36-755: Advanced Statistical Theory I **Fall 2017**

Lecture 16: October 25

Lecturer: Alessandro Rinaldo *Scribes: Yangyi Lu*

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.

16.1 Principal Component Analysis

16.1.1 Davis - Kahan Theorem

We show a theorem that provide a bound for the distance between two subspaces.

Theorem 16.1 *Davis - Kahan theorem*
Let Σ and $\hat{\Sigma}$ be $d \times d$ symmetric matrices with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_d$ and $\hat{\lambda_1}, \hat{\lambda_2}, \cdots, \hat{\lambda_d}$ respectively. Fix $1 \leq r \leq s \leq d$, let V and $\hat{V}$ be $d \times (s-r+1)$ matrices with columns corresponding to eigenvectors for $\lambda_J, J = 1, \ldots, s$ and $\hat{\lambda_J}, J = 1, \ldots, s$. Let

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

By convention, $\hat{\lambda}_0 = -\infty$, $\hat{\lambda}_{d+1} = \infty$.

Then, let $\mathcal{E} = range(V)$ and $\mathcal{F} = range(\hat{V})$, we have following bound,

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

Some intuitions behind this theorem:

1. We are interested in bounding distance between subspaces by leading r eigenvectors of Σ and $\hat{\Sigma} = \Sigma + E$.
2. When $\mathcal{E}$ and $\mathcal{F}$ can be regarded as close to each other?
   Take $x \in \mathcal{R}^d$, $\Pi_{\mathcal{E}} x$ is projection of x onto $\mathcal{E}$. If $\mathcal{F}$ is a good approximation to $\mathcal{E}$, then this quantity should be small.
   $$||\Pi_{\mathcal{E}} x - \Pi_{\mathcal{F}}(\Pi_{\mathcal{E}} x)|| = ||(\mathcal{I} - \Pi_{\mathcal{F}})\Pi_{\mathcal{E}} x||$$
   which means $\Pi_{\mathcal{F}}\Pi_{\mathcal{E}} \approx \Pi_{\mathcal{E}}\Pi_{\mathcal{E}} = \Pi_{\mathcal{E}}$. Last time we show,
   $$||sin\Theta(\mathcal{E}, \mathcal{F})||_F^2 = ||(\mathcal{I} - \Pi_{\mathcal{F}})\Pi_{\mathcal{E}} x||_2^2 = \frac{1}{2}||\Pi_{\mathcal{F}} - \Pi_{\mathcal{E}}||_2^2$$

A useful variant of D.K. theorem.

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

where $q = s - r + 1$. If $r = 1$, you are considering s leading eigenvalues of Σ and $\hat{\Sigma}$.

Also, $\exists O \in \mathcal{R}^{d\times d}$ orthogonal, s.t.

$$\min_{\epsilon\in\{1,-1\}} ||VO - \hat{V}||_F \leq 2^{3/2}\frac{||\Sigma - \hat{\Sigma}||_F}{\delta}$$

Example.
If $s = r = 1$, we have,

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

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

16.1.2 Applications of D.K theorem - spiked covariance model

Sparse PCA

$$\Sigma = \theta vv^T + I_d$$

where $v \in S^{d-1}$. Then eigenvalues of Σ are $1 + \theta, 1, \ldots, 1$.
We can view Σ as the covariance matrix of $\sqrt{\theta}\xi v + \epsilon$, where $\xi \sim N(0, 1), \epsilon \sim N(0, I_d)$ or $v(v^T y) + \epsilon$, where $y \sim N(0, \theta I_d)$. Eigen gap is $\lambda_1 - \lambda_2 = 1 + \theta - 1 = \theta$. Let $\hat{V}$ be leading eigenvector of $\hat{\Sigma}$,

$$\min_{\epsilon\in\{1,-1\}} ||\epsilon V - \hat{V}|| \leq \frac{2^{3/2}||\Sigma - \hat{\Sigma}||_{op}}{\theta}$$

$$\lesssim \frac{1+\theta}{\theta}\max\{\sqrt{\frac{d + log(1/\delta)}{n}}, \frac{d + log(1/\delta)}{n}\}$$

with prob $\geq 1 - \delta$.

