Homework 1
10-702/36-702 Statistical Machine Learning

Due: Friday Jan 27 3:00

Hand in to: Sharon Cavlovich GHC 8215.

1. (Review of Maximum Likelihood.) Let $X_1, \ldots, X_n$ be iid discrete random variables where $X_i \in \{1, \ldots, k\}$. Let $\theta = (\theta_1, \ldots, \theta_k)$ where $\theta_j = P(X_i = \theta_j)$. Hence, $\sum_{j=1}^k \theta_j = 1$.

   (a) Find the mle of θ.

   (b) Find the Fisher information matrix.

   (c) Find an asymptotic $1 - \alpha$ confidence interval for θ_1.

   (d) Prove that $\sum_{j=1}^k |\widehat{\theta}_j - \theta_j| \xrightarrow{P} 0$ as $n \to \infty$.

   (e) Now suppose we let $k = k(n)$ increase with n. (In this case, θ can change with n.) Show that, if $k(n)$ does not increase too quickly, then $\sum_{j=1}^{k(n)} |\widehat{\theta}_j - \theta_j| \xrightarrow{P} 0$ as $n \to \infty$.

2. Prove Theorem 1.52 from the notes.

3. Recall Bernstein's inequality. Let $X_1, \ldots, X_n$ be iid with mean μ and variance σ^2 and suppose that $|X_i| \leq c$. Then

$$\mathbb{P}(|\overline{X}_n - \mu| > \epsilon) \leq 2 \exp\left(- \frac{n\epsilon^2}{2\sigma^2 + \frac{2c\epsilon}{3}} \right)$$

   where $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$.

   (a) Show that, with probability at least $1 - 2\delta$,

$$|\overline{X}_n - \mu| \leq \sqrt{\frac{2\sigma^2 \log(1/\delta)}{n}} + \frac{2c \log(1/\delta)}{3n}.$$

   (b) Let $Y_1, \ldots, Y_n$ be iid random variables with bounded density f. Let $A_n = [-1/n, 1/n]$. Define $X_i = I(Y_i \in A_n)$. Let $\mu_n = \mathbb{E}(X_i) = \mathbb{P}(Y_i \in A_n)$. Use Bernstein's inequality to show that

$$|\overline{X}_n - \mu_n| = O_P\left(\frac{1}{n}\right).$$

   Note that Hoeffding's inequality yields the weaker result $|\overline{X}_n - \mu_n| = O_P\left(\frac{1}{\sqrt{n}}\right)$.

4. (Rademacher Complexity.) Prove Lemma 1.70.

5. Bracketing Numbers. Let $\mathcal{F}$ be the set of monotone non-decreasing functions on $[0, 1]$ such that $0 \leq f(x) \leq 1$. Show that for any distribution P,

$$N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \leq \exp(C/\epsilon).$$

for some constant $C > 0$.