Instructions

- For your proofs, you may use any result covered in class, and its analysis, but please cite the result that you use.
- Each problem is worth 10 points.
- The assignment will be graded on clarity and correctness. If your arguments are not clear and have holes in them, I will assume that they are wrong.

Problem 1: PAC Learning

You are given an algorithm $A$ for learning a hypothesis class $H$. Given a parameter $0 < \epsilon < 1$ and $m(\epsilon)$ i.i.d samples from a separable data distribution $D$, $A$ produces a $h \in H$ such that $err_D(h) \leq \epsilon$ with probability at least $\frac{1}{2}$ over its input.

Given $A$ and a procedure $EX$ which when called outputs a fresh independent labeled example from a separable distribution $D$, how will you design a PAC-learning algorithm for $H$? Describe your procedure, and prove that it PAC-learns $H$. [Hint: You may need to run $A$ multiple times.]

Problem 2: VC Dimension

In learning equivalence relations over an universe $U$ with $n$ elements, the sample space $X = U \times U$. A hypothesis $h_{\Pi}$ corresponds to a disjoint partition $\Pi = \{U_1, \ldots, U_k\}$ of $U$ into $k$ parts; for $u$ and $v$ in $U$, $h_{\Pi}((u, v)) = 1$ if and only if $u$ and $v$ belong to some $U_j$ in $\Pi$. The hypothesis class $H$ is the set of all such hypotheses.

Show that the VC dimension of $H$ is $\Theta(n)$ – that is, there exist constants $c_1$ and $c_2$ such that $c_1 n \leq VC(H) \leq c_2 n$.

Problem 3: VC Dimension

Let $X = \mathbb{R}^2$, and let $H$ be the hypothesis class of all convex sets on the plane. In other words, each $h_K \in H$ corresponds to a convex set $K$, such that $h_K(x) = 1$ if $x$ is inside or on the boundary of $K$ and 0 otherwise. Show that $H$ has infinite VC dimension.