STAT 881
Spring 2008

HOMEWORK 2

1. For a fixed classification rule f and (X_i, Y_i) iid with $P_{X,Y}$, $i = 1, \ldots, n$, $nR_n(f) = \sum_{i=1}^n I(f(X_i) \neq Y_i)$ follows a binomial distribution $B(n, R(f))$. By the central limit theorem, we have

$$\frac{\sqrt{n}(R_n(f) - R(f))}{\sqrt{R(f)(1 - R(f))}} \to N(0, 1)$$

in distribution as $n \to \infty$. Find an approximate bound of $P(|R_n(f) - R(f)| \geq \epsilon)$ for n large enough that $n \geq R(f)(1 - R(f))/\epsilon^2$ by using the asymptotic distribution of the emprical error rate of f, and compare it with the bound given by Hoeffding's inequality.

Hint: You may use

$$1 - \Phi(x) \leq \frac{\phi(x)}{x} \text{ for } x > 0$$

and in particular, $1 - \Phi(x) \leq \phi(x)$ *for* $x \geq 1$ *without proof.*

2. Let $\mathcal{A}$ be a collection of measurable sets in $\mathbb{R}^d$ and $Z_1, \ldots, Z_n$ be iid with probability measure $\nu(A) = P(Z_1 \in A)$ for all measurable sets $A \subset \mathbb{R}^d$. Define the empirical measure $\nu_n(A) = \frac{1}{n} \sum_{i=1}^n I(Z_i \in A)$ and the nth shatter coefficient of $\mathcal{A}$, $s(\mathcal{A}, n) = \max_{(z_1,\ldots,z_n)\in(\mathbb{R}^d)^n} N_{\mathcal{A}}(z_1, \ldots, z_n)$, where $N_{\mathcal{A}}(z_1, \ldots, z_n) =$ the cardinality of $\{(I(z_1 \in A), \ldots, I(z_n \in A)) \mid A \in \mathcal{A}\}$.

   (a) Prove the Vapnik-Chervonenkis inequality, a generalization of the Glivenko-Cantelli theorem:

   For any probability measure ν *and class of sets* $\mathcal{A}$, *and for any* n *and* $\epsilon > 0$,

$$P(\sup_{A\in\mathcal{A}} |\nu_n(A) - \nu(A)| > \epsilon) \leq 8s(\mathcal{A}, n)e^{-n\epsilon^2/32}.$$

   (b) Verify that the inequality can be strengthened to

$$P(\sup_{A\in\mathcal{A}} |\nu_n(A) - \nu(A)| > \epsilon) \leq 8E\{N_{\mathcal{A}}(Z_1, \ldots, Z_n)\}e^{-n\epsilon^2/32}.$$

   Note that this upper bound depends on the distribution ν.

3. Let $\mathcal{A}$, $\mathcal{A}_1$, and $\mathcal{A}_2$ be classes of subsets of $\mathbb{R}^d$. Prove the following properties of the shatter coefficient.

   (a) $s(\mathcal{A}, n+m) \leq s(\mathcal{A}, n)s(\mathcal{A}, m)$.

   (b) For $\mathcal{A}_c = \{A^c : A \in \mathcal{A}\}$, $s(\mathcal{A}_c, n) = s(\mathcal{A}, n)$.

   (c) For $\mathcal{A} = \{A_1 \cap A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$, $s(\mathcal{A}, n) \leq s(\mathcal{A}_1, n)s(\mathcal{A}_2, n)$.

4. Consider a class of decision rules with only one parameter defined on $\mathcal{X} = (0, 2\pi)$, $\mathcal{F} = \{f(x) = I(\sin(\alpha x) \geq 0) \mid \alpha \in (0, \infty)\}$. Show that the Vapnik-Chervonenkis dimension of $\mathcal{F}$ is infinity. This illustrates that the V-C dimension of a class is not necessarily the same as the number of parameters of the class.

   Hint: Let $x_i = 10^{-i}$ for $i = 1, \ldots, n$. Given a subset of $\{x_1, \ldots, x_n\}$, let $y_i \in \{-1, 1\}$ indicate exclusion (-1) or inclusion (1) of x_i in the subset. Then by taking $\alpha = \pi\{1 + (1/2)\sum_{i=1}^{n}(1 - y_i)10^i\}$, show that $sign(\sin(\alpha x_i)) = y_i$ for all i.