HDP
(13) Matrix Bernstein Inequality

@ 滕佳烨

前情回顾:

$f(X) \to \mathbb{E}f(X)$ with f Lipschitz and $X \sim \mathbb{P}$

- sphere with Euclidean metric
- Gaussian RV with Euclidean metric
- Bernoulli RV with Hamming metric

→ Matrix Bernstein Inequality

前情回顾:

Bernstein’s inequality (3)

Theorem 2.8.4 (Bernstein’s inequality for bounded distributions). *Let $X_1, \dots, X_N$ be independent, mean zero random variables, such that $|X_i| \le K$ all i. Then, for every $t \ge 0$, we have*

$$\mathbb{P}\Big\{\Big|\sum_{i=1}^{N} X_i\Big| \ge t\Big\} \le 2\exp\Big(-\frac{t^2/2}{\sigma^2 + Kt/3}\Big).$$

Here $\sigma^2 = \sum_{i=1}^{N} \mathbb{E}\, X_i^2$ is the variance of the sum.

Matrix concentration (10)

Theorem 6.5.1 (Norms of random matrices with non-i.i.d. entries). *Let A be an $n \times n$ symmetric random matrix whose entries on and above the diagonal are independent, mean zero random variables. Then*

$$\mathbb{E}\,\|A\| \le C\sqrt{\log n} \cdot \mathbb{E} \max_i \|A_i\|_2,$$

where A_i denote the rows of A.

Matrix Bernstein Inequality:

Theorem 5.4.1 (Matrix Bernstein's inequality). *Let $X_1, \ldots, X_N$ be independent, mean zero, $n \times n$ symmetric random matrices, such that $\|X_i\| \le K$ almost surely for all i. Then, for every $t \ge 0$, we have*

$$\mathbb{P}\left\{\left\|\sum_{i=1}^{N} X_i\right\| \ge t\right\} \le 2n \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right).$$

Here $\sigma^2 = \left\|\sum_{i=1}^N \mathbb{E} X_i^2\right\|$ is the norm of the matrix variance of the sum.

$$\mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\| \lesssim \left\|\sum_{i=1}^{N} \mathbb{E} X_i^2\right\|^{1/2} \sqrt{1+\log n} + K(1+\log n).$$

考虑的是矩阵的**和**的norm，而不是矩阵的 norm

Hint: $f(x) = tr(\exp H + logX)$，拆分前N-1项与第N项，f(x)期望有上界

Matrix Hoeffding Inequality:

Exercise 5.4.12 (Matrix Hoeffding's inequality). Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent symmetric Bernoulli random variables and let $A_1, \ldots, A_N$ be symmetric $n \times n$ matrices (deterministic). Prove that, for any $t \geq 0$, we have

$$\mathbb{P}\left\{ \left\| \sum_{i=1}^{N} \varepsilon_i A_i \right\| \geq t \right\} \leq 2n \exp(-t^2/2\sigma^2),$$

where $\sigma^2 = \left\| \sum_{i=1}^{N} A_i^2 \right\|$.

Khintchine's Inequality:

$$\mathbb{E} \left\| \sum_{i=1}^{N} \varepsilon_i A_i \right\| \leq C\sqrt{1 + \log n} \left\| \sum_{i=1}^{N} A_i^2 \right\|^{1/2}.$$

Covariance Estimation

Theorem 5.6.1 (General covariance estimation). *Let X be a random vector in $\mathbb{R}^n$, $n \geq 2$. Assume that for some $K \geq 1$,*

$$\|X\|_2 \leq K \left(\mathbb{E}\,\|X\|_2^2\right)^{1/2} \quad \textit{almost surely.} \tag{5.16}$$

Then, for every positive integer m, we have

$$\mathbb{E}\,\|\Sigma_m - \Sigma\| \leq C\left(\sqrt{\frac{K^2 n \log n}{m}} + \frac{K^2 n \log n}{m}\right)\|\Sigma\|.$$

Exercise 5.6.4 (Tail bound). Our argument also implies the following high-probability guarantee. Check that for any $u \geq 0$, we have

$$\|\Sigma_m - \Sigma\| \leq C\left(\sqrt{\frac{K^2 r(\log n + u)}{m}} + \frac{K^2 r(\log n + u)}{m}\right)\|\Sigma\|$$

with probability at least $1 - 2e^{-u}$. Here $r = \operatorname{tr}(\Sigma)/\|\Sigma\| \leq n$ as before.

Take-away Messages

1. Matrix Bernstein's Inequality (randomness on X)

$$P\left(\left||\sum X_i|\right| \geq t\right) \leq n \exp(-\frac{t^2}{\sigma^2 + Kt})$$

2. Matrix Hoeffding Inequality (randomness on ϵ)

$$P\left(\left||\sum \epsilon X_i|\right| \geq t\right) \leq n \exp(-\frac{t^2}{2\sigma^2})$$

3. Covariance Estimation

$$\left||\Sigma_m - \Sigma|\right| \leq K \sqrt{\frac{r \log n}{m}}$$

@ 滕佳烨

Thanks!