Prof. Friedrich Eisenbrand
Nicolai Hähnle

Discussions from: October 8, 2009

Mathematics of Machine Learning

Fall 2009

Assignment Sheet 2

Exercises marked with a $\star$ can be handed in for bonus points. Due date is October 22.

Exercise 1 ($\star$ = exercise 7 from previous assignment sheet)
Consider the following *two-oracle* variant of the PAC model: when $c \in \mathscr{C}$ is the target concept, there are separate and arbitrary distributions $\mathscr{D}_c^+$ over only the positive examples of c and $\mathscr{D}_c^-$ over only the negative examples of c. The learning algorithm must find a hypothesis h satisfying $\Pr_{x \in \mathscr{D}_c^+}[h(x) = 0] \le \varepsilon$ and $\Pr_{x \in \mathscr{D}_c^-}[h(x) = 1] \le \varepsilon$. In other words, the learning algorithm may now explicitly request either a positive or a negative example, but must find a hypothesis with small error on both distributions.

Let $\mathscr{C}$ be any concept class and $\mathscr{H}$ be any hypothesis class. Let h_0 and h_1 be representations of the identically 0 and identically 1 functions, respectively. Prove that $\mathscr{C}$ is efficiently PAC learnable using $\mathscr{H} \cup \{h_0, h_1\}$ in the original one-oracle model if and only if $\mathscr{C}$ is efficiently PAC learnable using $\mathscr{H}$ in the two-oracle model.

Exercise 2
Let $\mathscr{C}$ and $\mathscr{H}$ be a concept class and a hypothesis class, respectively, over the instance space X with $X_n = \{0,1\}^n$, and let $\mathscr{C}$ be efficiently PAC learnable using $\mathscr{H}$.

1. Show that for every $c \in \mathscr{C}_n$ there exists a hypothesis $h \in \mathscr{H}_n$ with $h(x) = c(x)$ for all $x \in X_n$.
2. What if $X_n = \mathbb{R}^n$?

Exercise 3
Let k and l be natural numbers, $l \ge k$, and let us make the following hardness assumption:

> There does not exist a randomized algorithm that takes as input a parameter $0 < \delta \le 1$ and a graph G that is either k-colorable or not l-colorable, that runs in time polynomial in $1/\delta$ and the size of G, and that correctly decides with probability at least $1-\delta$ whether G is k-colorable or not l-colorable.

Generalize the argument from the lecture to show that, under this assumption, k-term DNF formulae are not efficiently PAC learnable using l-term DNF formulae.

Exercise 4 ($\star$)
The mint of Cheatistan uses biased coin blanks but does not always stamp the coins with

the same orientation. As a result, there are two types of coins: for one type of coin, the probability of turning up heads is $p = 1/3$, for the other type of coin this probability is $p = 2/3$. Given a coin, we want to determine its type in the following way: we toss the coin m times. If it turns up heads at least $m/2$ times, we say that it is of type $p = 2/3$, otherwise we say that it is of type $p = 1/3$.

Using Chernoff bounds (see exercise 6), show that we need at most $m = O(\ln(1/\delta))$ many coin tosses in order to have probability less than δ of determining the type of coin incorrectly.

Exercise 5 ($\star$)
Consider the following variations of the efficient PAC learning definition, where learning algorithms have

1. constant success probability $p > 0$.
2. parameterized success probability at least $1-\delta$, and their running time depends polynomially on $1/\delta$ (this is the original definition).
3. parameterized success probability at least $1-\delta$, and their running time depends polynomially on $\ln(1/\delta)$.

Show that all these definitions are equivalent, that is: a concept class is efficiently PAC learnable according to one of those definitions if and only if it is efficiently learnable in all of them.

Hint: Show "3 $\implies$ 2" and "2 $\implies$ 1" first. The hard part is showing "1 $\implies$ 3": your "definition 3 algorithm" B should run a "definition 1 algorithm" A several times, each time testing whether A returned a good hypothesis. Use Chernoff bounds to analyze your hypothesis test similar to exercise 4.

Exercise 6
Let $X_1, \ldots, X_n$ be independent random variables with $-1 \leq X_i \leq 1$ and $\mathbb{E}[X_i] = 0$ for every i. Let $X = \sum_{i=1}^n X_i$. In this exercise, you will prove two variants of Chernoff bounds; each of the steps is self-contained.

1. Show that for every $\lambda > 0$ and every a, $\Pr[X \geq a] = \Pr[e^{\lambda X} \geq e^{\lambda a}] \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda a}}$.
2. Show that for every i and every $\lambda > 0$, $\mathbb{E}[e^{\lambda X_i}] \leq \frac{e^{\lambda}+e^{-\lambda}}{2} \leq e^{\lambda^2/2}$.
3. Show that $\mathbb{E}[e^{\lambda X}] \leq e^{\lambda^2 n/2}$ for every $\lambda > 0$.
4. Conclude that $\Pr[X \geq a\sqrt{n}] \leq e^{-a^2/2}$.
5. Let $Y_1, \ldots, Y_n$ be independent random variables with $0 \leq Y_i \leq 1$ for all i and let $Y = \sum_{i=1}^n Y_i$. Show that $\Pr[Y - \mathbb{E}[Y] \geq a\sqrt{n}] \leq e^{-a^2/2}$.