# Introduction to Machine Learning - COS 324

## Written Homework Assignment 3

*Due date: The minute after 11:59pm of 10/10/2017. Electronic submissions only.*

**Note:**

- Consulting with other students from this course is allowed. If you do so, clearly state whom you consulted with for each problem separately.
- Searching the internet or literature for solutions is **prohibited**.

1. State and re-prove the fundamental theorem for PAC learnability 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. You are welcome to use the handouts.

2. 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 $y$ and also takes the values 0 or 1. Notice 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 the question, you may assume that the attributes and labels are all $\{0, 1\}$-valued. Also, let $n$ be the number of attributes. Let $m$ be the number of examples. For instance, $n = 8$ and $m = 9$ in the table above. Assume, as usual, that training and test examples are generated independently at random according to some unknown distribution.

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

(b) Describe a simple algorithm that, given a dataset, efficiently (in time which is polynomial in $n$ and $m$) finds a consistent monomial, assuming that one exists.

(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 $\epsilon$ 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 make sure that the generalization error of a consistent monomial is at most 10% with 95% confidence?