates, but the discussions should only be at a high level, and you should write your solutions in your own words. For every question you have had discussions on, please mention explicitly whom you have discussed with; otherwise it may be counted as academic integrity violation.
- For detailed homework policies, please read the course syllabus, available on the course website.

This homework is due on Feb 15, 2022, 5pm MST, on gradescope.

Problem 1 (10pts)

Solve the following exercises from the lectures:

1. Given a random variable X with expectation μ, prove that for any $w \in \mathbb{R}$,

$$\mathbb{E}\left[(X-w)^2\right] = \mathbb{E}\left[(X-\mu)^2\right] + (\mu - w)^2.$$

   (Note that this implies that $\mu = \operatorname{argmin}_{w \in \mathbb{R}} \mathbb{E}\left[(X-w)^2\right]$.)

2. Suppose $\lambda \in \mathbb{R}$, and we have a continuous random variable Y with probability density function (pdf) $p_Y(y)$, and we let another random variable Z have the following probability density function:

$$P_Z(y) = \frac{P_Y(y) e^{\lambda y}}{\int_{\mathbb{R}} P_Y(y) e^{\lambda y} dy};$$

   Prove that

$$\mathbb{E}[Z] = \frac{\mathbb{E}[e^{\lambda Y} Y]}{\mathbb{E}[e^{\lambda Y}]}, \quad \mathbb{E}\left[Z^2\right] = \frac{\mathbb{E}[e^{\lambda Y} Y^2]}{\mathbb{E}[e^{\lambda Y}]}.$$

3. In the analysis of the closure algorithm for PAC learning rectangles, we defined "boundary region" R_j's, $j = 1, \ldots, 4$, each of which has probability exactly $\frac{\epsilon}{4}$. Suppose S is a set of m iid training examples drawn iid from distribution D. Define event E as

$$E = \left\{ \text{ for every } j = 1, \ldots, 4, \ S \text{ has some example in } R_j \ \right\}$$

   Show that if $m \ge \frac{4}{\epsilon} \ln \frac{4}{\delta}$, then $\mathbb{P}(E) \ge 1 - \delta$. (Hint: define $E_j = \left\{ \ S \text{ has some example in } R_j \ \right\}$; try to calculate the probabilities of E_j's exactly.)

Problem 2 (10pts)

Consider the cost-sensitive classification (CSC) problem: each example is represented by a pair (z, c), where $z \in \mathcal{Z}$ is its feature part, and $c = (c(1), ..., c(K)) \in [0, M]^K$ is its cost vector, where for $k \in [K] := \{1, \ldots, K\}$, $c(k)$ represents the cost of predicting the example with class k. Given a classifier $h : \mathcal{Z} \to [K]$, the cost of h on example (z, c) is defined as $c(h(z))$. The performance of a classifier h on a distribution D is measured by its expected cost $L(h, D) = \mathbb{E}_{(z,c)\sim D}\left[c(h(z))\right]$.

1. Given a binary classification example (x, y) (where $y \in \{1, 2\}$), how can we construct a CSC example (z, c), such that $\mathbf{1}(h(x) \neq y)$, the 0-1 classification error indicator of a binary classifier h on (x, y), equals its cost on (z, c)?

2. Suppose we have a finite hypothesis class $\mathcal{H}$, and are given a set S of m CSC examples $(z_1, c_1), \ldots, (z_m, c_m)$ drawn iid from D. Define the ERM as $\hat{h} = \text{argmin}_{h\in\mathcal{H}} L(h, S)$, where $L(h, S) = \frac{1}{m}\sum_{i=1}^{m} c_i(h(z_i))$. For any $\delta > 0$, give an upper bound on $L(\hat{h}, D) - \min_{h'\in\mathcal{H}} L(h', D)$, that holds with probability $1 - \delta$ and goes to zero with $m \to \infty$. Justify your answer.

Problem 3 (10pts)

Suppose we have an algorithm $\mathcal{B}$ that learns hypothesis class $\mathcal{H}$ in the following sense. There exists a function $g : (0, 1) \to \mathbb{N}$, such that for any distribution D, for any $\epsilon > 0$, if $\mathcal{B}$ draws $m \geq g(\epsilon)$ iid training examples from D, then with probability at least $\frac{1}{10}$, $\mathcal{B}$ returns a classifier $\hat{h}$ whose excess generalization error on D $(\text{err}(\hat{h}, D) - \min_{h\in\mathcal{H}} \text{err}(h, D))$ is at most ϵ.

Now, given $\mathcal{B}$, and the ability to draw fresh training examples from the data distribution, how can you design an algorithm $\mathcal{A}$ that (ϵ, δ)-agnostic PAC learns $\mathcal{H}$ for any $\epsilon, \delta > 0$? What is your $\mathcal{A}$'s sample complexity? (Hint: you may want to design $\mathcal{A}$ so that it calls $\mathcal{B}$ multiple times.)

Problem 4 (10pts)

1. Let $\mathcal{H}$ be the class of signed intervals in $\mathbb{R}$, that is, $\mathcal{H} = \left\{h_{a,b,s}, a \leq b, s \in \{-1, 1\}\right\}$, where

$$h_{a,b,s}(x) = \begin{cases} s, & x \in [a, b], \\ -s, & x \notin [a, b]. \end{cases}$$

   What is the VC dimension of $\mathcal{H}$?

2. Let $n \in \mathbb{N}$. Define the class of univariate polynomial threshold functions

$$\mathcal{H}_n = \left\{\mathbf{1}(p(x) > 0) : p \text{ is a polynomial of degree} \leq n\right\}$$

   (where $x \in \mathbb{R}$). What is the VC dimension of $\mathcal{H}$?

Problem 5 (2pts)

1. How much time did it take you to complete this homework?

2. Do you have any suggestions for this course so far (e.g. homework length / easiness, lecture pace, etc)?