STAT 499: Undergraduate Research

Week 5: Concentration and the Gaussian tail example

Facheng Yu Compiled on: 2024-02-05, 06:04:25

Last week we learn about the consistency theorem of Lagrangian Lasso:

Theorem 5.1 *Assume that θ^* has its support on S and $\mathbf{X}$ satisfies the RE condition. for any solution $\widehat{\theta}$ of the Lagrangian Lasso with $\lambda_n \geq 2\|\frac{\mathbf{X}^T w}{n}\|_\infty$, we have*

$$\|\widehat{\theta} - \theta^*\|_2^2 \leq \frac{3}{k}\sqrt{s}\lambda_n.$$

Gaussian variable. A random variable X with mean μ and variance σ^2 is said to be Gaussian if its density f satisfies

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\},$$

and we denote X as

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Moment generating function. The moment generating function (MGF) of X is defined as

$$M_X(t) = \mathbb{E}[e^{tX}].$$

For Gaussian variable $X \sim \mathcal{N}(\mu, \sigma^2)$, its MGF is

$$M_X(t) = \mathbb{E}[e^{tX}] = e^{\mu t + \frac{\sigma^2}{2}t^2}.$$

Lemma 5.2 (Markov inequality) *For a non-negative random variable X with $\mathbb{E}[X] < \infty$, it holds that, for any $t > 0$,*

$$\mathbb{P}(X > t) \leq \frac{\mathbb{E}[X]}{t}.$$

Proof: Note that

$$\mathbb{P}(X > t) = \mathbb{E}[\mathbb{1}_{\{X>t\}}] \leq \mathbb{E}[\frac{X}{t}\mathbb{1}_{\{X>t\}}] \leq \mathbb{E}[\frac{X}{t}] = \frac{\mathbb{E}[X]}{t}.$$

■

Lemma 5.3 *Suppose that $X \sim \mathcal{N}(0, \sigma^2)$. it holds that, for any $t > 0$,*

$$\mathbb{P}(X \geq t) \leq e^{\frac{t^2}{2\sigma^2}}.$$

Proof: By Markov inequality and the MGF of a Gaussian variable, for any $t > 0$ and $s > 0$,

$$\mathbb{P}(X \geq t) = \mathbb{P}(e^{sX} \geq e^{st}) \leq e^{-st}E[e^{sX}] = e^{-ts + \frac{\sigma^2}{2}s^2}.$$

Since $-ts + \frac{\sigma^2}{2}s^2$ takes its minimum at $s = \frac{t}{\sigma^2}$, we have

$$\mathbb{P}(X \geq t) \leq e^{-\frac{2t^2}{\sigma^2}}.$$

■

Proposition 5.4 *Suppose that* $X_1, \ldots, X_d \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. *Then we have*

$$\mathbb{P}(\max\{X_1, \ldots, X_d\} \geq t) \leq d \exp\left\{-\frac{t^2}{2\sigma^2}\right\}$$

Proof:

$$\begin{aligned}\mathbb{P}(\max\{X_1, \ldots, X_d\} \geq t) &\leq \mathbb{P}(\cup_{i=1}^d \{X_i \geq t\}) \\ &\leq \sum_{i=1}^d \mathbb{P}(X_i \geq t) \\ &\leq d \exp\left\{-\frac{t^2}{2\sigma^2}\right\}.\end{aligned}$$

■

Example 7.14 in Wainwright's book. Consider the classical linear Gaussian model, where $w \in \mathbb{R}^n$ has i.i.d. $\mathcal{N}(0, \sigma^2)$ entries.

Consider $\mathbf{X} \in \mathbb{R}^{n \times d}$ is fixed. Suppose that $\mathbf{X}$ satisfies the RE condition, and that it is C-column normalized, i.e.,

$$\max_{j=1,\ldots,d} \frac{\|\mathbf{X}(\cdot, j)\|_2}{\sqrt{n}} \leq C.$$

Thus, the random variable $\|\frac{\mathbf{X}^T w}{n}\|_\infty$ corresponds to the absolute maximum of d zero-mean Gaussian variables, each with variance at most $\frac{C^2\sigma^2}{n}$, since

$$var(\frac{\mathbf{X}(\cdot, j)^T w}{n}) = \frac{\sigma^2}{n^2} \sum_{i=1}^n \mathbf{X}(i, j)^2 = \sigma^2 \frac{\|\mathbf{X}(\cdot, j)\|_2^2}{n^2} \leq \frac{C^2\sigma^2}{n}.$$

The standard Gaussian tail bounds states that, for any $j \in \{1, \ldots, d\}$,

$$P\left(\left|\frac{\mathbf{X}(\cdot, j)^T w}{n}\right| \geq t\right) \leq 2\exp\{-\frac{nt^2}{2C^2\sigma^2}\} \text{ for all } t > 0.$$

Thus, for all $\delta > 0$,

$$\begin{aligned}P\left(\left\|\frac{\mathbf{X}^T w}{n}\right\|_\infty \geq C\sigma\left(\sqrt{\frac{2\log d}{n}} + \delta\right)\right) &\leq \sum_{j=1}^d P\left(\frac{\mathbf{X}(\cdot, j)^T w}{n} \geq C\sigma\left(\sqrt{\frac{2\log d}{n}} + \delta\right)\right) \\ &\leq 2d \exp\left\{-\frac{nC^2\sigma^2(\sqrt{\frac{2\log d}{n}} + \delta)^2}{2C^2\sigma^2}\right\} \\ &\leq 2e^{-n\delta^2/2}.\end{aligned}$$

If we set $\lambda_n = 2C\sigma(\frac{2\log d}{n} + \delta)$, this means that $\lambda_n \geq 2\left\|\frac{\mathbf{X}^T w}{n}\right\|_\infty$ with the probability at least $1 - 2e^{-n\delta^2/2}$. Then the theorem implies that with the probability at least $1 - 2e^{-n\delta^2/2}$, we have

$$\|\widehat{\theta} - \theta^*\|_2 \leq \frac{3}{k}\sqrt{s}\lambda_n = \frac{6C\sigma}{k}\sqrt{s}\left\{\frac{2\log d}{n} + \delta\right\}.$$

Note: If we take $\delta = \left(\frac{1}{n}\right)^{\frac{1}{2}-\alpha}$ for some $\alpha > 0$, then with the probability at least $1 - 2e^{-n^{2\alpha}/2}$, it holds that

$$\|\widehat{\theta} - \theta^*\|_2 \leq \frac{3}{k}\sqrt{s}\lambda_n = \frac{6C\sigma}{k}\sqrt{s}\left\{\frac{2\log d}{n} + \frac{1}{n^{1/2-\alpha}}\right\},$$

which would converge to zero with the rate slightly slower than $1/\sqrt{n}$.