

## 8803 Machine Learning Theory

Homework # 5

Due: April 22nd 2010

---

This homework is due by the start of class on April 22nd. You can either submit the homework via the course page on T-Square or hand it in at the beginning of the class on April 22nd. Start early!

### 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. Learning Decision Lists in the SQ model

Give an algorithm to learn the class of *decision lists* in the SQ model (and argue correctness for your algorithm). Be clear about what specifically the queries  $\chi$  are and the tolerances  $\tau$ . Ensure that your algorithm does not ask for conditional probabilities like  $\Pr[A|B]$ , but asks for  $\Pr[A \wedge B]$ .

#### 2. Consistency Problem for 2-term DNF formulas

Prove that the consistency problem for 2-term DNF formulas is NP-hard.

Hint 1: reduce from the NP-complete Set-Splitting problem (also called the hypergraph 2-colorability problem), which is the following. You are given  $m$  subsets  $S_1, S_2, \dots, S_m$  of the set  $\{1, \dots, n\}$ . You want to know if there is a coloring of the numbers  $1, \dots, n$  in which each number is colored either red or blue, so that none of the sets  $S_i$  has all its elements the same color. For instance, if the sets are  $\{1, 2, 3\}$ ,  $\{2, 3\}$ , and  $\{1, 2\}$ , a legal coloring might give the numbers 1, 2, and 3 the colors red, blue, and red respectively.

Hint 2: Think about what it means if examples 011111 and 101111 are positive but 001111 is negative.