HDP

(23) Matrix Deviation Inequality

@ 滕佳烨

前情回顾

The upper bound of $\mathbb{E} \sup_{t\in T} X_t$

1. Dudley's inequality

$$\mathbb{E}\sup_{t\in T} X_t \le CK \int_0^\infty \sqrt{\log \mathcal{N}(T,d,\varepsilon)}\, d\varepsilon.$$

2. Generic Chaining Bound with γ_2 (tighter but harder to calculate)

$$\mathbb{E}\sup_{t\in T} X_t \le CK\gamma_2(T,d).$$

The lower bound of $\mathbb{E} \sup_{t\in T} X_t$

1. Sudakov's minoration inequality

$$\mathbb{E}\sup_{t\in T} X_t \ge c\varepsilon\sqrt{\log \mathcal{N}(T,d,\varepsilon)}.$$

VC dim

- A way to model the complexity of function class
- Covering number can be bounded by VC dim $(1/\epsilon)^d$
- Empirical process can be bounded by VC dim $(\sqrt{d/n})$

Matrix deviation inequality

Theorem 9.1.1 (Matrix deviation inequality). *Let A be an $m \times n$ matrix whose rows A_i are independent, isotropic and sub-gaussian random vectors in $\mathbb{R}^n$. Then for any subset $T \subset \mathbb{R}^n$, we have*

$$\mathbb{E} \sup_{x \in T} \Big| \|Ax\|_2 - \sqrt{m}\|x\|_2 \Big| \leq CK^2 \gamma(T).$$

Here $\gamma(T)$ is the Gaussian complexity introduced in Section 7.6.2, and $K = \max_i \|A_i\|_{\psi_2}$.

Remark: the results includes random process, concentration, geometry

$\gamma(T)$: Gaussian complexity

Hint: Talagrand's comparison inequality with condition

$$\|X_x - X_y\|_{\psi_2} \leq CK^2 \|x - y\|_2$$

Matrix deviation inequality (extension)

$$\mathbb{E}\sup_{x\in T}\Big|\|Ax\|_2-\sqrt{m}\|x\|_2\Big|\le CK^2\gamma(T).$$

Expectation version:

$$\mathbb{E}\sup_{x\in T}\Big|\|Ax\|_2-\mathbb{E}\,\|Ax\|_2\Big|\le CK^2\gamma(T).$$

Tail version:

$$\sup_{x\in T}\Big|\|Ax\|_2-\sqrt{m}\|x\|_2\Big|\le CK^2\left[w(T)+u\cdot\mathrm{rad}(T)\right] \tag{9.12}$$

Square version:

$$\mathbb{E}\sup_{x\in T}\Big|\|Ax\|_2^2-m\|x\|_2^2\Big|\le CK^4\gamma(T)^2+CK^2\sqrt{m}\,\mathrm{rad}(T)\gamma(T).$$

Application

$$\mathbb{E}\sup_{x\in T}\left|\|Ax\|_2-\sqrt{m}\|x\|_2\right|\leq CK^2\gamma(T).$$

Setting $T=S^{n-1}$: random matrix concentration

$$\sqrt{m}-CK^2(\sqrt{n}+u)\leq s_n(A)\leq s_1(A)\leq\sqrt{m}+CK^2(\sqrt{n}+u).$$

Setting $T=T-T$: random projection

$$\mathbb{E}\operatorname{diam}(PT)\leq\sqrt{\frac{m}{n}}\operatorname{diam}(T)+CK^2w_s(T).$$

Setting $T=\Sigma^{1/2}S^{n-1}$ (square version): covariance estimation with stable rank

$$\mathbb{E}\|\Sigma_m-\Sigma\|\leq CK^4\left(\sqrt{\frac{r}{m}}+\frac{r}{m}\right)\|\Sigma\|.$$

Setting $T=\mathcal{X}$: JL lemma on infinite set

$$\|x-y\|_2-\delta\leq\|Qx-Qy\|_2\leq\|x-y\|_2+\delta\quad \textit{for all } x,y\in\mathcal{X}$$

Take-away Messages

1. Matrix Deviation Inequality: random process, concentration, geometry

* General

@ 滕佳烨

Thanks!