ST 790 Assignment 3

David Elsheimer and Jimmy Hickey

10/15/2020

Instruction

This assignment consists of 3 problems. The assignment is due on **Friday, October 16** at 11:59pm EDT. Please submit your assignment electronically through the **Moodle** webpage. The assignment can be done as a group with at most 3 members per group (please include the name of the group members on the front page of the assignment).

Problem 1

In class, we only proved matrix concentration inequalities for symmetric matrices. Specifically, Theorem 9 in the lecture notes states that

Theorem (Tropp) Let $X_1, X_2, \ldots, X_n$ be independent mean 0 $d \times d$ **symmetric** random matrices and suppose that $\|X_i\| \leq M$ almost surely for all i. Define

$$\sigma^2 = \left\| \sum_{i=1}^{n} \mathbb{E}[X_i^2] \right\|$$

Then for every $t > 0$

$$\mathbb{P}\left(\left\| \sum_{i=1}^{n} X_i \right\| > t \right) \leq 2d \exp\left(\frac{-t^2}{\sigma^2 + Mt/3} \right)$$

Extend this result to general matrices with real-valued entries. That is, show the following corollary

Corollary Let $X_1, X_2, \ldots, X_n$ be independent mean 0 $d_1 \times d_2$ random matrices and suppose that $\|X_i\| \leq M$ almost surely for all i. Define

$$\sigma^2 = \max\left\{ \left\| \sum_{i=1}^{n} \mathbb{E}[X_i X_i^\top] \right\|, \left\| \sum_{i=1}^{n} \mathbb{E}[X_i^\top X_i] \right\| \right\}.$$

Then for every $t > 0$

$$\mathbb{P}\left(\left\| \sum_{i=1}^{n} X_i \right\| > t \right) \leq (d_1 + d_2) \exp\left(\frac{-t^2}{\sigma^2 + Mt/3} \right).$$

Hint Relate the singular values of a rectangular matrix M with the eigenvalues of its **symmetric dilation**

$$\begin{bmatrix} 0 & M \\ M^\top & 0 \end{bmatrix}.$$

Answer:

Note that the Hermitean dilation $H = \begin{bmatrix} 0 & M \\ M^\top & 0 \end{bmatrix}$ is symmetric. Then note that $H^2 = H \cdot H = \begin{bmatrix} MM^\top & 0 \\ 0 & M^\top M \end{bmatrix}$. This is a symmetric, block diagonal matrix. The eigenvalues of H^2 are the eigenvalues of $MM^\top, M^\top M$. In turn, these are the same as the quare of the singular values of $M, M^\top$. As such, the spectral radius of M is the same as the largest eigenvalues of H.

Where H is as previously defined, we can then consider $H = \sum_{i=1}^{n} H_i$, such that $H_i = \begin{bmatrix} 0 & X_i \\ X_i^\top & 0 \end{bmatrix}$.

Using this, the proof follows similarly to the proof for Tropp.

Note first that $\left\| \sum_{i=1}^{n} X_i \right\| = \left\| \sum_{i=1}^{n} H_i \right\| = \lambda_{max}(Z_n)$, where $Z_n = \sum_{i=1}^{n} H_i$.

Note further that $\sigma^2 = \max\left\{ \left\| \sum_{i=1}^{n} \mathbb{E}[X_i X_i^\top] \right\|, \left\| \sum_{i=1}^{n} \mathbb{E}[X_i^\top X_i] \right\| \right\} = \left\| \sum_{i=1}^{n} H_i^2 \right\|$.

Thus, the remainder of the proof is as follows

$$P\left(\left\| \sum_{i=1}^{n} H_i \right\| > t \right) \le e^{-st} E(\exp(s\lambda_{max}(Z_n)))$$

$$\le e^{-st} E(\text{tr } (\exp(\sum_{i=1}^{n} sH_i)))$$

$$= e^{-st} E(E(\text{tr } \exp\left(\sum_{i=1}^{n-1} sH_i + sH_n \right) |\{H_i\}_{i=1}^{n-1}))$$

$$\le e^{-st} E(\text{tr } \exp\left(\sum_{i=1}^{n-1} sH_i \right) + \log E(e^{sH_n}))$$

