18.S096: Homework Problem Set 3

Topics in Mathematics of Data Science (Fall 2015)

Afonso S. Bandeira

Due on November 3, 2015

Problem 3.1 *Given n i.i.d. non-negative random variables $x_1, \ldots, x_n$, show that*

$$\mathbb{E} \max_i x_i \lesssim \left(\mathbb{E}\left[x_1^{\log n} \right] \right)^{\frac{1}{\log n}}.$$

Problem 3.2 *Let $y_1, \ldots, y_n \in \mathbb{R}^d$ be i.i.d. random variables such that $\mathbb{E} y_k = 0$ and*

$$\mathbb{E} y_k y_k^T = I_{d\times d}.$$

Show that, if

$$\frac{\log d}{n} \left(\mathbb{E}\|y_1\|^{2\log n} \right)^{\frac{1}{\log n}} \lesssim 1,$$

then

$$\mathbb{E} \left\| \left(\frac{1}{n} \sum_{k=1}^{n} y_k y_k^T \right) - I_{d\times d} \right\| \lesssim \sqrt{\frac{\log d}{n}} \left(\mathbb{E}\|y_1\|^{2\log n} \right)^{\frac{1}{2\log n}}$$

Note that $\|\cdot\|$ denotes spectral norm, and $\lesssim$ means smaller up to constants.

Problem 3.3 *Given a centered*[1] *random symmetric matrix $X \in \mathbb{R}^{d\times d}$, we define*

$$\sigma = \sqrt{\|\mathbb{E}X^2\|},$$

and

$$\sigma_* = \max_{v:\ \|v\|=1} \sqrt{\mathbb{E}\left(v^T X v\right)^2},$$

1. *Show that $\sigma \geq \sigma_*$.*
2. *If X has independent entries (except for the fact that $X_{ij} = X_{ji}$) such that $X_{ij} \sim \mathcal{N}\left(0, b_{ij}^2\right)$, show that*

[1] Meaning that $\mathbb{E}X = 0$.

- $\sigma^2 = \max_i \sum_{j=1}^{n} b_{ij}^2$
- $\sigma_* \leq 2 \max_{ij} |b_{ij}|$

Note that $\|\cdot\|$ *denotes spectral norm, and in expressions with* $\mathbb{E}$ *and a power, the power binds first. For example, by* $\mathbb{E}X^2$*, we mean* $\mathbb{E}\left[X^2\right]$ *and, by* $\mathbb{E}\left(v^T X v\right)^2$*, we mean* $\mathbb{E}\left[\left(v^T X v\right)^2\right]$

Problem 3.4 (Norm concentration of projection) *Let* $y_1, \ldots, y_d$ *be i.i.d standard Gaussian random variables and* $Y = (y_1, \ldots, y_d)$*. Let* $g : \mathbb{R}^d \to \mathbb{R}^k$ *be the projection into the first* k *coordinates and* $Z = g\left(\frac{Y}{\|Y\|}\right) = \frac{1}{\|Y\|}(y_1, \ldots, y_k)$ *and* $L = \|Z\|^2$*. It is clear that* $\mathbb{E}L = \frac{k}{d}$*. Prove that* L *is very concentrated around its mean*

- *If* $\beta < 1$,
$$\Pr\left[L \leq \beta \frac{k}{d}\right] \leq \exp\left(\frac{k}{2}(1 - \beta + \log \beta)\right).$$
- *If* $\beta > 1$,
$$\Pr\left[L \geq \beta \frac{k}{d}\right] \leq \exp\left(\frac{k}{2}(1 - \beta + \log \beta)\right).$$

1. *Prove*
$$\Pr\left[\sum_{i=1}^{k} y_i^2 \leq \beta \frac{k}{d} \sum_{i=1}^{d} y_i^2\right] \leq \frac{(1 - 2tk\beta)^{-(d-k)/2}}{(1 - 2t(k\beta - d))^{k/2}}$$
for any $t > 0$ *such that* $1 - 2tk\beta > 0$ *and* $1 - 2t(k\beta - d) > 0$.
(Hint: Prove, and use, that $\mathbb{E}(e^{sX^2}) = \frac{1}{\sqrt{1-2s}}$*, for* $-\infty < s < 1/2$*, where* X *is standard normal.)*
2. *Find a suitable* t *and conclude the proof of the inequality for* $\beta < 1$.
3. *Use the same idea as above to prove the* $\beta > 1$ *case.*

18.S096 Topics in Mathematics of Data Science
Fall 2015