36-755, Fall 2016
Homework 5

Due Mon Oct 31 by 5:00pm in Jisu's mailbox

1. **Matrix Algebra Problems. We have used these facts in class. Now it is time to prove them.**

   (a) Problem 8.3 (You may assume the result of Problem 8.1 as given).

   (b) Recall the spiked covariance model: $\Sigma = \theta v v^\top + I_d$, where $\theta > 0$ and $v \in \mathbb{S}^{d-1}$. Let $\hat{v}$ be another unit vector in $\mathbb{S}^{d-1}$. Show that

   $$v^\top \Sigma v - \hat{v}^\top \Sigma \hat{v} = \theta \sin^2(\angle(v, \hat{v}))$$

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

   (c) Show that

   $$\left\| \hat{v}\hat{v}^\top - vv^\top \right\|_F^2 = 2\sin^2(\angle(v, \hat{v})),$$

   where, for a matrix $A = (A_{i,j})$, $\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$.

2. **Reading exercise**. Skim through the article "Models as Approximations A Conspiracy of Random Regressors and Model Deviations Against Classical Inference in Regression", available here:
   http://www-stat.wharton.upenn.edu/~buja/PAPERS/Buja_et_al_A_Conspiracy-rev1.pdf

   (a) Prove equations (7) and (8) in the paper. That is, prove the following. Let X be a zero mean d-dimensional vector and Y a random variable. The joint distribution of (X, Y) is P. Let $\Sigma = \mathbb{E}[XX^\top]$ and $\alpha = \mathbb{E}[Y \cdot X]$. Show that $\beta = \Sigma^{-1}\alpha$ is the solution to

   $$\text{argmin}_{\beta \in \mathbb{R}^d} \mathbb{E}\left[(Y - X^\top \beta)^2\right]$$

   and to

   $$\text{argmin}_{\beta \in \mathbb{R}^d} \mathbb{E}\left[(\mu(X) - X^\top \beta)^2\right],$$

   where $\mu(X) = \mathbb{E}[Y|X]$.

   (b) Did this article contain results that were surprising to you? Think about fitting a linear regression model to a model where $\mu(X)$ is non-lineat: what happens? Write no more than 2 paragraphs.

3. Exercise 4.3.

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

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

   where $D = \max_{a \in \mathcal{A}} \|a\|$. Use this result to prove Lemma 4.1 of Chapter 4. (In proving both claims it will be OK to get different constants).

5. **Another symmetrization inequality**

Let $\mathcal{A}$ be a countable class of sets in $\mathbb{R}^d$ and $X = (X_1, \ldots, X_n) \overset{i.i.d.}{\sim} P$. Let P_n^X be the empirical measure corresponding to the sample X. Let $Y = (Y_1, \ldots, Y_n) \overset{i.i.d.}{\sim} P$ be a ghost sample, independent of Y, and P_n^Y the corresponding empirical measure. Prove that

$$\mathbb{P}\left(\sup_{A\in\mathcal{A}} |P_n^X(A) - P(A)| > \epsilon\right) \le 2\mathbb{P}\left(\sup_{A\in\mathcal{A}} |P_n^X(A) - P_n^Y(A)| > \epsilon/2\right),$$

for $n\epsilon^2 \ge 2$.

Proceed as follows:

(a) Show that, if $n\epsilon^2 \ge 2$, for all $A \in \mathcal{A}$,

$$\mathbb{P}\left(|P_n^X(A) - P(A)| > \epsilon\right) \le 1/2.$$

(b) Prove the following claim: let $(Z_k, k = 1, 2, \ldots)$ be a sequence (finite or infinite) of random variables and $(Z'_k, k = 1, 2, \ldots)$ be a ghost sequence with the same distribution and independent of it. Suppose that $\mathbb{P}(|Z_k| > \epsilon/2) = \mathbb{P}(|Z'_k| > \epsilon/2) \le 1/2$ for all k. Then

$$\mathbb{P}\left(\sup_k |Z_k| > \epsilon\right) \le 2\mathbb{P}\left(\sup_k |Z_k - Z'_k| > \epsilon/2\right).$$

Hint: Define the events: $A_1 = \{|Z_i| > \epsilon\}$ *and for* $k \ge 2$,

$$A_k = \{|Z_1| \le \epsilon, \ldots, |Z_{k-1}| \le \epsilon, |Z_k| > \epsilon\}.$$

Then,

$$\frac{1}{2}\mathbb{P}\left(\max_k |Z_k| > \epsilon\right) = \frac{1}{2}\sum_k \mathbb{P}(A_k).$$

Proceed...

6. Exercise 4.10.

7. Exercise 4.13.