$$\le \cdots$$

$$\le e^{-st} \text{tr } \exp\left(\sum_{i=1}^{n} \log E(e^{sH_i}) \right), \text{ [Via corollary 13.]}$$

$$\le e^{-st} \text{tr } \exp\left(\sum_{i=1}^{n} \frac{s^2/2}{1 - |s|M/3} E(H_i^2) \right), \text{ [Via Proposition 14.]}$$

$$\le (d_1 + d_2) e^{-st} \lambda_{max}\left(\exp(\frac{s^2/2}{1 - |s|M/3} \sum_{i=1}^{n} E(H_i^2)) \right), \text{ [Via Golden-Thompson.]}$$

$$\le (d_1 + d_2) e^{-st} \exp\left(\frac{s^2/2}{1 - |s|M/3} \sum_{i=1}^{n} \lambda_{max}(E(H_i^2)) \right)$$

$$= (d_1 + d_2) e^{-st} \exp\left(\frac{s^2/2}{1 - |s|M/3} \left\| \sum_{i=1}^{n} H_i^2 \right\| \right)$$

$$= (d_1 + d_2) e^{-st} \exp\left(\frac{\sigma^2 s^2/2}{1 - |s|M/3} \right)$$

For $s = \frac{t}{\sigma^2 + Mt/3}$, this bound is $(d_1 + d_2) \exp(\frac{-t^2}{\sigma^2 + Mt/3}) \exp\left(t^2/2 \right)$ which is always greater than or equal to $(d_1 + d_2) \exp(\frac{-t^2}{\sigma^2 + Mt/3})$.

Thus for independent matrices X_i with dimension $d_1 \times d_2$, $P(\left\|\sum_{i=1}^n X_i\right\| > t) \le (d_1 + d_2) \exp(\frac{-t^2}{\sigma^2 + Mt/3})$.

Problem 2

To continue with the theme of the previous problem, in our class we only prove matrix concentration inequalities for the case when the summands are bounded almost surely in spectral norm. See for example the statement of the result in the previous problem. For this problem, you are asked to derive a matrix concentration inequality in the setting where the matrix powers of X_i satisfy a Bernstein-like tail conditions. In particular, prove the following result.

Theorem Let $X_1, X_2, \ldots, X_n$ be independent mean 0, symmetric random $d \times d$ matrices. Furthermore, assume that there exists a constant $R > 0$ and symmetric matrices $A_1, A_2, \ldots, A_n$ such that for any i

$$\mathbb{E}[X_i^p] \preceq \frac{p!}{2} R^{p-2} A_i^2, \quad \text{for } p = 2, 3, 4, \ldots.$$

Here $A \preceq B$ if $B - A$ is a positive semidefinite matrix. Now define

$$\sigma^2 = \left\| \sum_{i=1}^n A_i^2 \right\|$$

Then for every $t > 0$

$$\mathbb{P}\left(\left\| \sum_{i=1}^n X_i \right\| > t \right) \leq 2d \exp\left(\frac{-t^2/2}{\sigma^2 + Rt} \right)$$

Next apply the above result to derive a tail bound for the following (weighted) random graph model. Let P be a $n \times n$ symmetric matrix with non-negative, finite entries p_{ij}. Let A be a $n \times n$ symmetric matrix whose entries a_{ij} for $i \leq j$ are independent and that $a_{ij} \sim \text{Pois}(p_{ij})$. Sketch a tail bound for $\|A - P\|$ similar in flavor to Theorem 12 in the set of lecture notes for matrix concentration inequalities. (You don't have to pay too much attention to constant factors; we are only interested in important quantities similar to the $\sqrt{\Delta \log n}$ in the statement of Theorem 12 there).

