36-709: Advanced Statistical Theory	**Spring 2019**

Lecture 22: April 9, 2019

Lecturer: Alessandro Rinaldo *Scribes: Yihang Shen*

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

This lecture's notes illustrate some uses of various LaTeX macros. Take a look at this and imitate.

22.1 Announcements before class

Final exam: last day of classes, THU, May 2, 1 hour and 20 mins (not 3 hours this time), open books, open notes, open laptops (no internet). Some practice exams will be sent soon.

A new homework will be posted soon.

22.2 Distance between linear sub-spaces

Last time we talked about one issue with PCA: how well the eigenspace approximate the original covariance matrix. Today we will discuss the distance between linear sub-spaces first.

Let $\mathcal{E}$ and $\mathcal{F}$ be d-dimensional sub-spaces in R^p, let $P_{\mathcal{E}}$ and $P_{\mathcal{F}}$ be the projection matrices onto $\mathbb{E}$ and $\mathbb{F}$.

$$P_{\mathbb{E}^\perp} = \mathbb{I} - P_{\mathcal{E}}$$

$$P_{\mathbb{E}^\perp} = \mathbb{I} - P_{\mathcal{E}}$$

$\mathbb{E}$ and $\mathbb{F}$ are $p \times d$ matrices with orthonormal columns, whose column ranges are $\mathcal{E}$ and $\mathcal{F}$.

Then we have:

$$P_{\mathcal{E}} = \mathbb{E}\mathbb{E}^\perp$$

$$P_{\mathcal{F}} = \mathbb{F}\mathbb{F}^\perp$$

Let v_1 and v_2 be unit vectors in R^p, the angle between v_1 and v_2 is:

$$\angle(v_1, v_2) = cos^{-1}|v_1^T v_2|$$

That is:

$$cos(\angle(v_1, v_2)) = |v_1^T v_2|$$

Definition 22.1 *The canonical/principle angle between two sub-spaces $\mathcal{E}$ and $\mathcal{F}$ are the $\theta_1 = cos^{-1}(\sigma_1)$, ..., $\theta_d = cos^{-1}(\sigma_d)$, where $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_d \geq 0$ are the singular values of $\mathbb{E}^T\mathbb{F}$ or $\mathbb{F}^T\mathbb{E}$. Therefore, we have:*

$$\mathbb{E}^T\mathbb{F} = Ucos(\Theta)V$$

$$\Theta = \begin{bmatrix} \theta_1 & & 0 \\ & \ddots & \\ 0 & & \theta_d. \end{bmatrix}$$

We can also have an equivalent definition:

Definition 22.2 *The first canonical angle is:*

$$cos^{-1}(\max_{x \in \mathcal{E}} \max_{y \in \mathcal{F}} |x^T y|)$$

subject to:

$$||x|| = ||y|| = 1$$

For $k = 2, ..., d$, the k-th canonical angle is:

$$cos^{-1}(\max_{x \in \mathcal{E}} \max_{y \in \mathcal{F}} |x^T y|)$$

subject to:

$$||x|| = ||y|| = 1 \quad \textit{and} \quad y^T y_i = x^T x_i = 0 \quad \textit{for any i=1,...,k-1}$$

where (x_k, y_k) is a pair realizing the k-th canonical angle.

We could equivalently give another definition:

Definition 22.3 *Let $\theta_i = sin^{-1}(s_i)$ for $i = 1, ..., d$, where $s_1 \geq ... \geq s_d \geq 0$ are the singular values of $P_{\mathcal{E}}P_{\mathcal{F}^\perp}$ or $P_{\mathcal{F}}P_{\mathcal{E}^\perp}$. ($P_{\mathcal{E}}P_{\mathcal{F}^\perp} = Usin(\Theta)V^T$)*

Now given the definition of the canonical angles, we can define the distance between sub-spaces $\mathcal{E}$ and $\mathcal{F}$:

Definition 22.4 *The distance between $\mathcal{E}$ and $\mathcal{F}$ is $||sin(\Theta)||_F = ||sin(\Theta(\mathcal{E}, \mathcal{F}))||_F$.*

Then we have:

$$||sin(\Theta)||_F = ||P_{\mathcal{E}}P_{\mathcal{F}^\perp}||_F^2 = ||P_{\mathcal{E}}(\mathbb{I} - P_{\mathcal{F}})||_F^2 = ||P_{\mathcal{F}}(\mathbb{I} - P_{\mathcal{E}})||_F^2 = \frac{1}{2}||P_{\mathcal{F}} - P_{\mathcal{E}}||_F^2$$

22.3 Davis-Kahan Theorem

Theorem 22.5 *(Davis-Kahan Theorem) Let Σ and $\hat{\Sigma}$ be $p \times p$ symmetric matrices with eigenvalues:*

