EE378B Inference, Estimation, and Information Processing

More on graph clustering and random matrices

Andrea Montanari Lecture 9-10 - Due on 2/13/2017

Paper submissions are preferred. You can submit the homework in class, or in the box of the second floor in Packard (there is a box for EE378B). You can use any programming language. Submit your code as part of the homework. Tables and figures are preferred to help to illustrate your results. There are several sub-questions for each question. Make sure that it is easy for readers to figure out which sub-questions you are answering.

This week there will be two options: a numerical/implementation option, and a theoretical option. You only need to solve one of the two problems!

Option 1: Graph clustering via relaxation

Given a graph $G = (V = [n], E)$, let $A_G = (A_{ij})_{i,j \leq n}$ be its adjacency matrix. We will try embed the graph in $r$ dimensions, by finding vectors $\sigma_1, \ldots, \sigma_n$ that solve

$$\max F(\sigma) = \sum_{i,j=1}^{n} A_{ij} \langle \sigma_i, \sigma_j \rangle - \frac{\lambda}{n} \left\| \sum_{i=1}^{n} \sigma_i \right\|_2^2,$$

subject to $\sigma_i \in \mathbb{R}^r$, $\|\sigma_i\|_2 = 1$, for all $i \in [n]$.

Equivalently, the objective function can be written as

$$F(\sigma) = \sum_{i,j=1}^{n} \left( A_{ij} - \frac{\lambda}{n} \right) \langle \sigma_i, \sigma_j \rangle.$$

Here is a concrete implementation of this idea, for the case of $k = 2$ clusters. (Below $S^{r-1} = \{ x \in \mathbb{R}^r : \|x\|_2 = 1 \}$ denotes the unit sphere.)

<table>
<thead>
<tr>
<th>RELAXATION-BASED CLUSTERING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong> : Graph $G = (V = [n], E)$, number of sweeps $T$</td>
</tr>
<tr>
<td><strong>Output</strong> : Vertex labels $\hat{\sigma} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_n)$, $\hat{\sigma}_i \in {1, \ldots, k}$</td>
</tr>
<tr>
<td>1: Initialize $\sigma$ uniformly at random on $S^{r-1}$;</td>
</tr>
<tr>
<td>2: For $t \in {1, \ldots, nT}$;</td>
</tr>
<tr>
<td>3: Draw $i(t) \sim \text{Unif}({n})$;</td>
</tr>
<tr>
<td>4: Let $\sigma_{i(t)}^{\text{new}} = \arg \max_{\sigma_{i(t)}} F(\sigma)$;</td>
</tr>
<tr>
<td>5: Set $\sigma_{i(t)} \leftarrow \sigma_{i(t)}^{\text{new}}$;</td>
</tr>
<tr>
<td>6: Regarding $\sigma \in \mathbb{R}^{r \times n}$ as a matrix, compute its top right singular vector $v_1 = v_1(\sigma) \in \mathbb{R}^n$;</td>
</tr>
<tr>
<td>7: Return vector $\hat{\sigma} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_n)$, where $\hat{\sigma}<em>i = \text{sign}(v</em>{1,i})$</td>
</tr>
</tbody>
</table>

A few remarks:

- The optimum at step 4 can be written in closed form. You should derive this expression and implement it.
- It is a good idea to store $A_G$ in sparse format.
The algorithm depends on 3 parameters:

1. The cost parameter \( \lambda \). A larger \( \lambda \) forces a more balanced embedding. Setting \( \lambda \gtrsim d \) (with \( d \) the average degree of \( G \)) should be sufficient.
2. The dimension \( r \). In practice \( r \in \{2, \ldots, 20 \} \) should work well.
3. The number of sweeps \( T \) (note that the number of iterations is \( nT \)). You can decide on this by monitoring the objective function, but \( T \approx 100 \) should be sufficient.

The homework consists in implementing the above algorithm and test it on the synthetic and real data. Namely:

(a) Generate 20 random graphs from the stochastic block model with \( n = 5,000 \), \( b = 10 \) and each of \( a \in \{10, 15, 20, 25, 30\} \). Apply Relaxation-based clustering and report the empirical overlap for each choice of the parameters \((n,a,b)\), averaged over the 20 samples.

(b) Apply the same algorithm to the polblogs dataset used for Homework 2. Compare these results with the ones of Homework 2.

Option 2: More on random matrices

This exercise aims at developing a more refined version of the \( \epsilon \)-net method to bound the operator norm of random matrices.

We say that a centered random variable \( X \) (with \( \mathbb{E} X = 0 \)) is \( b \)-sub-exponential if, for all \( \lambda \) with \( |\lambda| \leq 1/b \),
\[
\mathbb{E}\{e^{\lambda X}\} \leq e^{\lambda^2 b^2/2}.
\] (4)

There are other equivalent ways to define sub-exponential random variables. It is useful to recall following Bernstein inequality for sub-exponential random variables.