Answer:

The first portion of this problem requires a proof follows similarly to the proof for theorem 11, with some minor adjustments. However, to avoid any confusion from skipping steps, the full proof of this portion is provided. Note $Z_n = \sum_{i=1}^n X_i$, $E(X_i^2) = A_i^2$.

$$P(\left\|\sum_{i=1}^{n} X_i\right\| > t) \le e^{-st} E(\exp(s\lambda_{max}(Z_n)))$$

$$\le e^{-st} E(\text{tr } (\exp(\sum_{i=1}^{n} sX_i)))$$

$$= e^{-st} E(E(\text{tr } \exp\left(\sum_{i=1}^{n-1} sX_i + sX_n\right) |\{X_i\}_{i=1}^{n-1}))$$

$$\le e^{-st} E(\text{tr } \exp\left(\sum_{i=1}^{n-1} sX_i\right) + \log E(e^{sX_n}))$$

$$\le \cdots$$

$$\le e^{-st}\text{tr } \exp\left(\sum_{i=1}^{n} \log E(e^{sX_i})\right), \text{ [Via corollary 13.]}$$

$$\le e^{-st}\text{tr } \exp\left(\sum_{i=1}^{n} \frac{s^2/2}{1-|s|M/3} E(X_i^2)\right), \text{ [Via Proposition 14.]}$$

$$\le de^{-st}\lambda_{max}\left(\exp(\frac{s^2/2}{1-|s|M/3} \sum_{i=1}^{n} E(X_i^2))\right), \text{ [Via Golden-Thompson.]}$$

$$\le de^{-st}\exp\left(\frac{s^2/2}{1-|s|M/3} \sum_{i=1}^{n} \lambda_{max}(E(X_i^2))\right)$$

$$= de^{-st}\exp\left(\frac{s^2/2}{1-|s|M/3}\left\|\sum_{i=1}^{n} A_i^2\right\|\right)$$

$$= de^{-st}\exp\left(\frac{\sigma^2 s^2/2}{1-|s|M/3}\right)$$

For $s = \frac{t/2}{\sigma^2+Rt}$, where $R = M/3$, the theorem holds for the above bound.

Thus for independent matrices X_i with dimension $d \times d$, $P(\left\|\sum_{i=1}^{n} X_i\right\| > t) \le 2d\exp(\frac{-t^2/2}{\sigma^2+Rt})$.

Using this result, the second part of the problem involves finding a bound for the weighted graph model as defined in the problem. Note the following:

$$\zeta_{ij} = \begin{cases} e_i e_j^\top + e_j e_i^\top & i \ne j \\ e_i e_i^\top & i = j \end{cases}$$

For $\zeta_{ij}, e_1, \ldots, e_n$ are the basic vectors of $\mathbb{R}^n$.

Note $A - P = \sum_{i \le j}(a_{ij} - p_{ij})\zeta_{ij}$. Let $X_{ij} = (a_{ij} - p_{ij})\zeta_{ij}$.

Then $P(\left\|A - P\right\| \ge t) \le 2n\exp\left(\frac{-t^2/2}{\sigma^2+Rt}\right)$.

$$\sigma^2 = \left\| E(X_{ij}^2) \right\|$$

$$= p_{ij}\zeta_{ij}^2$$

$$\sum_{i \leq j} E(X_{ij}^2) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}e_i e_i^\top$$

$$\sigma^2 = \max_i \sum_j p_{ij} = \Delta$$

$$P(\left\| A - P \right\| \geq t) \leq 2n\exp\left(\frac{-t^2/2}{\sigma^2 + Rt}\right)$$

$$= 2n\exp\left(\frac{-t^2}{2\sigma^2 + 2Rt}\right)$$

Let $c > 0$, assume $\eta \in [n^{-c}, 1/2]$. Then there exists C independent from n, P such that for $\Delta > C\log n$, $RT = 2\sqrt{\Delta\log(2n/\eta)} < \Delta$. Then,

$$P(\left\| A - P \right\| \geq t) \leq 2n\exp\left(\frac{-t^2}{2\Delta + 2\Delta}\right)$$

$$\leq 2n\exp\left(\frac{-t^2}{4\Delta}\right)$$

$$= 2n\exp(-\log(2n\eta))$$

$$= \eta$$

Thus η is a bound for $P(\left\| A - P \right\| \geq t)$.

Problem 3

Sketch a proof for the weak recovery of SBMs for the spectral clustering using the normalized Laplacian. In particular, consider the algorithm on page 1881 of Rohe et al.. Note that you do not need to use the eigenvectors associated with the largest k eigenvalues (in modulus), but rather the eigenvectors associated with the d largest eigenvalues where $d = \text{rk}(\mathbf{B})$ which is assumed known.

