15-854 Machine Learning Theory

Homework # 1 **Due: January 21, 1998**

Groundrules:

- Homeworks will generally consist of *exercises*, easier problems designed to give you practice, and *problems*, that may be harder, and/or somewhat open-ended. You should do the exercises by yourself, but you may work with a friend on the harder problems if you want. One exception: no fair working with someone who has already figured out (or already knows) the answer. If you work with a friend, then write down who you are working with.
- If you've seen a problem before (sometimes I'll give problems that are "famous"), then say that in your solution. It won't affect your score, I just want to know. Also, if you use any sources other than the textbook, write that down too. It's fine to look up a complicated sum or inequality or whatever, but don't look up an entire solution.

Exercises:

1. **The difficulty of 2-term DNF.** Prove that the consistency problem for 2-term DNF formulas (e.g., formulas like $x_1\bar{x}_2x_3 \vee x_2x_3\bar{x}_4$) is NP-hard. (This was shown by Pitt and Valiant [1].) Hint: reduce from the NP-complete Set-Splitting problem, in analogy to the neural nets reduction discussed in class.
2. **Mistake-bound model.** Give an algorithm to learn 2-CNF formulas over n boolean features in the mistake-bound model. Your algorithm should run in polynomial-time per example (so the "halving algorithm" is not allowed). How many mistakes does it make at most?

Problems:

3. **Mistake-bound model.** Give an algorithm that learns the class of 1-decision lists in the mistake-bound model, with mistake bound $O(nL)$ where n is the number of variables and L is the length of the shortest decision list consistent with the data. The algorithm should run in polynomial time per example.

   Hint: the hypotheses used by your algorithm need not be decision lists exactly.

References

[1] Leonard Pitt and Leslie G. Valiant. Computational limitations on learning from examples. *Journal of the ACM*, 35(4):965–984, 1988.