Recent advances in unsupervised learning: fundamental limits and efficient algorithms.

Lecture ①

Today

(i) Some examples.

(ii) The statistical questions.

(iii) Information theoretic lower bounds.

Example ① (Community detection)

Data: $G = ([n], E)$

Model: $\sigma = (\sigma_i)_{i=1}^n$, $\sigma_i \overset{iid}{\sim} \begin{cases} +1 & \text{wp } 1/2 \\ -1 & \text{wp } 1/2 \end{cases}$. Fix $p_n, q_n \in [0,1]$

(SBM) $\mathbb{P}[\{i,j\} \in E \mid \sigma] = \begin{cases} p_n & \text{if } \sigma_i = \sigma_j \\ q_n & \text{o.w.} \end{cases}$ If $p_n \geq q_n$ - assortative; $q_n > p_n$ - disassortative

Target: Recover the latent community structure σ from the graph G.

Comments:

(i) History - 80's — Holland, Laskey, Leinhardt. (social sciences)
90's — planted partition model (Computer science)

(ii) Challenging regime — $p_n = a/n$, $q_n = b/n$, $a \geq b$. (constant ind. of n)

Example ② ($\mathbb{Z}_2$-synchronization)

Data: $M \in \mathbb{R}^{n \times n}$, $M = M^T$.

Model: $\sigma = (\sigma_i) \in \{\pm 1\}^n$, $\sigma_i \overset{iid}{\sim} \begin{cases} +1 & \text{wp } 1/2 \\ -1 \end{cases}$. $M = \frac{\lambda}{n} \sigma \sigma^T + W$.

$W = (W_{ij}) \in \mathbb{R}^{n \times n}$, $W = W^T$, $W_{ij} \sim \begin{cases} N(0, 1/n) & \text{if } i<j \\ N(0, 2/n) & \text{if } i=j. \end{cases}$

Target: Recover σ from the matrix M.

Comments

(i) Normalization: $W_{ij} \sim \begin{cases} N(0, 1/n) & \text{if } i<j \\ N(0, 2/n) & \text{if } i=j \end{cases} \Rightarrow \|W\|_2 = O(1).$

$\left\| \frac{1}{n} \sigma\sigma^T \right\|_2 = 1.$ Normalization ensures that signal & noise are on the same level.

(ii) Motivation:

a) PCA — Recover a low rank signal under additive noise.

b) Group synchronization problem:

$\mathcal{G}$ - group (eg $SO(3)$) $\quad \sigma_i \in \mathcal{G}$.

Data: $\sigma_i^{-1}\sigma_j$ observed under noise.

Recover the latent rotations σ_i (upto a global rotation $o \in \mathcal{G}$)

(iii) Connections with community detection

A - adjacency matrix of SBM. $\quad G \sim G(n, a/n, b/n)$

$A = A - \mathbb{E}A + \mathbb{E}A \qquad \mathbb{E}A = \left[\begin{array}{c|c} a/n & b/n \\ \hline & a/n \end{array}\right] \qquad \mathbb{E}[A^{cen}_{ij}] = 0$

$A - \frac{d}{n} 11^T = \frac{a-b}{n}\sigma\sigma^T + A^{cen} \qquad A^{cen} = A - \mathbb{E}A \qquad \mathbb{E}\left[(A^{cen}_{ij})^2\right]$

$= \frac{a}{2n}\left(1-\frac{a}{n}\right) + \frac{b}{2n}\left(1-b/n\right) + o$

$$\frac{A - \frac{d}{n}11^T}{\sqrt{2(a+b)}} \overset{d}{\cong} \frac{\lambda}{n}\sigma\sigma^T + M^{cen} \qquad \lambda = \frac{a-b}{\sqrt{2(a+b)}}, \qquad d = \frac{a+b}{2}$$

$\cancel{\mathrm{Var}}\,\mathbb{E}(A_{ij} \mid \sigma_i\sigma_j = 1) = a/n$

$\rightarrow$ Similar to $\mathbb{Z}_2$-synchronization.

Example ③ (Community detection with side information)

Data: $G = ([n], E)$, $b_i \in \mathbb{R}^p$, $1 \le i \le n$.

