36-755, Fall 2017
Homework 5

Due Wed Nov 15 by 5:00pm in Jisu's mailbox

1. In this exercise you will fill in some of the details from the proof of the upper bound for sparse PCA under a spike covariance model.

   (a) For $p \geq 1$ the Schatten p-norm of a $n \times m$ matrix A is the ℓ_p norm of its singular values:

   $$\|A\|_p = \left(\sum_{i=1}^{r} \sigma_i^p \right)^{1/p},$$

   where $\sigma \geq \sigma_2 \geq \ldots \sigma_r \geq 0$ are the singular values of A and $r = \min\{m, n\}$. Prove the non-commutative Hölder inequality for conformal matrices A and B:

   $$|\mathrm{tr}(A^\top B)| \leq \|A\|_p \|B\|_q,$$

   for all $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$.

   (b) Let u and v be two unit norm vectors in $\mathbb{R}^d$. Show that

   $$\sqrt{2}\|uu^\top - vv^\top\|_{\mathrm{op}} = \|uu^\top - vv^\top\|_F = \sqrt{2 - 2(u^\top v)^2} = \sqrt{2\sin^2\left(\angle(u, v)\right)}.$$

   where $\angle(u, v) = \cos^{-1}(|u^\top v|)$

2. This is a result that I cited when discussing spectra clustering for stochastic block models. 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 $y^\top A x$ 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

   $$\|A\|_{\mathrm{op}} = \max_{x \in \mathbb{R}^m, x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{y \in \mathbb{S}^{n-1}, x \in \mathbb{S}^{m-1}} y^\top A x.$$

   Show that, for some $C > 0$,

   $$\mathbb{E}\left[\|A\|_{\mathrm{op}}\right] \leq C\left(\sqrt{n} + \sqrt{m}\right).$$

   (c) Find a concentration inequality for $\|A\|_{\mathrm{op}}$.

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

3. Exercise 8.4

4. Suppose we observe an i.i.d. sample $X_1, \ldots, X_n$ from the mixture distribution

   $$\frac{1}{2}P_1 + \frac{1}{2}P_2,$$

where $P_1 = N_d(\mu, I_d)$ and $P_2 = N_d(-\mu, I_d)$, with $\mu \in \mathbb{R}^d$ a non-zero vector. Ours task is to cluster the sample points into two groups, where points in the same group originated from the same component of the mixture (i.e. Either P_1 or P_2)

We will use spectral clustering: compute the leading eigenvector $\hat{\nu}$ of the empirical covariance matrix and cluster the points depending on the sign of $X_i^\top \hat{\nu}$, $i = 1, \dots, n$.

Use the Davis-Kahan theorem to derive an upper bound on the proportion of misclustered nodes.

5. Exercise 4.3

6. **Massart's finite class Lemma** Let $\mathcal{A}$ be a finite subset of $\mathbb{R}^n$ and let $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ be a vector of i.i.d. Rademacher variables. Show that

$$\mathbb{E}\left[\frac{1}{n}\sup_{a \in \mathcal{A}} a^\top \epsilon\right] \leq D\frac{\sqrt{2\log|\mathcal{A}|}}{n}$$

where $D = \max_{a \in \mathcal{A}} \|a\|$. Use this result to prove Lemma 4.1 of Chapter 4. (In proving both claims it is OK if you get get different constants; know however that the constants in Lemma 4.1 are sub-optimal).