Homework 8
Due: Apr 25, 2019 at 2:30pm

Please follow these guidelines when writing up your CS 155 homeworks.
A major goal of this class is that you learn how to mathematically analyze probabilistic processes and events. Hence, make sure your proofs are rigorous, i.e., that every step is adequately justified. Try to keep your proofs as simple and as concise as possible, keeping in mind that you will potentially have to examine multiple proof strategies in order to achieve this. Make sure to CLEARLY STATE YOUR ASSUMPTIONS at the beginning of any solution, when necessary. As we are doing probabilistic analysis, we shall require that you clearly define the probability space you are working with and identify the events within it that you will analyze.

You are allowed to discuss these problems with other students, but you have to leave any meeting with other students without any written material about the problems, and the writeups must be done by yourself without help from others. If you have any questions regarding how these guidelines apply to a particular problem or what they mean in general, please email the TAs.

Problem 1

Properties of Rademacher averages. In this question, we prove several convenient properties of Rademacher averages. They are mostly uninteresting in isolation, but can be used to construct bounds on complicated function families.

a. Linearity: For any $\alpha, \beta \in \mathbb{R}$, $F \subseteq \mathcal{X} \to \mathbb{R}$, take $\alpha F + \beta = \{\alpha f(\cdot) + \beta : f \in F\}$. Let $z$ be any sample of size $m$. Show that

$$\hat{R}_m(\alpha F + \beta, z) = |\alpha| \hat{R}_m(F, z).$$

Hint: the multiplicative and additive aspects of this problem can be handled separately.

b. Monotonicity: Suppose $F_1, F_2 \subseteq \mathcal{X} \to \mathbb{R}$. Show that

$$\hat{R}_m(F_1, z) \leq \hat{R}_m(F_1 \cup F_2, z).$$
Problem 2

Absolute Symmetrization. Some authors use an alternative definition of the Rademacher average, which takes an absolute value inside the supremum. Taking $z \sim D^m$, and $\mathcal{F} \subseteq \mathcal{X} \rightarrow \mathbb{R}$, prove the following symmetrization inequality without using the standard Rademacher symmetrization inequality:

$$E_z \left[ \sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{i=1}^{m} f(z_i) - E[f] \right| \right] \leq 2E_z,\sigma \left[ \sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{i=1}^{m} \sigma_i f(z_i) \right| \right].$$

Hint: Sometimes the less you know about an object, the easier it is to prove something about it. The above result can be shown not just for the absolute value, but for any function that is convex and subadditive (a function $g$ is subadditive if $g(a+b) \leq g(a) + g(b)$). If you get stuck, you may consult a proof of the standard symmetrization inequality, theorem 14.20 in the textbook, for guidance.

Problem 3

Monte-Carlo Estimation of Rademacher Averages. The definition of the empirical Rademacher average contains an expectation over an exponentially large space of possible $\sigma$. In this problem, we show that Monte-Carlo estimation of this expectation is very accurate.

Suppose $\sigma$ is drawn i.i.d. Rademacher, and define the following Monte-Carlo estimator of the empirical Rademacher average

$$\hat{R}_m^1(\mathcal{F}, z, \sigma) = \sup_{f \in \mathcal{F}} \frac{1}{m} \sum_{i=1}^{m} \sigma_i f(z_i).$$

Hint: If you get stuck, consult theorem 14.21 in the textbook for guidance.

a. Show that for any $z \sim D^m$, $\mathcal{F} \subseteq \mathcal{X} \rightarrow [-1, 1]$,

$$\Pr \left( \hat{R}_m(\mathcal{F}, z) \geq \hat{R}_m^1(\mathcal{F}, z, \sigma) + \epsilon \right) \leq e^{-m\epsilon^2/2} .$$

b. Now show that

$$\Pr \left( \sup_{f \in \mathcal{F}} \left( E[f] - \frac{1}{m} \sum_{i=1}^{m} f(z_i) \right) \geq 2\hat{R}_m^1(\mathcal{F}, z, \sigma) + 5\epsilon \right) \leq 3e^{-m\epsilon^2/2} ,$$
and then conclude that for $0 < \delta < 1$,

$$\Pr_{\sigma,z} \left( \sup_{f \in \mathcal{F}} \left( E[f] - \frac{1}{m} \sum_{i=1}^{m} f(z_i) \right) \leq 2 \hat{R}^1_m(\mathcal{F}, z, \sigma) + 5 \sqrt{\frac{2 \ln \left( \frac{2}{\delta} \right)}{m}} \right) \geq 1 - \delta .$$

**Problem 4**

We can apply Theorem 14.21 to compute a bound on the sample complexity of agnostic learning a binary classification. Suppose we have a domain $X$ and a concept class $C$ with VC dimension $d$.

a. For each hypothesis $h \in C$, we define a function $f_h(x)$ on $X$ as follows: $f_h(x) = 1$ if $h$ correctly classifies $x$, and $-1$ otherwise. (This is the same definition as in section 14.6.3 of the textbook.) Find a sample size $m_1$ such that the empirical Rademacher average of $\mathcal{F} = \{f_h\}_{h \in C}$ is at most $\epsilon/4$ for any sample.

*Hint:* the bound on the empirical Rademacher average in section 14.6.3 might help.

b. Use Theorem 14.21 together with part a to find a sample size $m$ such that with probability $1 - \delta$, the expectations of all the functions in $\mathcal{F}$ are estimated within error $\epsilon$.

c. Compare your bound to the bound you would obtain using VC dimension (i.e., the $\epsilon$-sample theorem).