Mehryar Mohri
Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 1
Due: February 22, 2010

A. PAC Learning

- Let $X = \mathbb{R}^2$ with orthonormal basis (e_1, e_2) and consider the set of concepts defined by the area inside a right triangle ABC with two sides parallel to the axes, with $\overrightarrow{AB}/AB = e_1$ and $\overrightarrow{AC}/AC = e_2$, and $AB/AC = \alpha$ for some positive real $\alpha \in \mathbb{R}_+$. Show, using similar methods to those used in the lecture slides for the axis-aligned rectangles, that this class can be (ϵ, δ)-PAC-learned from training data of size $m \geq (3/\epsilon)\log(3/\delta)$.

B. Biased coins

- Professor Moent has two coins in his pocket, coin x_A and coin x_B, both slightly biased as follows:

  $$\Pr[x_A = 0] = 1/2 - \epsilon/2 \quad \Pr[x_B = 0] = 1/2 + \epsilon/2,$$

  where $0 < \epsilon < 1$ is a small positive number and where 0 denotes head and 1 tail. He likes to play the following game with his students. He randomly picks a coin $x \in \{x_A, x_B\}$ from his pocket, tosses it m times, then reveals the sequence of 0s and 1s he obtained and asks which coin was tossed.

  The goal of this exercise is to determine how large m needs to be for a student's coin prediction error to be at most $\delta > 0$.

  1. Let S be a sample of size m. Professor Moent's best student, Oskar, plays according to the decision rule $f_o \colon \{0,1\}^m \to \{x_A, x_B\}$ defined by $f_o(S) = x_A$ iff $N(S) < m/2$, where $N(S)$ is the number of 0's in sample S.

     Suppose m is even, then show that

     $$error(f_o) \geq \frac{1}{2} \Pr\left[N(S) \geq \frac{m}{2} \middle| x = x_A\right]. \tag{1}$$

  2. Assuming m even, use the inequalities given in the appendix to show that

     $$error(f_o) > \frac{1}{4}\left[1 - \left[1 - e^{-\frac{m\epsilon^2}{1-\epsilon^2}}\right]^{\frac{1}{2}}\right]. \tag{2}$$

3. Argue that if m is odd, the probability can be lower bounded by the one for $m+1$ and conclude that for both odd and even m,

$$error(f_o) > \frac{1}{4}\left[1 - \left[1 - e^{-\frac{2\lceil m/2\rceil \epsilon^2}{1-\epsilon^2}}\right]^{\frac{1}{2}}\right]. \tag{3}$$

4. Using this bound, how large should m be for Oskar's error to be at most δ, where $0 < \delta < 1/4$. What is the asymptotic behavior of this lower bound as a function of ϵ?
5. Show that no decision rule $f\colon \{0,1\}^m \to \{x_a, x_B\}$ can do better than Oskar's rule f_o. Conclude that the lower bound of the previous question applies to all rules.

Appendix

0.1 Binomial inequality

Let B be a binomial (m,p) random variable with with $p \leq 1/2$. Then, for $mp \leq k \leq m(1-p)$,

$$\Pr[B \geq k] \geq \Pr\left[N \geq \frac{k - mp}{\sqrt{mp(1-p)}}\right], \tag{4}$$

where N is in standard normal form.

0.2 Tail bound

If N is a random variable following the standard normal distribution, then for $u > 0$,

$$\Pr[N \geq u] \geq \frac{1}{2}\left(1 - \sqrt{1 - e^{-u^2}}\right). \tag{5}$$