In particular, proceed as follows

- Relate the eigenvectors of $\mathcal{L}(\mathbf{P}) = \mathbf{T}^{-1/2}\mathbf{P}\mathbf{T}^{-1/2}$ to the block structures. Here $\mathbf{T}$ is a diagonal matrix with $T_{ii} = \sum_j p_{ij}$ is the **expected degree** of vertex i.

- Next invoke matrix concentration inequality for $\|\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P})\|$.

- Then invoke the Davis-Kahan theorem to bound the difference between the top eigenvectors of $\mathcal{L}(\mathbf{A})$ and the top eigenvectors of $\mathcal{L}(\mathbf{P})$.

- Finally, show that the above difference yield that the **proportion** of mis-clustered vertices (assuming clustering is done via K-means) will converges to 0.

Answer: We will **very** closely follow the "Implication for our spectral clustering story" from the notes which shows theorem 4 for weak recvover of SBMs.

Take $A \sim SBM(B, \tau)$. Then our normalized Laplacian is $\mathcal{L}(A) = D^{-1/2}AD^{-1/2}$. Here D is the diagonal matrix with $D_{ii} = \sum_j w_{ij}$ where w_{ij} is the edge weight between vertex i and vertex j. Notice that T is the diagonal matrix with th expected degrees as diagonal entries. That is, $E(D_{ii} = T_{ii})$.

Further take Z to be the $n \times K$ matrix such that $Z_{ik} = I(\tau(i) = k)$. Then we define $P = ZBZ^T$. Then, $\mathcal{L}(P) = T^{-1/2}PT^{-1/2}$.

We can order our eigenvalues similar to as we did in the notes. Let $|\widehat{\lambda}_1| \geq |\widehat{\lambda}_2| \geq \ldots$ be the eigenvalues of $\mathcal{L}(A)$ in decreasing modulus. Similarly take $|\lambda_1| \geq |\lambda_2| \geq \ldots$ to be the eigenvalues of $\mathcal{L}(P)$ in decreasing modulus. Here $\lambda_s(P) = 0$ for $s > d = \text{rank}(\mathcal{L}(P))$.

Recall that $\mathcal{L}(P) = T^{-1/2}PT^{-1/2}$. Since T is invertible, it must be full rank (and so must be $T^{-1/2}$). Thus, $\text{rank}(\mathcal{L}(P)) = \text{rank}(P) = d$.

Let us definte the sets $S_1 = \{\widehat{\lambda}_j : j \geq d+1\}$ and $S_2 = \{\lambda_j : 1 \leq j \leq d\}$. Now we can apply Weyl's inequality to say

$$\max_j |\lambda_j - \widehat{\lambda}_j| \leq \|\mathcal{L}(A) - \mathcal{L}(P)\| \Rightarrow |\widehat{\lambda}_h| \leq \|\mathcal{L}(A) - \mathcal{L}(P)\| \text{ for all } j \geq d+1.$$

Let's take $\widehat{U}$ and U to be the matrices whose columns are the top d eigenvectors of $\mathcal{L}(A)$ and $\mathcal{L}(P)$ respectively. Take $\lambda_* = \min\{|\lambda| : \lambda \in S_2\}$. Now suppose $\lambda_* > \|\mathcal{L}(A) - \mathcal{L}(P)\|$. We can apply the Davis-Kahan Theorem to say

$$\min_{W \in \mathcal{O}(p)} \|\widehat{U}W - U\|_F \leq \frac{\sqrt{2d}\|\mathcal{L}(A) - \mathcal{L}(P)\|}{\lambda_* - \|\mathcal{L}(A) - \mathcal{L}(P)\|}.$$

We will proceed by bounding λ_*.

