36-755, Fall 2016
Homework 3

Due Wed Oct 5 by 5:00pm in Jisu's mailbox

1. A random matrix A of dimension $n \times m$ is sub-Gaussian with parameter σ^2, written as $A \in SG_{m,n}(\sigma^2)$, when $y^\top Ax$ is $SG(\sigma^2)$ for any $y \in \mathbb{S}^{n-1}$ and $x \in \mathbb{S}^{m-1}$. You may assume that $\mathbb{E}[A] = 0$ (or otherwise replace A by $A - \mathbb{E}[A]$).
   (a) Suppose that the entries of A are independent variables that are $SG(\sigma^2)$. Show that $A \in SG_{m,n}(\sigma^2)$.
   (b) Let $A \in SG_{n,m}(\sigma^2)$ and recall that the operator norm of A is
   $$\|A\|_{\text{op}} = \max_{x \in \mathbb{R}^m, x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{y \in \mathbb{S}^{n-1}, x \in \mathbb{S}^{m-1}} y^\top Ax.$$
   Show that, for some $C > 0$,
   $$\mathbb{E}\left[\|A\|_{\text{op}}\right] \leq C\left(\sqrt{n} + \sqrt{m}\right).$$
   (c) Find a concentration inequality for $\|A\|_{\text{op}}$.

   Hint: work with a 1/4 net for $\mathbb{S}^{n-1}$ and a 1/4 net for $\mathbb{S}^{m-1}$.

2. **Median and sample quantiles.**
   (a) Suppose that $(X_1, \ldots, X_n)$ is an i.i.d. sample from a distribution P (if you like, you may assume P to be absolutely continuous). Let $X_{(1)} \leq X_{(2)} < \ldots < X_{(n)}$ be the order statistics and set $\alpha \in (0,1)$. Determine a $1-\alpha$ confidence interval for the median of P of the form
   $$\left(X_{(k_1)}, X_{(k_2)}\right)$$
   for some choice of $k_1 < k_2$. Determine k_1 and k_2 by relating this problem to a $\text{Bin}(n, 1/2)$ distribution and use concentration.
   (b) Consider the same setting as the previous exercise and let F be the c.d.f. of P and $p \in (0,1)$. The pth quantile and p-th sample quantile are, respectively,
   $$\xi_p = \inf\{x \colon F(x) \geq p\}$$
   and
   $$\hat{\xi}_p = \inf\{x \colon F_n(x) \geq p\},$$
   res[ectively, where F_n is the sample c.d.f. (i.e. $F_n(x) = \frac{1}{n}\sum_{i=1}^n 1(X_i \leq x)$). Show that, for any $\epsilon > 0$,
   $$\mathbb{P}\left(|\hat{\xi}_p - \xi_p| > \epsilon\right) \leq 2\exp\left\{-2n\delta_\epsilon^2\right\},$$
   where $\delta_\epsilon = \min\left\{F(x_p + \epsilon) - p, p - F(\xi_p - \epsilon)\right\}$.
   Write, for instance, $\mathbb{P}\left(\hat{\xi}_p > \xi_p + \epsilon\right) = \mathbb{P}\left(p > F_n(\xi_p + \epsilon)\right)$. Then, notice that $F_n(x)$ is a sum of i.i.d. Bernoulli and use Hoeffding yet again...

3. Consider the linear regression model

$$Y = X\theta^* + \epsilon$$

where $\theta \in \mathbb{R}^d$, X is fixed and $\epsilon \in \mathbb{R}^n$ consists of independent zero-mean variables with finite variance. The ridge estimator is defined as

$$\hat{\theta}_{\text{ridge}} = \hat{\theta}_{\text{ridge}}(\lambda) = \text{argmin}_{\theta \in \mathbb{R}^d} \left\{ \frac{1}{n} \|Y - X\theta\|^2 + \lambda \|\theta\|^2 \right\},$$

where $\lambda > 0$.

   (a) Show that $\hat{\theta}_{\text{ridge}}$ is uniquely defined for any $\lambda > 0$ and find a closed-form expression.

   (b) Compute the bias of $\hat{\theta}_{\text{ridge}}$.

4. Let $(\mathcal{X}, d)$ be a metric space and, for $\delta > 0$, let $N(\delta)$ and $M(\delta)$ denote the δ-covering and δ-packing number, respectively. Show that

$$M(2\delta) \leq N(\delta) \leq M(\delta).$$