STSCI 6940 : Topics in high-dimensional inference.

Lecture 1 Date : 02/09

Interested in signal recovery and detection in high-dimensional regime / high noise regime.

signal : low-dimensional structure. (signal sparse, low-rank, ---).

Two points of focus :

Information-theoretic limits ! (when is est / det possible at all?).
Algorithmic aspect :
Can we do this efficiently up to the IT limit?

General situation : $x \in \mathbb{R}^n$

$$x \xrightarrow{\text{noisy channel}} y \in \mathbb{R}^n$$

observe y and then try to recover x.

A blend of tools : information theory, statistical physics, probability th. computer science (algorithms).

- A toy model : rank-one spiked Wigner model

(dense model) $$\underset{\substack{\uparrow \\ \text{rank-one signal} \\ \text{(spiked)}}}{Y = \sqrt{\frac{\lambda}{n}} \cdot x x^T} + \underset{\substack{\uparrow\, \mathbb{R}^{n\times n} \\ \text{noise.}}}{W}$$

clean theory that emerged in the last decade, heavily inspired by spin glass theory.

(sparse model) stochastic block model / reconstruction on trees.

reconstruction on general graphs i.e. beyond tree-like graphs.

information - percolation bounds.

- Algorithmic approaches:
  1. Spectral methods (RMT).
  2. Approximate message passing.
  3. variational inference.
  4. MCMC methods for sampling
     Glauber dynamics / Langevin dynamics.

- "Lower bounds"/ evidence of hardness:
  - low degree likelihood ratio method.
  - overlap gap phenomenon.

- For projects: List in course's webpage.

- Plan: 27 meetings.

  - 13 meetings will be devoted to an exposition of the foundations.
  - second half: presentations by students.

pick one or two papers from the list
by Friday 19.

Evaluation: 1. Homework (30%).
2. Presentation (30%).
3. written report (40%).

office hours wednesdays 3-5 pm.
Slack channel.

- Material: Léo Miolane's PhD thesis.
  Talagrand's book (spin glasses).
  Research papers.

Lecture 1:

The additive Gaussian channel:

$x \in \mathbb{R}^n$ $\quad x \sim P_0$ finite second moment.

$$y = \sqrt{\lambda} \cdot x + z \qquad , z \sim N(0, I) \text{ indep of } x.$$

setting where model is known, λ is known, P_0 is known.

- $\hat{x} : \mathbb{R}^n \to \mathbb{R}^n$

Mean squared error:

$$MSE(\hat{x}) := \mathbb{E}\left[\| \hat{x}(y) - x \|_2^2\right].$$

Thm: Best estimator is $\hat{x}(y) = \mathbb{E}[x \mid y]$.

its error is

$$MMSE(\lambda, P_0) = \mathbb{E}\left[\| \mathbb{E}[x \mid y] - x \|_2^2\right].$$

Take $\hat{x}$ any estimator

Pf: $\mathbb{E}[\|\hat{x} - x\|_2^2] = \mathbb{E}[\|\hat{x} - \mathbb{E}[x|y] + \mathbb{E}[x|y] - x\|_2^2]$

$$= \mathbb{E}[\|\hat{x} - \mathbb{E}[x|y]\|_2^2] + \mathbb{E}[\|\mathbb{E}[x|y] - x\|_2^2]$$

$$+ 2\,\underbrace{\mathbb{E}[(\hat{x}(y) - \mathbb{E}[x|y])(\mathbb{E}[x|y] - x)]}$$

$$= \mathbb{E}[\mathbb{E}[(\qquad\qquad) | y]]$$

$$= 0$$

$$MSE(\hat{x}) \geq MSE(\mathbb{E}[x|y]).$$

$\Rightarrow$ $\hat{x} = \mathbb{E}[x|y]$ minimizes the MSE.

interesting object : $\mathbb{P}(\cdot | y)$.