Model: $\sigma = (\sigma_i)$, $\sigma_i \overset{iid}{\sim} \begin{cases} +1 & \text{wp } 1/2 \\ -1 \end{cases}$ $\mathbb{P}[A_{ij} = 1 | \sigma] = \begin{cases} a/n & \text{if } \sigma_i = \sigma_j \\ b/n & \text{o.w.} \end{cases}$

$u \in \mathbb{R}^p$. $\quad b_i = \sqrt{\frac{\mu}{n}} \sigma_i u + \frac{z_i}{\sqrt{p}}$ $\quad u \sim N(0, I_p/p)$, $z_i \sim N(0, I_p)$ ind.

$\mu > 0$

Asymptotic regime: $a, b = O(1)$, $p/n \to 1/\gamma \in (0, \infty)$.

Comments: (i) Additional covariates at each node. How to combine with graph info to perform signal recovery?

Statistical questions:

(i) **Detection**: Can we detect the presence of the signal?

(ii) **Recovery**: " " recover the hidden signal?

(iii) **Optimal recovery**: How to achieve "optimal" recovery?

[Addiotionally: can the above questions be solved 'efficiently' (eg in polynomial time)?]

(i) **Detection**. $\mathbb{Z}_2$-synchronization: $M = \frac{\lambda}{n} \sigma \sigma^T + W$.

$H_0: \lambda = 0$ vs. $H_1: \lambda > 0$.

Defn: (Consistent detection) We say that consistent detection is possible if $\exists \{T_n : n \ge 1\}$ s.t. $\mathbb{E}_{H_0}(T_n) + \mathbb{E}_{H_1}(1 - T_n) \to 0$ as $n \to \infty$.

Q. For which $\lambda > 0$ is consistent detection possible?

(ii) Recovery:

Q. Can we estimate the underlying signal?

Ex: Community detection $G \sim G(n, a/n, b/n)$

Fact: If $a, b = O(1)$, G contains $O(n)$ isolated vertices.
For isolated vertices, can only do random guessing!

Defn (Weak recovery) We say weak recovery is possible if
$\exists\, \hat{\sigma}$ s.t. $\frac{1}{n}|\langle \hat{\sigma}, \sigma \rangle| \geq \varepsilon$ whp.

(Intuition: Can we guess better than random?)

Comment:
(i) Other notions of recovery are possible eg exact recovery, almost exact recovery etc.

(iii) Optimal Recovery
All our examples are Bayesian problems. Given a loss function (eg squared loss), one studies the performance of the Bayes optimal estimator.
→ This is often computationally intractable!

Q. Can we design computationally efficient, Bayes optimal procedures?

Information theoretic lower bounds for detection.

Defn (Contiguity) Let P_n & Q_n be probability distributions on $(\mathcal{X}_n, \mathcal{F}_n)$. The sequence $\{Q_n : n \geq 1\}$ is contiguous to $\{P_n : n \geq 1\}$ if $P_n(E_n) \to 0$ implies $Q_n(E_n) \to 0$ $\forall \{E_n : n \geq 1\} \subseteq \mathcal{F}_n$.

NB: If $\{Q_n : n \geq 1\}$ is contiguous to $\{P_n : n \geq 1\}$ & $\{P_n : n \geq 1\}$ is contiguous to $\{Q_n : n \geq 1\}$, we say that $\{P_n : n \geq 1\}$ & $\{Q_n : n \geq 1\}$ are mutually contiguous.

Lemma If $\{Q_n : n \geq 1\}$ is contiguous to $\{P_n : n \geq 1\}$, then consistent detection is impossible.

Q. How to establish contiguity?

Lemma If $\mathbb{E}_{P_n}\left[\left(\frac{dQ_n}{dP_n}\right)^2\right]$ exists & remains bounded as $n \to \infty$, then $\{Q_n : n \geq 1\}$ is contiguous to $\{P_n : n \geq 1\}$.

Thm (Detection threshold for $\mathbb{Z}_2$-synchronization)
Let $M = \frac{\lambda}{n} \sigma\sigma^T + W$. $P_n = H_0$, $Q_n \sim P_\lambda$.
If $\lambda < 1$, then $\{Q_n : n \geq 1\}$ is contiguous to $\{P_n : n \geq 1\}$.