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$$

and:

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ... \geq \hat{\lambda}_p$$

Fix $1 \leq r \leq s \leq p$, and let $d = s - r + 1$, let V and $\hat{V}$ be $p \times d$ matrices consisting of the eigenvectors of Σ and $\hat{\Sigma}$ corresponding to eigenvalues $\lambda_r, ..., \lambda_s$ and $\hat{\lambda_r}, ..., \hat{\lambda_s}$. Let:

$$\delta = \inf\{|\lambda - \hat{\lambda}|, \lambda \in [\lambda_s, \lambda_r], \hat{\lambda} \in (-\infty, \hat{\lambda}_{s+1}] \bigcup [\hat{\lambda}_{r-1}, +\infty)\}$$

Here we define $\hat{\lambda}_0 = \lambda_0 = +\infty$ and $\hat{\lambda}_{p+1} = \lambda_{p+1} = -\infty$.

If $\delta > 0$, then:

$$||sin(\Theta(\mathcal{E}, \mathcal{F}))||_F \leq \frac{||\Sigma - \hat{\Sigma}||_F}{\delta}$$

The Frobenius Norm $|| \cdot ||_F$ used in the theorem can be replaced by operator norm $|| \cdot ||_{op}$.

Let's assume γ_n be s.t. $||\Sigma - \hat{\Sigma}||_{op} \leq \gamma_n$ with high probability, and also assume that:

$$|\hat{\lambda}_{s+1} - \lambda_s| \geq \lambda_s - \lambda_{s+1} - \gamma_n \geq 0$$

and:

$$|\hat{\lambda}_{r-1} - \lambda_r| \geq \lambda_{r-1} - \lambda_r - \gamma_n \geq 0$$

The eigen-gap $\delta^* = \min\{\lambda_s - \lambda_{s+1}, \lambda_{r-1} - \lambda_r\} \geq \gamma_n$, then the bound is $\frac{||\Sigma - \hat{\Sigma}||_F}{\delta^* - \gamma_n}$.

Typically, when $r = 1$ and $s = d < p$, which gives $\delta^* = \lambda_d - \lambda_{d+1}$, Yu, Wang & Samworth (A useful variant of the Davis-Kahan theorem for statisticians) proved that:

$$||sin(\Theta(\mathcal{E}, \mathcal{F}))||_F \leq \frac{2 \min\{\sqrt{d}||\Sigma - \hat{\Sigma}||_{op}, ||\Sigma - \hat{\Sigma}||_F\}}{\min\{\lambda_s - \lambda_{s+1}, \lambda_{r-1} - \lambda_r\}}$$

Also, $\exists O$ be $d \times d$ orthogonal matrix such that:

$$||\hat{V}O - V||_F \leq \frac{2^{\frac{3}{2}} \min\{\sqrt{d}||\Sigma - \hat{\Sigma}||_{op}, ||\Sigma - \hat{\Sigma}||_F\}}{\min\{\lambda_s - \lambda_{s+1}, \lambda_{r-1} - \lambda_r\}}$$

if $r = s = 1$, then:

$$sin\angle(v_1, \hat{v}_1) \leq \frac{2||\Sigma - \hat{\Sigma}||_F}{\lambda_1 - \lambda_2}$$

and:

$$\min_{\epsilon \in \{-1,1\}} ||\epsilon \hat{v}_1 - v_1|| \leq 2^{\frac{3}{2}} \frac{||\Sigma - \hat{\Sigma}||_F}{\lambda_1 - \lambda_2}$$

22.3.1 Spiked Covariance Model

Let $p \times p$ matrix $\Sigma = \theta v v^T + \mathbb{I}_p$, $\theta > 0$ and $||v|| = 1$, then we can know that $1 + \theta$ is the leading eigenvalue and v is the leading eigenvector.

Σ is the covariance matrix of $v(v^T X) + Z$ where $Z \sim \mathcal{N}_p(0, \mathbb{I})$ and $X \sim \mathcal{N}_p(0, \mathbb{I})$, $Z \perp X$.

Let $Y_1, Y_2, .., Y_n \sim (0, \Sigma)$ are sub-Gaussian random variables with parameter $||\Sigma||_{op} = 1 + \theta$, then we have:

$$\min_{\epsilon \in \{-1,1\}} ||\hat{v}\epsilon - v|| \leq C \frac{1+\theta}{\theta} \max\{\sqrt{\frac{p + \ln \frac{1}{\delta}}{n}}, \frac{p + \ln \frac{1}{\delta}}{n}\}$$

with probability at least $1 - \delta$. Here, $\hat{v}$ is the leading eigenvector of $\Sigma_{i=1}^{n} \frac{y_i y_i^T}{n}$.