STA 561: Homeworks 5 and 6 (Due April 5 midnight)

Reminder: work together! Share ideas, brainstorm, explain/verify your answers but write up your own work. Your homework should be submitted as pdf file generated using either latex or an python notebook. (Problems 1 and 2 will be considered HW 4 while problems 3 and 4 will be considered HW 5).

1. (Some simple calculations.) Suppose that $\Omega \in \mathbb{R}^{k\times p}$ is populated with Rademacher random variables (i.e., $\Omega_{i,j} \sim \text{Uniform}\{-1, 1\}$ for all i, j and the entries are mutually independent). For any $\boldsymbol{x} \in \mathbb{R}^p$, show that $\mathbb{E}||k^{-1/2}\Omega\boldsymbol{x}||^2 = ||\boldsymbol{x}||^2$. Suppose we want to generate Ω by filling it with i.i.d. entries from some distribution Q, give simple conditions on Q so that $\mathbb{E}||k^{-1/2}\Omega\boldsymbol{x}||^2 = ||\boldsymbol{x}||^2$ for all $\boldsymbol{x} \in \mathbb{R}^p$.

2. (Streamy McStreamface.) Assume the observed data are $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n$ which comprise i.i.d. pairs $(\boldsymbol{X}, Y)$ where $\boldsymbol{X} \in \mathbb{R}^p$ and $Y\mathbb{R}$. Let $\Omega \in \mathbb{R}^{k\times p}$ be a random projection matrix which satisfies the conditions you gave in the preceding problem. Define

$$\widehat{\boldsymbol{\beta}}_n^{\Omega} = \Omega^{\intercal} \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^k} \mathbb{P}_n \left\{Y - (\Omega\boldsymbol{X})^{\intercal}\boldsymbol{\beta}\right\}^2 .$$

Suppose we are interested in computing the averaged projected least squares estimator:

$$\widehat{\boldsymbol{\beta}}_n^{\text{ave}} = \frac{1}{B}\sum_{b=1}^{B} \widehat{\boldsymbol{\beta}}_n^{\Omega^{(b)}},$$

where $\Omega^{(1)}, \ldots, \Omega^{(B)}$ are i.i.d. projection matrices of dimension $\mathbb{R}^{k\times p}$. Give an algorithm that can compute $\widehat{\boldsymbol{\beta}}_n^{\text{ave}}$ with a single pass through the data with storage requirement $O(p^2)$ regardless of B.

3. (Randy.) The reduced dimension, k, is a potentially important tuning parameter. One approach to possibly mitigating sensitivity to this choice is to average a large number of estimators across randomly generated values of k. That is, we generate $k^1, \ldots, k^B \sim P(k)$, where P is a distribution on $\{k_{\min}, k_{\min} + 1, \ldots, k_{\max}\}$, and subsequently compute $\Omega^{(b)} \in \mathbb{R}^{k^{(b)}\times p}$ and

$$\widehat{\boldsymbol{\beta}}_n^{\text{ave}} = \frac{1}{B}\sum_{b=1}^{B} \widehat{\boldsymbol{\beta}}_n^{\Omega^{(b)}}.$$

Given a streaming estimator of this version of $\widehat{\boldsymbol{\beta}}_n^{\text{ave}}$ and provide an implementation of your algorithm in python or R.

4. (k-dim embedding.) We proved that random projections approximately preserve inner products. Given data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ define $f_n : \mathbb{R}^{k \times p} \to \mathbb{R}_+$ by

$$f_n(\Omega) = \sum_{i,j} \left\{ \boldsymbol{x}_i^\intercal \boldsymbol{x}_j - (\Omega \boldsymbol{x}_i)^\intercal (\Omega \boldsymbol{x}_j) \right\}^2 .$$

A natural alternative to random projections is to derive the k-dim embedding

$$\widehat{\Omega}_n = \arg \min_{\Omega \in \mathbb{R}^{k \times p}} f_n(\Omega).$$

Design and conduct a simulation study to compare random normal projections with the k-dim embedding in terms of the objective f_n and the out-of-sample MSE of the least-squares estimator fit to the projected inputs.