10-702 Statistical Machine Learning: Assignment 6
Due Friday, April 15

Hand in to Sharon Cavlovich, GHC (Gates Hillman Center) 8215 by 3:00.

1. We introduced covering numbers in class. A related idea is bracketing. Let $F$ be a set of functions. If $\ell$ and $u$ are two functions, we define the bracket

$$[\ell, u] = \left\{ f : \ell(x) \leq f(x) \leq u(x) \text{ for all } x \right\}.$$ 

We say that $[\ell, u]$ is an $\epsilon$ bracket in $L_r(P)$ if

$$\int |u(x) - \ell(x)|^r dP(x) \leq \epsilon^r.$$ 

The bracketing number $N_1(\epsilon, F, L_1(P))$ is the smallest number of $\epsilon$ brackets needed to cover $F$.

(a) Suppose that $N_1(\epsilon, F, L_1(P)) < \infty$ for every $\epsilon > 0$. Also suppose that $\|f\|_\infty \leq B < \infty$ for every $f \in F$. Show that

$$\sup_{f \in F} |P_n(f) - P(f)| \overset{P}{\to} 0.$$ 

(b) Let $F$ be the set of all indicator functions of the form $f_t = I_{(-\infty, t]}$ for $t \in \mathbb{R}$. Show that

$$N_1(\epsilon, F, L_1(P)) \leq \frac{C}{\epsilon}$$

for some $C > 0$.

(c) Show that

$$\sup_t |\widehat{F}_n(t) - F(t)| \overset{P}{\to} 0$$

where $F(t) = \mathbb{P}(X \leq t)$ is the cdf and $\widehat{F}_n(t) = n^{-1} \sum_{i=1}^n I(X_i \leq t)$ is the empirical cdf.

2. Let $X \geq 0$ be a positive random variable, with moment generating function that exists in an interval around zero. Given some $\delta > 0$, show that

$$\inf_{k=0,1,2,...} \frac{\mathbb{E}[|X|^k]}{\delta^k} \leq \inf_{\lambda \geq 0} \frac{\mathbb{E}[\exp(\lambda X)]}{\exp(\lambda \delta)}.$$ 

What does the above result suggest for the Chernoff bounding technique?
3. (a) Let \( Y \sim N(\theta, 1) \) where \( \theta \in \{-a, a\} \). Define a loss function \( L(\theta, \hat{\theta}) = 1 \) if \(|\theta - \hat{\theta}| \geq a\) and \( L(\theta, \hat{\theta}) = 0 \) otherwise. Find an exact expression for the minimax risk.

(b) Let \( p_\theta \) denote a multivariate normal density with mean \( \theta = (\theta_1, \ldots, \theta_d) \) and covariance equal to the identity matrix. Let \( a > 0 \) and define \( \Theta = \{\xi_1, \ldots, \xi_d\} \) where \( \xi_1 = (a, 0, \ldots, 0), \xi_2 = (0, a, \ldots, 0), \ldots, \xi_d = (0, \ldots, 0, a) \). Find \( a \) such that

\[
\inf_\hat{\theta} \max_{\xi \in \Theta} P_\xi (\hat{\xi} \neq \xi) \geq \frac{1}{2}
\]

Find an estimator that achieves (or comes close to achieving) this risk.