Stat 210B Homework Assignment 3 (due March 3)

1. Show that for all $1 \le p < \infty$, we have:

$$H_p(\epsilon, Q, \mathcal{F}) \le H_{p,B}(\epsilon, Q, \mathcal{F})$$

for all ϵ. Show that if Q is a probability measure, we have:

$$H_{p,B}(\epsilon, Q, \mathcal{F}) \le H_\infty(\epsilon/2, \mathcal{F}).$$

2. Show that the Cramér-von Mises statistic, $n\int(F_n - F)^2 dF$, is a continuous function of the empirical process (with respect to the supremum norm). Show that the distribution of this statistic is independent of F, for all continuous F.

3. Let $\mathcal{H} \subset L^2(\mathbb{P})$ be a function class, where $\mathbb{P}$ is some measure defined on the real line. Now consider the following class of additive models:

$$\mathcal{F}_{d,\mathcal{H}} = \{f : \mathbb{R}^d \to \mathbb{R} \mid f(X_1, X_2, ..., X_d) = \sum_{j=1}^{d} h_j(X_j) \text{ where } h_j \in \mathcal{H} \text{ for each } j\},$$

where X_j, $j = 1, 2, ..., d$ are i.i.d and distributed according to the measure $\mathbb{P}$ ($X = (X_1, X_2, ..., X_d) \sim \mathbb{P}^d$). Prove the following upper bound for the covering entropy on $\mathcal{F}_{d,\mathcal{H}}$ with respect to the norm induced by $L^2(\mathbb{P}^d)$:

$$\log N_2(\epsilon;\ \mathbb{P}^d;\ \mathcal{F}_{d,\mathcal{H}}) \le d \log N_2(\frac{\epsilon}{\sqrt{d}};\ \mathbb{P};\ \mathcal{H}).$$

Would the bound hold if the X_j's were identically distributed but not necessarily independent?

4. Derive the VC dimension for the following class of sets in $\mathbb{R}^n$:

$$\{x \in S_{a_1,a_2,...,a_n,b} \mid \sum_{i=1}^{n} a_i x_i \le b;\ \text{where } a_1, a_2, ..., a_n, b \in \mathbb{R}\}.$$

5. Consider the function class $\mathcal{F} = \{f_\alpha : \mathbb{R} \to \mathbb{R} \mid f(x) = \alpha x \text{ where } |\alpha| \le R\}$. Let $\mathbb{P}$ denote the standard Gaussian measure (i.e. $X \sim \mathbb{P}$ means that $X \sim \mathcal{N}(0,1)$).

   (a) Determine $\log N_1(\epsilon;\ \mathbb{P};\ \mathcal{F})$ exactly.

   (b) Prove that

$$\mathbb{P}\Bigg(\log N_1(\epsilon;\ \mathbb{P}_n;\ \mathcal{F}) \ge \log N_1(\frac{\epsilon}{2};\ \mathbb{P};\ \mathcal{F})\Bigg) \to 0.$$

   Hint: You may find the following bound useful:
   For a standard χ^2 random variable with d degrees of freedom Z_d, we have the following tail bound:

$$\mathbb{Q}[Z_d \ge d(1+t)] \le \exp\Bigg(-\frac{3}{16} d t^2\Bigg).$$

   (c) Now use Theorem 24 from Pollard (page 25) to prove a uniform law of large numbers for $\mathcal{F}$.