10-806 Foundations of Machine Learning and Data Science

Homework # 5 **Due: December 9, 2015**

Groundrules:

- Your work will be graded on correctness, clarity, and conciseness. You should only submit work that you believe to be correct; if you cannot solve a problem completely, you will get significantly more partial credit if you clearly identify the gap(s) in your solution. It is good practice to start any long solution with an informal (but accurate) proof summary that describes the main idea.
- You may collaborate with others on this problem set and consult external sources. However, you must *write your own solutions* and *list your collaborators/sources* for each problem.

Problems:

1. [40 pts] **Distance and Disagreement.**
   (a) [20 pts] Prove that for any distribution D, the distance function $d(h, h') = \Pr_D(h(x) \neq h'(x))$ satisfies triangle inequality.
   Note that d is clearly symmetric, so this means it is a (semi)metric.
   (b) [20 pts] Let C be the class of linear separators through the origin, and let D be the uniform distribution over $X = \{x \in \mathbb{R}^d : ||x|| \leq 1\}$, i.e., over points in the unit ball in $\mathbb{R}^d$. Let h be the linear separator $x_1 \geq 0$. What is $DIS(B(h, \frac{\epsilon}{\sqrt{d}}))$? In other words, describe what the region looks like or what points are in it.
   (c) [10 pts extra credit] What approximately is $\Pr(DIS(B(h, \frac{\epsilon}{\sqrt{d}})))$?
2. [30 pts] **Active Learning Lower Bound.** Prove that active learning of intervals on the line over the uniform distribution on $[0, 1]$, needs $\Omega(1/\epsilon)$ label requests. Specifically, give a distribution over target functions such that for any active learning algorithm A that makes less than c/ϵ label requests (for your choice of constant $c > 0$), if the target function is chosen at random from the distribution over targets, then with probability at least $1/2$, A outputs a hypothesis of error $\geq \epsilon$. You may assume $\epsilon \leq c'$ for sufficiently small constant c'.
3. [30 pts] **Another Active Learning Lower Bound.** Assume that we are learning a linear separator $w^* \cdot x \geq 0$ and that the distribution is uniform over the surface of the unit sphere in $\mathbb{R}^d$. Show that $\Omega(d \log(1/\epsilon))$ is a lower bound on the number of label requests needed by any active learning algorithm to achieve error at most ϵ with probability $\geq 1/2$. A fact you may wish to use (you don't need to prove this) is that for some constant $c > 0$, for any $\epsilon' > 0$, it is possible to construct a set of $(\frac{c}{\epsilon'})^d$ vectors in $\mathbb{R}^d$ such that the angle between any two of the vectors is at least ϵ'.
4. [20 pts Extra Credit] **Class Conditional Queries.** A *class conditional query* is a more powerful form of active learning, where the algorithm specifies a *subset* S of the unlabeled examples and a label ℓ and asks "is there any example of label ℓ in S? If so, give me one."

So, if S has size 1, then this is just like standard active learning (in the case of binary labels) but this can be more powerful because you can use larger sets.

Give an example of a class $\mathcal{C}$ of functions that (in the realizable case) can be learned to error $\leq \epsilon$ with probability $\geq 1-\delta$ from $O(\frac{1}{\epsilon}\log\frac{1}{\delta})$ unlabeled examples and *just one class conditional query*, and yet, for some distribution on examples, would require $\Omega(1/\epsilon)$ label requests to learn for standard active learning.

5. [20 pts Extra Credit] **Class Conditional Queries, contd.** Show that any class $\mathcal{C}$ of VC-dimension d can be learned to error $\leq \epsilon$ with probability $\geq 1-\delta$ from $m = O(\frac{1}{\epsilon}(d\log\frac{1}{\epsilon}+\log\frac{1}{\delta}))$ unlabeled examples and $O(d\log(m/d))$ CCQs.