Restricted Strong Convexity of LASSO (cont'd)

2022年2月15日 9:16

Recall that $X_{ij} \sim \mathcal{N}(0,1)$, $V(r) = \{ \Delta \in \mathbb{R}^d : \|\Delta\|_2 = 1, \|\Delta\|_1 \leq r \}$.

<u>Proposition 1</u>:

$$\mathbb{E}\left[\inf_{\Delta \in V(r)} \frac{\|X\Delta\|_2}{\sqrt{n}} \right] \geq 3\left[\frac{1}{4} - \sqrt{\frac{\log d}{n}}\, r \right]$$

whenever $V(r) \neq \phi$.

<u>Proof</u> (cont'd) Using Gordon's comparison inequality, we showed

$$\mathbb{E}\Big[- \underbrace{\tilde{Q}(r,X)}_{= \inf_{\Delta \in V(r)} \|X\Delta\|_2} \Big] = \mathbb{E}\Big[\sup_{\Delta \in V(r)} \inf_{u \in S^{n-1}} \underbrace{u^T X \Delta}_{= Y_{u,\Delta}} \Big] \leq \underbrace{-\mathbb{E}[\|g_1\|_2]}_{\leq -\frac{3}{4}\sqrt{n}} + \underbrace{r \cdot \mathbb{E}[\|g_2\|_\infty]}_{\leq 3r\sqrt{\log d}},$$

where $g_1 \sim \mathcal{N}(0, I_n)$ and $g_2 \sim \mathcal{N}(0, I_d)$.

<u>1°</u>: $\mathbb{E}[\|g_1\|_2^2] = \mathbb{E}\left[\sum_{i=1}^n (g_1)_i^2 \right] = n$, so we expect

$$\mathbb{E}[\|g_1\|_2] \approx \sqrt{n}$$

(Exercise) Show that $\mathbb{E}[\|g_1\|_2] \geq \frac{3}{4}\sqrt{n}$ for, say, $n \geq 10$.

<u>2°</u>: To estimate $\mathbb{E}[\|g_2\|_\infty]$, we use

$$\mathbb{E}[\|g_2\|_\infty] = \int_0^\infty \Pr(\|g_2\|_\infty > t)\, dt$$

Here,

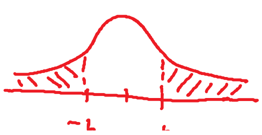
$$\Pr(\underbrace{\|g_2\|_\infty}_{= \max_i |(g_2)_i|} > t) = \Pr\left[|(g_2)_1| > t \text{ or } |(g_2)_2| > t \text{ or } \cdots \text{ or } |(g_2)_d| > t \right]$$

$$\leq \sum_{i=1}^d \Pr(|(g_2)_i| > t) \quad \text{(Union bound)}$$

$$= d \cdot \Pr(|g| > t), \quad g \sim \mathcal{N}(0,1)$$

($(g_2)_i$ are iid)

(Exercise) $\Pr(|g| > t) \leq \sqrt{\frac{2}{\pi}} \frac{1}{t} \exp(-t^2/2) \quad \forall t > 0$



Putting these together, $\mathbb{E}\left[\|g_2\|_\infty\right] \le 3\sqrt{\log d}$

Proposition 2: Let $r>0$ be s.t. $V(r) \neq \emptyset$. Define

$t(r) = \frac{1}{4} + 3\sqrt{\frac{\log d}{n}}\, r$. Then,

$$\Pr\left[\; \left| \underbrace{Q(r,X)}_{= \inf\limits_{\Delta\in V(r)} \frac{\|X\Delta\|_2}{\sqrt{n}}} - \mathbb{E}\left[\underline{Q(r,X)} \right] \right| \geq \frac{1}{2} t(r) \right] \le 2\exp\left(-n t^2(r)/8\right)$$

Fact (Concentration of measure for Lipschitz functions of Gaussians)

Let $g \sim \mathcal{N}(0, I_m)$ and $F: \mathbb{R}^m \to \mathbb{R}$ be Lipschitz with parameter $L>0$:

$$|F(u) - F(v)| \le L\|u-v\|_2 \quad \forall u,v$$

Then, for all $t>0$,

$$\Pr\left[\,|F(g) - \mathbb{E}F(g)| \geq t\right] \le 2\exp\left(-t^2/2L^2\right).$$

Proof: Our goal is to show that $X \mapsto Q(r,X)$ is Lipschitz.

$$\sqrt{n}\,\left|Q(r,X) - Q(r,Y)\right| = \left| \inf_{\Delta\in V(r)} \|X\Delta\|_2 - \inf_{\Delta\in V(r)} \|Y\Delta\|_2 \right|$$

$$= \left| \inf_{\Delta\in V(r)} \|X\Delta\|_2 - \|Y\hat{\Delta}\|_2 \right|$$

where $\hat{\Delta}$ is an optimal solution to $\inf\limits_{\Delta\in V(r)} \|Y\Delta\|_2$

$$\le \left| \|X\hat{\Delta}\|_2 - \|Y\hat{\Delta}\|_2 \right|$$

$$\le \sup_{\Delta\in V(r)} \|(X-Y)\Delta\|_2$$

$$\le \underbrace{\|X-Y\|}_{\text{Spectral norm of } X-Y} \cdot \underbrace{\sup_{\Delta\in V(r)} \|\Delta\|_2}_{=1} \le \|X-Y\|_F$$

It follows that $Q(r,\cdot)$ is $\frac{1}{\sqrt{n}}$-Lipschitz

It follows that $Q(r, \cdot)$ is $\frac{1}{\sqrt{n}}$-Lipschitz.