Let's look a general eigenvector v of $\mathcal{L}(P)$ with eigenvalue λ. We want to relate these eigenvalues to eigenvalues of a function of the block structure, B.

$$\ell(P)v = \lambda v$$
$$T^{-1/2}PT^{-1/2}v = \lambda v$$
$$T^{-1/2}(ZBZ^T)T^{-1/2}v = \lambda v$$
$$BZT^{-1/2} \cdot T^{-1/2}(ZBZ^T)T^{-1/2}v = BZT^{-1/2} \cdot \lambda v$$
$$BZT^{-1}Z(BZ^T T^{-1/2}v) = \lambda(BZ^T T^{-1/2}v)$$

Thus, $\mathcal{L}(P)$ and $BZ^T T^{-1}Z$ share the same eigenvalues. Notice that this is similar to what we found in lecture, except for the T^{-1}. Remember that T is diagonal, so this gives us

$$BZ^T T^{-1}Z = B \times diag(\frac{n_1}{T_{11}}, \frac{n_2}{T_{22}}, \dots, \frac{n_k}{T_{kk}}).$$

Now assume $n_k = \Theta(n)$ as n increases. This restricts all blocks to grow at the same rate, n. Take $d_{min} = \min_i \sum_j p_{ij} = \min_i T_{ii}$. Then we have

$$\lambda_* = \Theta(n \cdot \lambda_{min}(B) \cdot d_{min}).$$

Where the $\min_i T_{ii}$ term comes from our additional diagonal terms. Then we can apply the Oliveira's concentraion of Laplacian matrices inequality (theorem 15) to say

$$P[\|\mathcal{L}(A) - \mathcal{L}(P)\|] > 14d_{min}^{-1/2}\sqrt{\log(4n/\eta)}.$$

Which implies

$$P[\|\mathcal{L}(A) - \mathcal{L}(P)\|] = \sqrt{\frac{\log(n)}{d_{min}}}.$$

Suppose now that $B = \rho_n \widetilde{B}$ where $\widetilde{B}$ is fixed in n with ρ_n possible converging to 0 as $n \to \infty$. With $\widetilde{B}$ fixed, B depends on n only through ρ_n. So we have the following with high probability.

$$\lambda_* = \Theta(nd_{min}^{-1}\rho_n) \text{ and } \|\mathcal{L}(A) - \mathcal{L}(P)\| = \Theta(\sqrt{nd_{min}^{-1}\rho_n})$$

Further, with high probability we have

$$\lambda_* > \|\mathcal{L}(A) - \mathcal{L}(P)\|.$$

We can now bound our Davis-Kahan inequality with high probability.

$$\min_{W \in \mathcal{O}(p)} \|\widehat{U}W - U\|_F \leq \frac{\sqrt{2d}\|\mathcal{L}(A) - \mathcal{L}(P)\|}{\lambda_* - \|\mathcal{L}(A) - \mathcal{L}(P)\|} = O\left(\sqrt{d \cdot d_{min}^{-1}}(n\rho_n)^{-1/2}\right)$$

Thus, there exists an orthogonal matrix W_* such that, on average

$$\|W_*\widehat{U}_i - U_i\| = O\left(\sqrt{d \cdot d_{min}^{-1}}(n\rho_n)^{-1/2}\right).$$

Note that clustering using $\widehat{U}$ is "equivalent" to clustering using $\widehat{U}W$ for any orthogonal W.

Let us now consider U, which contains K distinct rows. Since the columns of U are orthonormal, we have for all i with $Z_i \neq Z_j$

$$\|U_i - Uj\| = \Omega(n^{-1/2}).$$

Let S be the set of indices such that

$$\|W_* \widehat{U}_i - U_i\| > \min_{k:Z_k \neq Z_i} \|W_* \widehat{U}_i - U_k\|.$$

Then $i \in S$ only if $\|W_* \widehat{U}_i - U_i\| = \Omega(n^{-1/2})$.

We therefore have

$$\min W \in \mathcal{O}(p) \|\widehat{U}W - U\|_F^2 \geq \sum_{i \in S} \|W_* \widehat{U}_i - U_i\|^2 \tag{1}$$

$$\geq \sum_{i \in S} \Omega(n^{-1}) \tag{2}$$

$$= |S| \times \Omega(n^{-1}). \tag{3}$$

Finally, we can use the bound the we derived previously to say

$$\frac{|S|}{n} = O(dd_{min}^{-1}(n\rho)^{-1}) = o(1)$$

provided that $n\rho \to \infty$ as $n \to \infty$.