Homework 5
Due: Mar 14, 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
Consider the range space \((X, R)\), where \(X\) is an interval \([a, b]\) and \(R\) is the collection of all intervals \([x, y]\) \(\subseteq [a, b]\).

Let \(S = \{x_1, \ldots, x_n\}\) be a finite set of points in \(X\), where \(x_1 < x_2 < \cdots < x_n\), and \(|S| = n\).

a. Which sets are in \(\Pi_R(S)\) (the projection of \(R\) on \(S\))?

b. Compare the number of sets you found in part a to the bound obtained using Sauer’s Lemma.

Problem 2
A function \(f : \{0, 1\}^n \rightarrow \{0, 1\}\) is symmetric if its value is uniquely determined by the number of 1’s in its input.
Let $X = \{0, 1\}^n$, and let $\mathcal{C}$ be the concept class defined by the set of all symmetric functions on $\{0, 1\}^n$. (That is, if $f$ belongs to $\mathcal{C}$, then a point $x \in X$ is classified as positive by $f$ if $f(x) = 1$.)

a. What is the VC-dimension of $(X, \mathcal{C})$?

b. Consider the uniform distribution over $X$: each of the $n$ bits is chosen independently at random with $Pr(1) = Pr(0) = 1/2$. The theorem we proved in class claims that we can $(\epsilon, \delta)$-PAC learn $\mathcal{C}$ on this distribution with a training set of size $m = O\left(\frac{n}{\epsilon} \ln\left(\frac{n}{\epsilon}\right) + \frac{1}{\epsilon} \ln\left(\frac{1}{\delta}\right)\right)$. Show that for any fixed value $\gamma > 0$, if $n$ is large, then this training set is unlikely to contain any points in $X$ with less than $(1 - \gamma)^n$ 1’s or more than $(1 + \gamma)^n$ 1’s. Why does this fact does not contradict the theorem?