Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 1 – Solution
February 6, 2007

A. Senate Laws

[30 points]

1. [15 points] The true error in the consistent case is bounded as follows:

$$error_D(h) \leq \frac{1}{m}(\log |H| + \log \frac{1}{\delta}). \quad (1)$$

   For $\delta = .05$, $m = 200$ and $H = 2800$, $error_D(h) \leq 5.5\%$.

2. [15 points] The true error in the inconsistent case is bounded as:

$$error_D(h) \leq \widehat{error_D}(h) + \sqrt{\frac{1}{2m}(\log 2|H| + \log \frac{1}{\delta})}. \quad (2)$$

   For $\delta = .05$, $\widehat{error_D}(h) = m'/m = .1$, $m = 200$ and $H = 2800$, $error_D(h) \leq 27.05\%$.

B. PAC Learning of Hyper-rectangles

[30 points]

1. The proof in the case of hyper-rectangles is similar to the one given in class. The algorithm selects the tightest axis-aligned hyper-rectangle containing all the sample points. For $i \in [1, 2n]$, select a region r_i such that $\Pr_D[r_i] = \epsilon/(2n)$ for each edge of the hyper-rectangle R. Assuming that $\Pr_D[R - R'] > \epsilon$, argue that R' cannot meet all r_is, so it must miss at least one. The probability that none of the m sample points falls into region r_i is $(1 - \epsilon/2n)^m$. By the union bound, this shows that

$$\Pr[error(R') > \epsilon] \leq 2n(1 - \epsilon/2n)^m \leq 2ne^{-\frac{\epsilon m}{2n}}. \quad (3)$$

   Setting δ to the right-hand side shows that for

$$m \geq \frac{2n}{\epsilon} \log \frac{2n}{\delta}, \quad (4)$$

   with probability at least $1 - \delta$, $error_D(R') \leq \epsilon$.

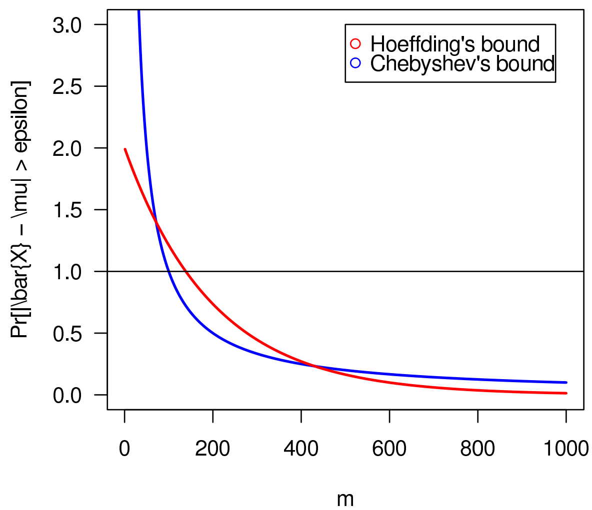

Figure 1: Comparison of Chebyshev's and Hoeffing's bound.

C. Bound Comparison

[40 points]

Let $X_1, \ldots, X_m$ be a sequence of random variables taking values in $[0, 1]$ with the same mean μ and variance $\sigma^2 < \infty$ and let $\overline{X} = \frac{1}{m}\sum_{i=1}^{m} X_i$.

1. [20 points] Since the random variables X_i are independent, the variance of $\overline{X}$ is the sum of the variances: $\text{Var}[\overline{X}] = m(\sigma^2/m^2) = \sigma^2/m$. For any $\epsilon > 0$, using Chebyshev's inequality (see lecture 1),

$$\Pr[|\overline{X} - \mu| > \epsilon] \leq \frac{\sigma^2}{m\epsilon^2}. \tag{5}$$

Using Hoeffding's inequality,

$$\Pr[|\overline{X} - \mu| > \epsilon] \leq 2e^{-2m\epsilon^2}. \tag{6}$$

Thus, Chebyshev's inequality is tighter for $\frac{\sigma^2}{m\epsilon^2} < 2e^{-2m\epsilon^2}$, that is for $\sigma \leq \sqrt{(2m\epsilon^2)\ e^{-2m\epsilon^2}}$.

2. [20 points] When X_i takes values in $\{0, 1\}$, the variance of X_i is given by

$$\sigma^2 = \mathrm{E}[X_i^2] - \mathrm{E}[X_i]^2 = \mathrm{E}[X_i] - \mathrm{E}[X_i]^2, \tag{7}$$

since $X_i^2 = X_i$. For $\mu \in [0, 1]$, the function $\mu \mapsto \mu(1 - \mu)$ reaches its maximum for $\mu = \frac{1}{2}$. Thus, $\sigma^2 \leq \frac{1}{4}$. Chebyshev's inequality can then be simplified into:

$$\Pr[|\overline{X} - \mu| > \epsilon] \leq \frac{1}{4m\epsilon^2}, \tag{8}$$

The two bounds are approximately equal for $m\epsilon^2 \approx 1.075$.

Figure 1 plots these inequalities for $\epsilon = .05$. Both bounds are vacuous for values of m less than 100. Chebyshev's inequality is tighter for $m < m_0 \approx 1.075/\epsilon^2 = 430$, Hoeffding's inequality tighter for larger values of m.