Hence, with $t(r) = \frac{1}{4} + 3\sqrt{\frac{\log d}{n}}\, r$

$$\Pr\left[\left| Q(r,X) - \underline{\mathbb{E}[Q(r,X)]} \right| \geq \frac{1}{2} t(r) \right]$$

$$\leq 2 \exp\left(-n t^2(r)/8\right). \qquad \geq 3\left[\frac{1}{4} - \sqrt{\frac{\log d}{n}}\, r\right]$$

Step 3: For a fixed $r > 0$, we showed that with high probability, $\left| Q(r,X) - \mathbb{E}[Q(r,X)] \right| \leq \frac{1}{2} t(r)$, which implies

$$Q(r,X) = \inf_{\Delta \in V(r)} \frac{\|X\Delta\|_2}{\sqrt{n}} \geq \mathbb{E}[Q(r,X)] - \frac{1}{2} t(r)$$

$$\geq \frac{3}{4} - 3\sqrt{\frac{\log d}{n}}\, r - \frac{1}{8} - \frac{3}{2}\sqrt{\frac{\log d}{n}}\, r$$

$$= \frac{5}{8} - \frac{9}{2}\sqrt{\frac{\log d}{n}}\, r = 1 - \frac{3}{2} t(r)$$

In other words, with high probability,

$$\frac{\|X\Delta\|_2}{\sqrt{n}} \geq 1 - \frac{3}{2} t(r) \qquad \forall \Delta \in V(r) \qquad \text{— } (*)$$

We need the above to hold uniformly in r. Observe

$$(*) \iff \sup_{\Delta \in V(r)} \left(1 - \frac{\|X\Delta\|_2}{\sqrt{n}}\right) \leq \frac{3}{2} t(r)$$

Now, consider the following result:

Fact: (Peeling argument)

Let $A \subseteq \mathbb{R}^d$ be a non-empty set, $f : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$, $h : \mathbb{R}^d \to \mathbb{R}_+$ be given, and $g : \mathbb{R} \to \mathbb{R}$ be non-negative and strictly increasing. Consider

$$\sup_{\substack{h(v) \leq r \\ v \in A}} f(v, X),$$

where X is a random vector. Suppose that $g(r) \geq \mu$ for all $r \geq 0$ and $\exists c > 0$ s.t. for all $r > 0$,

$$\Pr\left[\sup_{\substack{h(v) \leq r \\ v \in A}} f(v, X) \geq g(r)\right] \leq 2\exp\left(-c \cdot a\, g^2(r)\right)$$

for some $a > 0$. Then,

$$\Pr[\mathcal{E}] \leq \frac{2\exp(-4c\, a \cdot \mu^2)}{1 - \exp(-4c \cdot a \cdot \mu^2)}, \qquad \leftarrow \text{uniform in } r!$$

where $\mathcal{E} = \{ X : \exists v \in A \text{ s.t. } f(v, X) \geq 2g(h(v)) \}$.

Let us apply the fact to our setting.

$V(r) = \{\Delta : \|\Delta\|_2 = 1, \|\Delta\|_1 \leq r\}$

$= A \cap \{\Delta : \|\Delta\|_1 \leq r\}$

$$A = S^{d-1} = \{\Delta \in \mathbb{R}^d : \|\Delta\|_2 = 1\}$$

$$h(\Delta) = \|\Delta\|_1, \quad f(\Delta, X) = 1 - \frac{\|X\Delta\|_2}{\sqrt{n}},$$

$$g(r) = \frac{3}{2} t(r) = \frac{3}{2}\Big(\underbrace{\frac{1}{4} + 3\sqrt{\frac{\log d}{n}}\, r}_{t(r)}\Big)$$

Note that $g(r) \geq g(0) = \frac{3}{8} = \mu$. Moreover,

$$\Pr\left[\sup_{\substack{h(\Delta) \leq r \\ \Delta \in A}} f(\Delta, X) \geq g(r)\right] \overset{\text{step 2}}{\leq} 2\exp\left(-n t^2(r)/8\right)$$

$$= 2\exp\left(-n g^2(r)/18\right)$$

It follows that $\exists c_1, c_2 > 0$ s.t.

$$\Pr[\mathcal{E}] \leq c_1 \exp(-c_2 n),$$

where $\mathcal{E} = \left\{ \exists \Delta \in S^{d-1} : 1 - \frac{\|X\Delta\|_2}{\sqrt{n}} \geq 3t(\|\Delta\|_1) \right\}$

$$= \left\{ \exists \Delta \in S^{d-1} : \frac{\|X\Delta\|_2}{\sqrt{n}} \leq \frac{1}{4} - 9\sqrt{\frac{\log d}{n}}\, \|\Delta\|_1 \right\}$$