36-710: Advanced Statistical Theory **Fall 2018**

Lecture 13: February 26

Lecturer: Alessandro Rinaldo *Scribes: Ojash Neopane*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

13.1 Preliminaries

Definition 13.1 (Operator Norm of a Symmetric Matrix) *Let $\mathbf{A} \in S^{d\times d}$ be a symmetric $d \times d$ matrix. Then the operator norm of $\mathbf{A}$, denoted $||\mathbf{A}||_{op}$ is defined as*

$$||\mathbf{A}||_{op} \coloneqq \max_{x\in\mathbb{S}^{d-1}} |x^T\mathbf{A}x| \tag{13.1}$$

Definition 13.2 (Sub-Gaussian Random Vector) *A vector $\mathbf{x} \in \mathbb{R}^d$ is vector sub-gaussian ($\mathbf{x} \in SG_d(\sigma^2)$ with parameter σ if for all $\mathbf{v} \in \mathcal{S}^{d-1}$*

$$\mathbb{E}[\exp(\mathbf{v}^T(\mathbf{x} - \mu))] \leq \exp(\lambda^2\sigma^2/2) \tag{13.2}$$

13.2 Concentration Inequalities for Covariance Matrices

Theorem 13.3 *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be an i.i.d sequence of σ sub-gaussian random vectors such that $\mathbb{V}[\mathbf{x}_1] = \Sigma$ and let $\hat{\Sigma}_n \coloneqq \frac{1}{n}\sum_{i=1}^n x_i x_i^T$ be the empirical covariance matrix. Then there exists a universal constant $C > 0$ such that, for $\delta \in (0, 1)$, with probability at least $1 - \delta$*

$$\frac{||\hat{\Sigma}_n - \Sigma||_{op}}{\sigma^2} \leq C \max\left\{\sqrt{\frac{d + \log(2/\delta)}{n}}, \frac{d + \log(2/\delta)}{n}\right\} \tag{13.3}$$

If $d/n \to 0$ then the confidence interval goes to 0 at a rate of $\sqrt{d/n}$ which is the minimax rate

Proof: We break the proof up into two steps:

1. Use a discretization argument to reduce the problem to task of computing the maximum of finitely many random variables
2. Use standard concentration inequalities

Step 1:

Lemma 13.4 *Let $\mathbf{A} \in S^{d\times d}$ and let N_ϵ be an ϵ-net of $\mathbb{S}^{d-1}$. Then*

$$||\mathbf{A}||_{op} \leq \frac{1}{1 - 2\epsilon} \max_{y\in N_\epsilon} |y^T\mathbf{A}y| \tag{13.4}$$

Proof of Lemma 13.4:
Let $y \in N_\epsilon$ satisfy $||x-y|| \le \epsilon$. Then

$$|x^{\mathbf{A}}x - y^T\mathbf{A}y| = |x^T\mathbf{A}(x-y) + y^T\mathbf{A}(x-y)| \tag{13.5}$$
$$\le |x^T\mathbf{A}(x-y)| + |y^T\mathbf{A}(x-y)| \tag{13.6}$$

Looking at $|x^T\mathbf{A}(x-y)|$ we have

$$|x^T\mathbf{A}(x-y)| \le ||\mathbf{A}(x-y)||||x|| \tag{13.7}$$
$$\le ||\mathbf{A}||_{op}\underbrace{||x-y||}_{\le\epsilon}\underbrace{||x||}_{=1} \tag{13.8}$$
$$\le ||\mathbf{A}||_{op}\epsilon \tag{13.9}$$

Applying the same argument to $|y^T\mathbf{A}(x-y)|$ gives us $|x^{\mathbf{A}}x - y^T\mathbf{A}y| \le 2\epsilon||\mathbf{A}||_{op}$. To complete the proof, we see that $||\mathbf{A}||_{op} = \max_{x\in\mathbb{S}^{d-1}} x^T\mathbf{A}x \le 2\epsilon||\mathbf{A}||_{op} + \max_{y\in N_\epsilon} y^T\mathbf{A}y$. Rearranging the equation gives $||\mathbf{A}||_{op} \le \frac{1}{1-2\epsilon}\max_{y\in N_\epsilon} y^T\mathbf{A}y$ as desired.

Step 2:
Applying Lemma 13.4 on $\hat{\Sigma}_n - \Sigma$ with $\epsilon = 1/4$ we have

$$||\hat{\Sigma}_n - \Sigma||_{op} \le 2\max_{v\in N_{1/4}} |v^T(\hat{\Sigma}_n - \Sigma)v| \tag{13.10}$$

Additionally, we know that $N_{1/4} \le 9^d$. From here, we can apply standard cocentration tools as follows:

$$\mathbb{P}(||\hat{\Sigma}_n - \Sigma||_{op} \ge t) \le \mathbb{P}(\max_{v\in N_{1/4}} |v^T(\hat{\Sigma}_n - \Sigma)v| \ge t/2) \tag{13.11}$$
$$\le |N_{1/4}|\mathbb{P}(|v_i^T(\hat{\Sigma}_n - \Sigma)v_i| \ge t/2) \tag{13.12}$$

We rewrite $v_i^T(\hat{\Sigma}_n - \Sigma)v_i$ as follows:

$$v_i^T(\hat{\Sigma}_n - \Sigma)v_i = \frac{1}{n}\sum_{i=1}^n (v_i^Tx_j)^2 - \mathbb{E}[(v_i^Tx_j)^2] \tag{13.13}$$
$$= \frac{1}{n}\sum_{i=1}^n z_j - \mathbb{E}[z_j] \tag{13.14}$$

where z_j's are independent and by assumption $v_i^Tx_j \in SG(\sigma^2)$ so that $z_j - \mathbb{E}[z_j] \in SE((16\sigma^2)^2, 16\sigma^2)$. Applying the sub-exponential tail bound gives us

$$\mathbb{P}(|v_i^T(\hat{\Sigma}_n - \Sigma)v_i| \ge t/2) \le 2\exp\left\{-\frac{n}{2}\min\left\{(\frac{t}{32\sigma^2)^2}, \frac{t}{32\sigma^2}\right\}\right\} \tag{13.15}$$

so that

$$\mathbb{P}(||\hat{\Sigma}_n - \Sigma||_{op} \ge t) \le 2\cdot 9^d\exp\left\{-\frac{n}{2}\min\left\{(\frac{t}{32\sigma^2)^2}, \frac{t}{32\sigma^2}\right\}\right\} \tag{13.16}$$

Inverting the bound gives the desired result ■

13.3 Matrix Concentration Inequalities

Theorem 13.5 (Matrix Bernstein) *Let $X_1, \dots, X_n$ be independent mean 0 symmetric $d \times d$ random matrices such that $||X||_{op} \le C$ almost surely. Then for any $t \ge 0$*

$$P(||\sum X_i|| > t) \le 2d\exp(-\frac{t^2}{2(\sigma^2 + Ct/3)}) \tag{13.17}$$

where $\sigma^2 = ||\sum \mathbb{E}[X_i]||_{op}$

Some applications of matrix concentration inequalities include:

- Solving Linear Systems
- Matrix Multiplication
- Sub Sampling
- Sparsification methods for spectral clustering
- Dimensionality Reduction
- Compressed Sensing
- Network Models

For a more in-depth discussion on these topics, refer to [tropp2012user, tropp2015introduction]

References

[tropp2012user] J. Tropp, "User-friendly tail bounds for sums of random matrices,"

[tropp2015introduction] J. Tropp, "An introduction to matrix concentration inequalities,"