8803 Machine Learning Theory

Homework # 3 **Due: March 9th 2010**

This homework is due by the start of class on March 9th. You can either submit the homework via the course page on T-Square or hand it in at the beginning of the class on March 9th.

Groundrules:

- Your work will be graded on correctness, clarity, and conciseness.
- 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. **A bad modification to Winnow.** Suppose that we modify Winnow so that it doubles its weights on positive examples even when it did *not* make a mistake. Show how this can cause the algorithm to make an unbounded number of mistakes, even if all examples *are* consistent with some disjunction.

2. **Balanced Winnow.** Here is a variation on the Winnow algorithm, called *Balanced Winnow.* First of all, we introduce a fake variable x_0 which is set to 1 in every example. For each variable x_i ($0 \leq i \leq n$), and each output value y (as usual, $y \in \{-, +\}$, but you can also use this algorithm for multi-valued outputs) we have a weight w_{iy}. All weights are initialized to 1. In addition, we are given parameters $\alpha > 1$ and $\beta < 1$. The algorithm proceeds as follows:

   (a) Given example x, predict the label y such that $\sum_i x_i w_{iy}$ is largest.

   (b) If the algorithm makes a mistake, predicting y' when then correct answer is y, then for each $x_i = 1$, multiply the weight w_{iy} by α, and multiply $w_{iy'}$ by β.

   Using $\alpha = 3/2$ and $\beta = 1/2$, prove that as with the standard Winnow algorithm, this algorithm makes at most $O(r \log n)$ mistakes on any disjunction (OR-function) of r variables.

3. **Sample complexity bounds.** For some learning algorithms, the hypothesis produced can be uniquely described by a small subset of k of the training examples. E.g., if you are learning an interval on the line using the simple algorithm "take the smallest interval that encloses all the positive examples," then the hypothesis can be reconstructed from just the outermost positive examples, so $k = 2$. For a conservative Mistake-Bound learning algorithm, you can reconstruct the hypothesis by just looking at the examples on which a mistake was made, so $k \leq M$, where M is the algorithm's mistake-bound. (In this case, you may also care about the *order* in which those examples arrived.)

   Prove a PAC guarantee based on k. Specifically, fixing a description language (reconstruction procedure), so for a given set S' of examples we have a well-defined hypothesis $h_{S'}$, show that

$$\Pr_{S \sim D^n} \left(\exists S' \subseteq S, |S'| = k, \text{ such that } h_{S'} \text{ has 0 error on } S - S' \text{ but true error } > \epsilon \right) \leq \delta,$$

so long as

$$n \geq \frac{1}{\epsilon}\left(k \ln n + \epsilon k + \ln \frac{1}{\delta}\right).$$

Hint: Think of S' as a subset of indices, and imagine drawing points in S by drawing those in S' first. Also, be clear in your analysis what the events are that you are taking the union bound over.

Note the similarity of the form of this bound to VC-dimension and other bounds we have seen. These are often called “compression bounds”.