

DS-GA 3001: Homework Problem Set 3

Special Topics in Data Science: (Spring 2018)

MATHEMATICS OF DATA SCIENCE: Graphs and Networks

Afonso S. Bandeira

`bandeira@cims.nyu.edu`

<http://www.cims.nyu.edu/~bandeira>

Shuyang Ling

`sling@cims.nyu.edu`

<http://www.cims.nyu.edu/~sling>

Due on April 5, 2018

This homework problem set is due on April 5, before class at the homework dropbox in the CDS reception, or by email to the graders.

If you have questions about the homework feel free to contact me or Shuyang, or stop by our office hours.

Try not to look up the answers, you'll learn much more if you try to think about the problems without looking up the solutions. If you need hints, feel free to email me or Shuyang.

You can work in groups but each student must write his/her own solution based on his/her own understanding of the problem. Please list, on your submission, the students you work with for the homework (this will not affect your grade).

Late submissions will be graded with a penalty of 10% per day late.

If you need to impose extra conditions on a problem to make it easier (or consider specific cases of the question, like taking n to be 2, e.g.), state explicitly that you have done so. Solutions where extra conditions were assumed, or where only special cases were treated, will also be graded (probably scored as a partial answer).

Random Matrices

Recall the definition of a standard gaussian Wigner Matrix W : a symmetric random matrix $W \in \mathbb{R}^{n \times n}$ whose diagonal and upper-diagonal entries are independent $W_{ii} \sim \mathcal{N}(0, 2)$ and, for $i < j$, $W_{ij} \sim \mathcal{N}(0, 1)$. This random matrix ensemble is invariant under orthogonal conjugation: $U^T W U \sim W$ for any $U \in O(n)$. Also, the distribution of the eigenvalues of $\frac{1}{\sqrt{n}}W$ converges to the so-called semicircular law with support $[-2, 2]$

$$dSC(x) = \sqrt{4 - x^2} 1_{[-2, 2]}(x).$$

(try it in Matlab, draw an histogram of the distribution of the eigenvalues of $\frac{1}{\sqrt{n}}W$ for, say $n = 500$.)

In the next problem, you will show that the largest eigenvalue of $\frac{1}{\sqrt{n}}W$ has expected value at most 2.¹ For that, we will make use of Slepian's Comparison Lemma.

Slepian's Comparison Lemma is a crucial tool to compare Gaussian Processes. A Gaussian process is a family of gaussian random variables indexed by some set T , more precisely is a family of gaussian random variables $\{X_t\}_{t \in T}$ (if T is finite this is simply a gaussian vector). Given a gaussian process X_t , a particular quantity of interest is $\mathbb{E}[\max_{t \in T} X_t]$. Intuitively, if we have two Gaussian processes X_t and Y_t with mean zero $\mathbb{E}[X_t] = \mathbb{E}[Y_t] = 0$, for all $t \in T$ and same variances $\mathbb{E}[X_t^2] = \mathbb{E}[Y_t^2]$ then the process that has the "least correlations" should have a larger maximum (think the maximum entry of vector with i.i.d. gaussian entries versus one always with the same gaussian entry). A simple version of Slepian's Lemma makes this intuition precise:²

In the conditions above, if for all $t_1, t_2 \in T$

$$\mathbb{E}[X_{t_1} X_{t_2}] \leq \mathbb{E}[Y_{t_1} Y_{t_2}],$$

then

$$\mathbb{E} \left[\max_{t \in T} X_t \right] \geq \mathbb{E} \left[\max_{t \in T} Y_t \right].$$

A slightly more general version of it asks that the two Gaussian processes X_t and Y_t have mean zero $\mathbb{E}[X_t] = \mathbb{E}[Y_t] = 0$, for all $t \in T$ but not

¹Note that, a priori, there could be a very large eigenvalue and it would still not contradict the semicircular law, since it does not predict what happens to a vanishing fraction of the eigenvalues.

²Although intuitive in some sense, this is a delicate statement about Gaussian random variables, it turns out not to hold for other distributions.

necessarily the same variances. In that case it says that: If or all $t_1, t_2 \in T$

$$\mathbb{E}[X_{t_1} - X_{t_2}]^2 \geq \mathbb{E}[Y_{t_1} - Y_{t_2}]^2, \quad (1)$$

then

$$\mathbb{E} \left[\max_{t \in T} X_t \right] \geq \mathbb{E} \left[\max_{t \in T} Y_t \right].$$

Problem 1.1 We will use Slepian's Comparison Lemma to show that

$$\mathbb{E} \lambda_{\max}(W) \leq 2\sqrt{n}.$$

1. Note that

$$\lambda_{\max}(W) = \max_{v: \|v\|_2=1} v^T W v,$$

which means that, if we take for unit-norm v , $Y_v := v^T W v$ we have that

$$\lambda_{\max}(W) = \mathbb{E} \left[\max_{v \in \mathbb{S}^{n-1}} Y_v \right],$$

2. Use Slepian to compare Y_v with $2X_v$ defined as

$$X_v = v^T g,$$

where $g \sim \mathcal{N}(0, I_{n \times n})$

3. Use Jensen's inequality to upperbound $\mathbb{E}[\max_{v \in \mathbb{S}^{n-1}} X_v]$.

Problem 1.2 Given a centered³ random symmetric matrix $X \in \mathbb{R}^{d \times d}$, we define

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

and

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

1. Show that $\sigma \geq \sigma_*$.

2. If X has independent entries (except for the fact that $X_{ij} = X_{ji}$) such that $X_{ij} \sim \mathcal{N}(0, b_{ij}^2)$, show that

$$\bullet \sigma^2 = \max_i \sum_{j=1}^n b_{ij}^2$$

³Meaning that $\mathbb{E} X = 0$.

- $\sigma_* \leq 2 \max_{ij} |b_{ij}|$

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 $\mathbb{E}[X^2]$ and, by $\mathbb{E}(v^T X v)^2$, we mean $\mathbb{E}[(v^T X v)^2]$

Multidimensional scaling

Problem 1.3 (Multidimensional Scaling) Suppose you want to represent n data points in \mathbb{R}^d and all you are given is estimates for their Euclidean distances $\delta_{ij} \approx \|x_i - x_j\|_2^2$. Multidimensional scaling attempts to find an d dimensions that agrees, as much as possible, with these estimates. Organizing $X = [x_1, \dots, x_n]$ and consider the matrix Δ whose entries are δ_{ij} .

1. Show that, if $\delta_{ij} = \|x_i - x_j\|_2^2$ then there is a choice of x_i (note that the solution is not unique, as a translation of the points will preserve the pairwise distances, e.g.) for which

$$X^T X = -\frac{1}{2} H \Delta H,$$

where $H = I - \frac{1}{n} \mathbf{1}\mathbf{1}^T$.

2. If the goal is to find points in \mathbb{R}^d , how would you do it (keep part 1 of the question, and Principal Component Analysis, in mind)?

(The procedure you have just derived is known as Multidimensional Scaling)

This motivates a way to embed a graph in d dimensions. Given two nodes we take δ_{ij} to be the square of some natural distance on a graph such as, for example, the geodesic distance (the distance of the shortest path between the nodes) and then use the ideas above to find an embedding in \mathbb{R}^d for which Euclidean distances most resemble geodesic distances on the graph. This is the motivation behind a dimension reduction technique called ISOMAP (J. B. Tenenbaum, V. de Silva, and J. C. Langford, Science 2000).

Problem 1.4 Given n i.i.d. non-negative random variables x_1, \dots, x_n , show that

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