Bayes rule:

$$d\mathbb{P}(x|y) = \frac{f(y|x) \cdot d\mathbb{P}_0(x)}{\tilde{Z}(y,\lambda)}.$$

$f(y|x)$: density of y given x.

$$y = \sqrt{\lambda} \cdot x + z.$$

$$y | x \sim N(\sqrt{\lambda}\, x, I).$$

$$f(y|x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\|y - \sqrt{\lambda}\, x\|_2^2}.$$

$$= \frac{1}{\sqrt{2\pi}} e^{\sqrt{\lambda}\, y^T x - \frac{\lambda}{2}\|x\|_2^2 - \frac{\lambda}{2}\|y\|_2^2}.$$

$$d\mathbb{P}(x|y) = \frac{1}{\sqrt{2\pi}} \frac{e^{\sqrt{\lambda}\, y^T.x - \frac{\lambda}{2}\|x\|_2^2 - \frac{\lambda}{2}\|y\|_2^2}}{\tilde{Z}(y,\lambda)} dP_0(x)$$

$\uparrow$
Gibbs / Boltzmann dist

$$= \frac{e^{\sqrt{\lambda}\, y^T . x - \frac{\lambda}{2}\|x\|_2^2}}{Z(y,\lambda)} dP_0(x)$$

$$Z(y,\lambda) = \int e^{\sqrt{\lambda}\, y^T . x - \frac{\lambda}{2}\|x\|_2^2} dP_0(x).$$

$\uparrow$
partition function.

$$H(x) = \sqrt{\lambda}\, y^T . x - \frac{\lambda}{2}\|x\|_2^2$$: Hamiltonian.
(log-likelihood fct)

- $\mathbb{E}[x|y]$.

- More generally : $f : \mathbb{R}^n \to \mathbb{R}$

$$\langle f(x) \rangle_\lambda = \mathbb{E}[f(x) | y].$$ Gibbs average.

- $F(\lambda) = \mathbb{E} \log Z(y;\lambda)$: Free energy.

- Mutual information : (x,y) jointly distributed.

$$I(y;x) = I(x;y) = D_{KL}(P_{x \times y} \| P_x \otimes P_y).$$

$$D_{KL}(\mathbb{P}, \mathbb{Q}) = \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{Q}}\right].$$

- small computation ;

$$I(x;y) = \frac{\lambda}{2}\mathbb{E}[\|x\|^2] - F(\lambda).$$

$$I(x;y) = D_{KL}\left(P_{x,y} \,\|\, P_x \otimes P_y \right).$$

$$= \mathbb{E}_{P_{x,y}}\left[\log \frac{dP_{x,y}}{d(P_x \otimes P_y)} \right].$$

$$= \mathbb{E}_{P_{x,y}}\left[\log \frac{dP_{x|y} \cdot \cancel{dP_y}}{dP_x \cdot \cancel{dP_y}} \right].$$

$$= \mathbb{E}_{P_{x,y}}\left[\log \frac{e^{\sqrt{\lambda}\, y^T . x - \frac{\lambda}{2}\|x\|_2^2}}{\mathcal{Z}(y;x)} \right].$$

$$= \mathbb{E}\left[\sqrt{\lambda}\, y^T . x - \frac{\lambda}{2}\|x\|_2^2 \right] - \underbrace{\mathbb{E} \log \mathcal{Z}(y,\lambda)}_{F(\lambda)}.$$

$$\sqrt{\lambda}\,\mathbb{E}[y^T . x] = \lambda\, \mathbb{E}[\|x\|_2^2] + \sqrt{\lambda}\, \underbrace{\mathbb{E}[z^T . x]}_{=0}$$

$(y = \sqrt{\lambda}\, x + z)$

$$I(y;x) = \frac{\lambda}{2}\mathbb{E}[\|x\|_2^2] - F(\lambda).$$

$\blacksquare$

- I – MMSE relation:

$$MMSE_{P_0}(\lambda) = \mathbb{E}\left[\| x - \mathbb{E}[x|y] \|_2^2\right].$$

For all $\lambda \geq 0$: $\boxed{\frac{d}{d\lambda} I(y;x) = \frac{1}{2} MMSE_{P_0}(\lambda)}$.