Proof:
$$\frac{dQ_n}{dP_n} = \frac{\sum_v \left(\frac{\sqrt{n}}{\sqrt{2\pi}}\right)^{\binom{n}{2}} e^{-\frac{n}{2}\sum_{i<j}(M_{ij} - \frac{\lambda}{n} v_i v_j)^2}}{\left(\frac{\sqrt{n}}{\sqrt{2\pi}}\right)^{\binom{n}{2}} e^{-\frac{n}{2}\sum_{i<j} M_{ij}^2}} = \frac{\sum_{v \in \{\pm 1\}^n} e^{-\frac{n}{2}\sum_{i<j} M_{ij}^2 - \frac{n}{2}\sum \frac{\lambda^2}{n^2} v_i^2 v_j^2 + \lambda \sum_{i<j} M_{ij} v_i v_j}}{e^{-n/2 \sum_{i<j} M_{ij}^2}}$$

$$\frac{dQ_n}{dP_n} = \sum_{v \in \{\pm 1\}^n} e^{\lambda \sum_{i<j} M_{ij} v_i v_j - \frac{\lambda^2}{2n} \frac{n(n-1)}{2}}$$

Thm $M = \frac{\lambda}{n}\sigma\sigma^T + W$.

If $\lambda < 1$, consistent detection is impossible.

Pf: Let $P_n \sim H_0$ & $Q_n \sim \mathbb{P}_\lambda$.

$$\frac{dQ_n}{dP_n} = \frac{1}{2^n}\sum_{v\in\{\pm 1\}^n} \frac{e^{-\frac{n}{4}\|M - \frac{\lambda}{n}vv^T\|_F^2}}{e^{-\frac{n}{4}\|M\|_F^2}}.$$

$$\|M - \frac{\lambda}{n}vv^T\|_F^2 = \|M\|_F^2 + \frac{\lambda^2}{n^2}\|vv^T\|_F^2 - \frac{2\lambda}{n}\langle M, vv^T\rangle$$

$$\|vv^T\|_F^2 = \langle vv^T, vv^T\rangle = n^2$$

$$\frac{dQ_n}{dP_n} = \mathbb{E}_v\left[e^{\frac{\lambda}{2}\langle M, vv^T\rangle - \frac{\lambda^2 n}{4}}\right]$$

$$\left(\frac{dQ_n}{dP_n}\right)^2 = \mathbb{E}_{v^1, v^2}\left[e^{\frac{\lambda}{2}\langle M,\ v^1(v^1)^T + v^2(v^2)^T\rangle - \frac{\lambda^2 n}{2}}\right]$$

$$\mathbb{E}_{P_n}\left[\left(\frac{dQ_n}{dP_n}\right)^2\right] = \mathbb{E}_{v^1, v^2}\left[e^{-\frac{\lambda^2 n}{2}}\ \mathbb{E}_{P_n}\left[e^{\frac{\lambda}{2}\langle M,\ v^1(v^1)^T + v^2(v^2)^T\rangle}\right]\right]$$

$$\mathbb{E}_{P_n}\left[e^{c\langle M, X\rangle}\right] = \mathbb{E}_{P_n}\left[e^{2c\sum_{i<j} M_{ij}X_{ij} + c\sum_{i=1}^{n} M_{ii}X_{ii}}\right]$$

$$= e^{\frac{2c^2}{n}\sum_{i<j} X_{ij}^2 + \frac{c^2}{n}\sum_i X_{ii}^2} = e^{c^2/n \|X\|_F^2}.$$

$$\mathbb{E}_{P_n}\left[e^{\frac{\lambda}{2}\langle M,\, v^1(v^1)^T + v^2(v^2)^T\rangle}\right] = e^{\frac{\lambda^2}{4n}\| v^1(v^1)^T + v^2(v^2)^T\|_F^2} = \star$$

$$\| v^1(v^1)^T + v^2(v^2)^T \|_F^2 = 2n^2 + 2(\langle v^1, v^2\rangle)^2.$$

$$\star = e^{\frac{\lambda^2 n}{2} + \frac{\lambda^2}{2n}\langle v^1, v^2\rangle^2}$$

$$\therefore \quad \mathbb{E}_{P_n}\left[\left(\frac{dQ_n}{dP_n}\right)^2\right] = \mathbb{E}_{v^1, v^2}\left[e^{\frac{\lambda^2 n}{2}\left(\frac{1}{n}\langle v^1, v^2\rangle\right)^2}\right].$$

<u>To prove</u>: If $\lambda < 1$,

$$\limsup_{n\to\infty} \mathbb{E}_{v^1, v^2}\left[e^{\frac{\lambda^2 n}{2}\left(\frac{1}{n}\langle v^1, v^2\rangle\right)^2}\right] \le M < \infty$$

for some M.