Assignment 1

Due Jan 31, beginning of class

Instructions. Apart from correctness, clarity and conciseness of your solutions are important. Please write your solutions legibly.

You may collaborate in groups of two or three, but you must write your solution on your own and should list the name of the persons with whom you collaborated. If you need to, you are welcome to come to my office hours to discuss the problems; I may give you some hints. Solutions to a number of problems here can be found in books, research papers, or elsewhere on the web. Please do not refer to these to solve the problems. If you happen to have seen a solution please mention this in your work. If you think there is an error in some problem or need a clarification, please let me know (`navin001@gmail.com`). The problem set is presently incomplete and I will be adding new problems. Full problem set will be posted by Jan 18.

Problem 1. *This problem asks you to prove some properties of Rademacher complexity.*

1. *Lemma 26.6 in [SSBD] shows that for any $c \in \mathbb{R}$, we have $R(cA) \leq |c|R(A)$. Should this be an equality? Prove your answer.*
2. *For sets $A, B \subseteq \mathbb{R}^m$ define their sum by $A + B := \{a + b | a \in A, b \in B\}$. Show that $R(A + B) = R(A) + R(B)$.*
3. *For a set $A \subseteq \mathbb{R}^m$, define $A - A := \{a - b | a, b \in A\}$. Show that $R(A - A)/2 = R(A)$.*
4. *If $A \subset B$, then $R(A) \leq R(B)$.*

Problem 2. *For this problem we work in the realizable setting of PAC learning (we are given $X, Y, \mathcal{H}$). Fix error ϵ, and confidence parameter $\delta_0 \in (0, 1)$ and fix a positive integer m. Suppose we have a PAC-learning algorithm A such that for any realizable distribution $\mathcal{D}$ (on $Z = X \times Y$), on input $S \sim \mathcal{D}^m$, with probability at least $1 - \delta_0$ algorithm A outputs a hypothesis h such that $L_{\mathcal{D}}(h) \leq \epsilon$. Suppose that the confidence parameter δ_0 is large and so the probability $1 - \delta$ is small, and we would like to design an algorithm that has smaller confidence parameter δ. Using A as a subroutine, design another learning algorithm that can achieve any desired confidence parameter $\delta > 0$ at the price of slightly larger error $\epsilon + \epsilon'$ and using more samples. Here $\epsilon' > 0$ is any given constant. Show that this can be done with only a logarithmic price in $1/\delta$ by proving that* $\mathrm{poly}(1/\delta_0, \log(1/\delta), 1/\epsilon')$ *samples suffice.*