Machine Learning and Artificial Intelligence - COS 402

Written Homework Assignment 2

Due Date: one week from announcement in class, due in class

(1) Consulting other students from this course is allowed. In this case - clearly state whom you consulted with for each problem separately.

(2) Searching the internet or literature for solutions is NOT allowed.

(3) Submit your homework in separate pages for the different questions, each including your name and email address (this is to help the graders). Typing solutions up is strongly advised.

I State and re-prove the fundamental theorem for statistical learning for finite hypothesis classes which we learned and proved in class. Explain and justify each transition in the proof from elementary facts in probability theory and combinatorics.

II Consider the following dataset:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y
1	1	0	0	0	1	0	1	1
1	0	0	0	0	0	0	0	0
0	0	0	1	1	1	1	0	0
0	1	0	0	1	0	0	0	0
1	0	0	0	0	1	0	1	0
0	1	1	0	1	1	0	1	1
1	1	0	1	0	1	0	1	1
0	1	0	1	0	1	0	0	1
0	0	0	0	0	0	1	1	0

In this formulation, there are eight attributes (or features or dimensions), $x_1, \ldots, x_8$, each taking the values 0 or 1. The label (or class) is given in the last column denoted y; it also takes the values 0 or 1. In class, it was noticed that the label y is 1 if and only if x_2 and x_6 are both equal to 1. Since attributes and labels are $\{0, 1\}$-valued, we can write this rule succinctly as $y = x_2 x_6$. In general, such a product of any number of attributes is called a *monomial*. (This includes the "empty" monomial, which, being a product of no variables, is always equal to 1.)

Throughout this problem, you can assume that the attributes and labels are all $\{0, 1\}$-valued. Also, let n be the number of attributes (for instance, $n = 8$ in the example above). Assume, as usual, that the training and test examples are generated independently at random according to the same distribution.

(a) Describe a simple algorithm that, given a dataset, will efficiently (in time which is polynomial in n and m – the number of examples) find a monomial consistent with it, assuming that one exists.

(b) What is the total number of monomials that can be defined on n attributes?

(c) Suppose you applied your algorithm to the dataset above, and that a consistent monomial was found. Use the bound derived in class to compute an upper bound on the generalization error of this monomial. Derive a bound that holds with 95% confidence (so that $\delta = 0.05$).

(d) Continuing the last question in which your algorithm is applied to data with $n = 8$ attributes, how many training examples would be needed to be sure the generalization error of a consistent monomial is at most 10% with 95% confidence?

III For this exercise we restrict ourselves to one dimensional functions, $d = 1$. Prove the equivalence of the two definitions of convexity shown in class. That is, we defined that $f : K \mapsto \mathbb{R}^d$ is convex if and only if $f(\frac{1}{2}x + \frac{1}{2}y) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y)$ for all $x, y \in K$. Show that f (assuming it is differentiable) is convex if and only if

$$f(x) \geq f(y) + f'(y)(x - y)$$