Fano's $\neq$, lower bounds for compressed sensing

Note Title 2015-12-02

Sparse linear model: $y = \underset{\substack{\uparrow \\ \text{sub-Gauss}}}{A}x + Z\cdot\sigma$ $\nwarrow N(0,I)$, $x \in K_s := \{x : \|x\|_0 \le s\}$

Recall: Let $\hat{x}_{\ell_1}$ be the soln to ℓ_1-min. as in last lecture. We showed that

$$\sup_{x\in K_s} P\left(\|\hat{x}_{\ell_1} - x\|_2^2 > \frac{C s \log(\frac{n}{s})\sigma^2}{m}\right) \le .02$$

provided $m \geq C s \log(\frac{n}{s})$

Today we show that no estimator can do much better. We use:

Thm (Synthesized Fano's $\neq$)

$$\boxed{\begin{array}{l}\text{Let } y = x + Z\cdot\sigma \quad x \in K \subseteq \mathbb{R}^n, \quad z \sim N(0,I). \\ \text{Suppose } \sqrt{\log P\left(K \cap \delta B_2^n, \frac{\delta}{10}B_2^n\right)} \geq C\cdot\frac{\delta}{\sigma} \\ \text{some } \delta > 0. \quad \text{Then} \\ \qquad \inf_{\hat{x}=\hat{x}(y)} \sup_{x\in K} P\left(\|\hat{x} - x\|_2 > \frac{\delta}{20}\right) \geq 0.5\end{array}}$$

Remark If K is a cone, then

$$P\left(K \cap \delta B_2^n, \frac{\delta}{10}B_2^n\right) = P\left(K \cap B_2, \frac{1}{10}B_2\right)$$

"If you can pack a bunch of points in a ball in K, then you can't distinguish one point from the heap."

Corollary: Suppose $y = Ax + Z\sigma$, K is a cone, and $\frac{A}{\sqrt{m}}$ is a near isometry on $K-K$, that is,

$$\frac{1}{2}\|x\|_2 \le \|Ax\|_2 \le 2\|x\|_2 \qquad \forall\, x \in K-K.$$

Suppose $z \sim N(0,I)\,\sigma$.

Let $\delta := \sqrt{\log P(K \cap B_2, c B_2)}$

Then

$$\inf_{\hat{x}(y)} \sup_{x \in K} P\left(\|\hat{x}(y) - x\| > \frac{c\,\delta\sigma}{\sqrt{m}}\right) > \frac{1}{2}$$

Proof: $\boxed{\text{Exercize}}$

We can use this to give minimax error for sparse linear model:

As shown in homework, $\sqrt{\log P(K_s \cap B_2, \frac{1}{10} B_2)} \geq c\sqrt{s \log(\frac{n}{s})}$.

$$\Rightarrow \boxed{\inf_{\hat{x}(y)} \sup_{x \in K_s} P\left(\|\hat{x}(y) - x\|^2 \geq c\,\frac{s \log(\frac{n}{s})}{m}\,\sigma^2\right) \geq \frac{1}{2}}$$

if A is a near isometry, which occurs w.p. ≥ 0.99 if $m \geq C s \log(\frac{n}{s})$.

Proof of synthesized Fano's $\neq$

Assume $\sigma = 1$ for simplicity.

Possible step 1: Give K a prior, i.e. let x be randomly distributed, supported on K.

Then for any estimator $\hat{x}(y)$, $\delta > 0$,

$$\sup_{x \in K} P(\|\hat{x}(y) - x\|_2 > \delta) \leq \underbrace{P_{x,y}}(\|\hat{x}(y) - x\|_2 > \delta)$$

probability taken over x & y.

Our step 1: Let $\mathcal{X}$ be a $\frac{\delta}{10}$-packing of $K \cap \delta B_2^n$.

Put a uniform prior on $\mathcal{X}$.

Then,

$$\inf_{\hat{x}} \sup_{x} P\left(\|\hat{x}(y) - x\|_2 \geq \frac{\delta}{20}\right) \geq \inf_{\hat{x}} \sup_{x \in \mathcal{X}} P(\hat{x}(y) \neq X)$$

$$\geq \inf P_{x,y} P(\hat{x}(y) \neq X)$$

$$= P_{XY}(\hat{x}_{ml}(y) \neq X)$$

where

$$\hat{x}_{ml} = \text{maximum likelihood estimator}$$
$$= \operatorname*{argmax}_{\bar{x} \in \mathcal{X}} \ell(y \mid \bar{x})$$

where $\ell(y \mid \bar{x}) := \frac{1}{(\sqrt{2\pi})^2} e^{-\frac{1}{2}\|y - x\|_2^2}$

$$\Rightarrow \hat{x}_{ml}(y) = \operatorname*{argmin}_{\bar{x} \in \mathcal{X}} \|\bar{x} - y\|_2$$

i.e. $P(\hat{x}_{ml}(Y) = X) = P(X \text{ is closest to } y)$

$\{X \text{ is closest to } y\}$

$$\left.\begin{aligned} &\Leftrightarrow \|X - y\|_2^2 \leq \|\bar{x} - y\|_2^2 \\ &\Leftrightarrow \|Z\|_2^2 \leq \|\bar{x} - X - Z\|_2^2 \\ &\Leftrightarrow \langle \bar{x} - X, Z \rangle \leq \tfrac{1}{2}\|\bar{x} - X\|_2^2 \end{aligned}\right\} \quad \forall \bar{x} \in \mathcal{X}$$

The RHS $\leq 2\delta^2$

Thus $P(\hat{x} \neq X) \geq P\left(\sup_{\bar{x} \in \mathcal{X}} \langle \bar{x} - X, Z \rangle \geq 2\delta^2\right)$ (*)

Condition on X. Then

$\sup_{\bar{x} \in \mathcal{X}} \langle \bar{x} - X, Z \rangle$ is a Gaussian process

Note $\mathbb{E} \sup_{\bar{x} \in \mathcal{X}} \langle \bar{x} - X, Z \rangle = w(\mathcal{X}) \geq c\delta\sqrt{\log|\mathcal{X}|}$ (1)

by Sudakov $\neq$.

Further, by Borel-TIS ≠, the process concentrates around its mean, that is

$$\Big\| \sup_{\bar{x}\in\mathcal{X}} \langle x-\bar{x}, z\rangle - w(\mathcal{X}) \Big\|_{\psi_2} \le C\cdot \max_{\bar{x}\in\mathcal{X}} \| \langle x-\bar{x}, z\rangle \|_{\psi_2}$$

$$\le C \max_{\bar{x}\in\mathcal{X}} \|x-\bar{x}\|_2$$

$$\le C\cdot\delta \qquad (2)$$

Combine (1) and (2) to give

$$\sup_{\bar{x}\in\mathcal{X}} \langle x-\bar{x}, z\rangle \underset{\text{w.p.} \geq \frac{1}{2}}{\geq} \frac{1}{2} w(\mathcal{X}) \geq C\,\delta\sqrt{\log|\mathcal{X}|}$$

(provided $|\mathcal{X}| > c$)

Thus, by (*), we see that

$P(\hat{x} \neq x) \geq \frac{1}{2}$ provided $\sqrt{\log|\mathcal{X}|} \geq C\cdot\delta$.

∎