18.S096: Homework Problem Set 1 (revised)

Topics in Mathematics of Data Science (Fall 2015)

Afonso S. Bandeira

Due on October 6, 2015 Extended to: October 8, 2015

This homework problem set is due on October 6, at the start of class.

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.

You can work in groups but each student must write his/her own solution based on his/her own understanding of the problem.

If you need to impose extra conditions on a problem to make it easier, state explicitly that you have done so. Solutions where extra conditions were assumed will also be graded (probably scored as a partial answer).

1.1 Linear Algebra

Problem 1.1 *Show the resut we used in class: If $M \in \mathbb{R}^{n\times n}$ is a symmetric matrix and $d \leq n$ then*

$$\max_{\substack{U\in\mathbb{R}^{n\times d}\\ U^TU=\mathrm{I}_{d\times d}}} \operatorname{Tr}\left(U^TMU\right) = \sum_{k=1}^{d} \lambda_k^{(+)}(M),$$

where $\lambda_k^{(+)}$ is the largest k-th eigenvalue of M.

1.2 Estimators

Problem 1.2 *Given $x_1, \cdots, x_n$ i.i.d. samples from a distribution X with mean μ and covariance Σ, show that*

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad \textit{and} \quad \Sigma_n = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \mu_n\right)\left(x_k - \mu_n\right)^T,$$

are unbiased estimators for μ and Σ, i.e., show that $\mathbb{E}\left[\mu_n\right] = \mu$ and $\mathbb{E}\left[\Sigma_n\right] = \Sigma$.

1.3 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 emsemble 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 $[-2,2]$

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

(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 $\mathbb{E}\left[\max_{t\in T} X_t\right]$. Intuitively, if we have two Gaussian processes X_t and Y_t with mean zero $\mathbb{E}\left[X_t\right] = \mathbb{E}\left[Y_t\right] = 0$, for all $t \in T$ and same variances $\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[Y_t^2\right]$ 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

[1]Note 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.

gaussian entry). A simple version of Slepian's Lemma makes this intuition precise:[2]

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

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

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 $\mathbb{E}\left[X_t\right] = \mathbb{E}\left[Y_t\right] = 0$, for all $t \in T$ but not necessarily the same variances. In that case it says that: If or all $t_1, t_2 \in T$

$$\mathbb{E}\left[X_{t_1} - X_{t_2}\right]^2 \geq \mathbb{E}\left[Y_{t_1} - Y_{t_2}\right]^2, \tag{1}$$

then

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

Problem 1.3 *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 $g \sim \mathcal{N}\left(0, \mathrm{I}_{n \times n}\right)$

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

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

Problem 1.4 *In this problem you'll derive the limit of the largest eigenvalue of a rank 1 perturbation of a Wigner matrix.*

For this problem, you don't have to justify all of the steps rigorously. You can use the same level of rigor that was used in class to derive the analogue result for sample covariance matrices. Deriving this phenomena rigorously would take considerably more work and is outside of the scope of this homework.

Consider the matrix $M = \frac{1}{\sqrt{n}}W + \beta vv^T$ *for* $\|v\|_2 = 1$ *and* W *a standard Gaussian Wigner matrix. The purpose of this homework problem is to understand the behavior of* $\lambda_{\max}(M)$. *Because* W *is invariant to orthogonal conjugation we can focus on understanding*

$$\lambda_{\max}\left(\frac{1}{\sqrt{n}}W + \beta e_1 e_1^T\right).$$

Use the same techniques as used in class to derive the behavior of this quantity.

(Hint: at some point, you'll probably have to integrate $\int_{-2}^{2}\frac{\sqrt{4-x^2}}{y-x}dx$. *You can use the fact that, for* $y > 2$, $\int_{-2}^{2}\frac{\sqrt{4-x^2}}{y-x}dx = \pi\left(y - \sqrt{y^2-4}\right)$ *(you can also use an integrator software, such as Mathematica, for this).*

1.4 Diffusion Maps and other embeddings

Problem 1.5 *The ring graph on* n *nodes is a graph where node* $1 < k < n$ *is connected to node* $k-1$ *and* $k+1$ *and node* 1 *is connected to node* n. *Derive the two-dimensional diffusion map embedding for the ring graph (if the eigenvectors are complex valued, try creating real valued ones using multiplicity of the eigenvalues). Is it a reasonable embedding of this graph in two dimensions?*

Problem 1.6 (Multidimensional Scaling *Revised*) *Suppose you want to represent* n *data points in* $\mathbb{R}^d$ *and all you are given is estimates for their Euclidean distances* $\delta_{ij} \approx \|x_i - x_j\|_2^2$. *Multiimensional scaling attempts to find an* d *dimensions that agrees, as much as possible, with these estimates. Organizing* $X = [x_1, \ldots, x_n]$ *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 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).*