**Theorem 1** (Bernstein’s inequality). Let \((X_i)_{i \leq N}\) be a sequence of independent centered random variables, where \( X_i \) is \( b_i \)-sub-exponential, and define \( b = (b_1, \ldots, b_N) \). Then there exists a universal constant \( c_0 \) such that, for all \( t \geq 0 \),
\[
P \left\{ \left| \sum_{i=1}^{N} X_i \right| \geq t \right\} \leq 2 \exp \left\{ -c_0 \min \left( \frac{t}{\|b\|_{\infty}}, \frac{t^2}{\|b\|_2^2} \right) \right\}.
\] (5)

Let \( X \in \mathbb{R}^{d_1 \times d_2} \) be a random matrix with independent centered \( b \)-sub-exponential entries. We are interested in bounding the operator norm \( \|X\|_{\text{op}} \). Using the naive \( \epsilon \)-net method may not give a desirable bound.

To give a better bound on the operator norm \( \|X\|_{\text{op}} \), we will construct a special \( \epsilon \)-net as follows. For \( L \) an integer, define the set
\[
S_L = \left\{ 0, 1, \frac{1}{2}, \frac{1}{2^2}, \ldots, \frac{1}{2^L} \right\}.
\] (6)

We then define
\[
N^n(L) = \left\{ x \in \mathbb{R}^n : \|x\|_2 \leq 1, \ x_i^2 \in S_L \right\}.
\] (7)

We further define \( \pi_{<\epsilon} : N^n(L) \to N^n(L) \) and \( \pi_{=\epsilon} : N^n(L) \to N^n(L) \) by
\[
\pi_{<\epsilon}(x)_i = x_i 1_{x_i^2 > 2-\epsilon},
\] (8)
\[
\pi_{=\epsilon}(x)_i = x_i 1_{x_i^2 = 2-\epsilon}.
\] (9)

We also let \( N^n_{\geq \epsilon} = \pi_{=\epsilon}(N^n(L)), N^n_{<\epsilon} = \pi_{<\epsilon}(N^n(L)) \).
(a) Give an example of a random variable that is sub-exponential but not sub-Gaussian (and prove your claim).

(b) Prove that, if \( L = \log_2 n + c_0 \) for \( c_0 \) a suitable constant, \( N(L) \) is an \( \epsilon_0 \)-net of the unit ball \( B_n^2(0,1) \) for some \( \epsilon_0 < 1/2 \). As a consequence, for suitably chosen \( L \),

\[
\Pr(\|X\|_{op} \geq t) \leq \Pr\left( \max_{x \in N^{d_1}(L)} \max_{x \in N^{d_2}(L)} |\langle u, Xv \rangle| \geq C(\epsilon_0) t \right). \tag{10}
\]

(c) Prove that, for \( c_1 \) a suitable constant

\[
|N\subseteq \ell| \lor |N_{< \ell}| \leq \left( \frac{c_1 n}{2} \right)^{2^\ell}. \tag{11}
\]

(d) Prove that, for any \( t \geq 0, u \in B_2^{d_1}(0,1) \) and \( v \in B_2^{d_2}(0,1) \),

\[
\Pr\left( |\langle u, Xv \rangle| \geq t \right) \leq 2 \exp \left\{ -c_0 \min \left( \frac{t}{b \cdot \|u\|_{\infty} \|v\|_{\infty}}, \frac{t^2}{b^2} \right) \right\}. \tag{12}
\]

(e) Show that, for any matrix \( M \in \mathbb{R}^{d_1 \times d_2} \) and \( u \in N^{d_1}(L), v \in N^{d_2}(L) \)

\[
\langle u, Mv \rangle = \sum_{\ell=0}^{L} \langle \pi_{=\ell}(u), M\pi_{=\ell}(v) \rangle + \sum_{\ell=0}^{L} \langle \pi_{=\ell}(u), M\pi_{<\ell}(v) \rangle + \sum_{\ell=0}^{L} \langle \pi_{<\ell}(u), M\pi_{=\ell}(v) \rangle. \tag{13}
\]

(f) Use the above results to upper bound the following probabilities, for \( \ell \in \{0, \ldots, L\} \):

\[
\Pr\left( \max_{u \in N^{d_1}(L), v \in N^{d_2}(L)} \langle \pi_{=\ell}(u), X\pi_{=\ell}(v) \rangle \geq t_\ell \right), \tag{14}
\]

\[
\Pr\left( \max_{u \in N^{d_1}(L), v \in N^{d_2}(L)} \langle \pi_{<\ell}(u), X\pi_{=\ell}(v) \rangle \geq t_\ell \right), \tag{15}
\]

\[
\Pr\left( \max_{u \in N^{d_1}(L), v \in N^{d_2}(L)} \langle \pi_{=\ell}(u), X\pi_{<\ell}(v) \rangle \geq t_\ell \right). \tag{16}
\]

(g) Combine the above elements to obtain a bound on \( \|X\|_{op} \).

---

\(^1\)It might be useful to remember that \( \binom{n}{k} \leq \left( \frac{ne}{k} \right)^k \).