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

Due: Friday Jan 21 3:00

Hand in to: Michelle Martin GHC 8001.

1. (Review of Maximum Likelihood.) Let $X_1, \ldots, X_n$ be iid random variables where $X_i \in \{1, 2, 3, 4\}$. Let $p_j = \mathbb{P}(X_i = j)$. Suppose there exists $0 < \theta < 6/11$ such that

$$p_1 = \theta, \ p_2 = \frac{\theta}{2}, \ p_3 = \frac{\theta}{3}, \ p_4 = \frac{6 - 11\theta}{6}.$$

   (a) Find the mle of $\theta$.

   (b) Find the Fisher information.

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

   (d) Prove that $\widehat{\theta} \xrightarrow{P} \theta$.

2. For any $t \in \mathbb{R}$ define $f_t(z) = \text{sign}(\sin(tz))$. Let $\mathcal{F} = \{f_t : \ t \in \mathbb{R}\}$. Show that $\mathcal{F}$ has infinite VC dimension. Hint: consider a set of points like $\{1/2, 1/4, \ldots, 1/2^n\}$.

3. Recall Bernstein's inequality. Let $X_1, \ldots, X_n$ be iid with mean $\mu$ and variance $\sigma^2$ 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 $1 - \delta$,

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

   (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.) Let $\mathcal{F} = \{f_1, \ldots, f_N\}$ where each $f$ is a binary function: $f(z) \in \{0, 1\}$. Show that

$$\mathcal{R}_n(\mathcal{F}) \leq 2\sqrt{\frac{\log N}{n}}.$$