Equivalently: $F'(\lambda) = \frac{1}{2}\mathbb{E}\left[\|\mathbb{E}[x|y]\|_2^2\right]$.

Proof: Lemma (Gaussian integration by parts / Stein's lemma).

$z \sim N(0,1)$. $f : \mathbb{R} \to \mathbb{R}$ differentiable, grows at most polynomially for $|x| \to +\infty$

$$|f(x)| \leq c|x|^k \quad \text{for some } k, c.$$

$$\mathbb{E}[z f(z)] = \mathbb{E}[f'(z)].$$

- $$F(\lambda) = \mathbb{E} \log \int e^{\sqrt{\lambda}\, y^T \cdot x - \frac{\lambda}{2}\|x\|_2^2} \, dP_0(x).$$

$$y = \sqrt{\lambda}\, x_0 + z.$$

$$= \mathbb{E} \log \int e^{\sqrt{\lambda}\, z^T \cdot x + \lambda x_0^T \cdot x - \frac{\lambda}{2}\|x\|_2^2} \, dP_0(x).$$

$$F'(\lambda) = \mathbb{E}\left[\frac{1}{Z(y,\lambda)} \int \left(\frac{1}{2\sqrt{\lambda}} z^T \cdot x + x_0^T \cdot x - \frac{1}{2}\|x\|_2^2 \right) e^{H(x)} dP_0(x) \right]$$

First term:

$$\frac{1}{2\sqrt{\lambda}} \mathbb{E}\left[\frac{1}{Z(y,\lambda)} \cdot z^T \cdot \int x\, e^{H(x)} dP_0(x) \right].$$

$$= \frac{1}{2\sqrt{\lambda}} \mathbb{E}\left[\sum_{i=1}^{n} \frac{d}{dz_i} \left(\frac{1}{Z(y,\lambda)} \int x\, e^{H(x)} dP_0(x) \right) \right].$$

$$= \frac{1}{2\sqrt{\lambda}} \mathbb{E}\Bigg[\sqrt{\lambda} \int \|x\|^2 e^{H(x)} dP_0(x) \frac{1}{Z(y,\lambda)}$$

$$- \sqrt{\lambda} \frac{\left(\int x\, e^{H(x)} dP_0(x) \right)^T \left(\int x\, e^{H(x)} dP_0(x) \right)}{Z(y,\lambda)^2} \Bigg]$$

$$= \frac{1}{2} \underbrace{\mathbb{E}\left[\mathbb{E}[\|x\|^2 | y] \right]} - \frac{1}{2} \mathbb{E}\left[\|\mathbb{E}[x|y]\|^2 \right].$$

$$= \frac{1}{2} \mathbb{E}\left[\|x\|^2 \right] - \frac{1}{2} \mathbb{E}\left[\|\mathbb{E}[x|y]\|^2 \right].$$

$$F'(\lambda) = \frac{1}{2}\mathbb{E}[\|x\|^2] - \frac{1}{2}\mathbb{E}[\|\mathbb{E}[x|y]\|^2]$$

$$+ \mathbb{E}[x_0^T \cdot \mathbb{E}[x|y]] - \frac{1}{2}\mathbb{E}[\|x\|^2]$$

$$\mathbb{E}[x_0^T \cdot \mathbb{E}[x|y]] = \mathbb{E}[\mathbb{E}[x_0^T \cdot \mathbb{E}[x|y] \,|\, y]]$$

$$= \mathbb{E}[\mathbb{E}[x|y]^T \cdot \mathbb{E}[x|y]]$$

$$= \mathbb{E}[\|\mathbb{E}[x|y]\|_2^2].$$

$$F'(\lambda) = \frac{1}{2}\mathbb{E}[\|\mathbb{E}[x|y]\|_2^2].$$

$\blacksquare$