16.2 Sparse PCA

Again, for sparse spiked covariance model where $\theta > 0, v \in S^{d-1}, ||v||_0 \leq k \leq d/2$,

$$\Sigma = \theta vv^T + I_d$$

We estimate the eigenvector corresponding to the lartest eigenvalue: $\hat{v}$ using

$$\hat{v} = argmax_{z\in S^{d-1}} z^T\hat{\Sigma}z$$

with $||z||_0 \leq k' \leq d/2, k' \geq k$.

Theorem 16.2 *Assume $X_1, \cdots, X_n$ are of zero-mean and covariance Σ, $X_i \in SG_d(||\Sigma||_{op})$. Then:*

$$\min_{\epsilon\in\{1,-1\}} ||\epsilon v - \hat{v}|| \lesssim \frac{1+\theta}{\theta}Cmax\{\eta_n, \eta_n^2\}$$

with prob $\geq 1 - \delta$ for some C, where $\eta_n = \sqrt{\frac{(k'+k)log(\frac{ed}{k'+k}) + log(1/\delta)}{n}}$.

Proof: For all $\hat{v}$,

$$\begin{aligned} v^T\Sigma v - \hat{v}^T\Sigma\hat{v} &= \theta(1-\cos^2(\angle(v,\hat{v}))) \\ &= \theta\sin^2(\angle(v,\hat{v}))) \end{aligned}$$

Then,

$$\begin{aligned} v^T\Sigma v - \hat{v}^T\Sigma\hat{v} &= v^T\hat{\Sigma}v - \hat{v}^T\Sigma\hat{v} - \hat{v}(\hat{\Sigma}-\Sigma)v \\ &\leq \hat{v}^T\hat{\Sigma}\hat{v} - \hat{v}^T\Sigma\hat{v} - \hat{v}(\hat{\Sigma}-\Sigma)v \\ &= \hat{v}^T(\hat{\Sigma}-\Sigma)\hat{v} - v^T(\hat{\Sigma}-\Sigma)v \\ &=< \hat{\Sigma}-\Sigma, \hat{v}\hat{v}^T - vv^T > \end{aligned} \tag{16.1}$$

where $< A, B >= tr(A^TB)$. Also, observe that the Frobenius norm $||A||_F^2 =< A, A >$.

Now, we know that $v, \hat{v}$ are k, k'-sparse respectively, which implies that $\exists S \subset \{1, 2, \ldots, d\}$ with $|S| \leq k+k'$,s.t.

$$Equation\ 16.1 =< \hat{\Sigma}_S - \Sigma_S, \hat{v}_S\hat{v}_S^T - v_Sv_S^T >$$

where $\hat{\Sigma_S}, \Sigma_S$ are sub-matrices of $\hat{\Sigma}, \Sigma$ respectively rows/columns in S and $\hat{v}_S, v_S$ are sub-vectors of $\hat{v}, v$ respectively with entries in S.

Now we have,

$$\begin{aligned} < \hat{\Sigma}-\Sigma, \hat{v}\hat{v}^T - vv^T > &\leq ||\hat{\Sigma}_S - \Sigma_S||_{op}||\hat{v}_S\hat{v}_S^T - v_Sv_S^T||_1 \\ &\leq ||\hat{\Sigma}_S - \Sigma_S||_{op}\sqrt{2}||\hat{v}_S\hat{v}_S^T - v_Sv_S^T||_2 \\ &= ||\hat{\Sigma}_S - \Sigma_S||_{op}\sqrt{2}||\hat{v}\hat{v}^T - vv^T||_2 \\ &= ||\hat{\Sigma}_S - \Sigma_S||_{op}\sqrt{2}||\hat{v}\hat{v}^T - vv^T||_F \\ &= ||\hat{\Sigma}_S - \Sigma_S||_{op}\sqrt{2}\sqrt{2-2(v^T\hat{v})^2} \\ &= ||\hat{\Sigma}_S - \Sigma_S||_{op}\sqrt{2}\sqrt{2\sin^2(\angle(v,\hat{v}))} \\ &= 2||\hat{\Sigma}_S - \Sigma_S||_{op}\sin(\angle(v,\hat{v})) \end{aligned} \tag{16.2}$$

Combine Equation 16.1 and Equation 16.2, we get the following result:

$$\theta\sin(\angle(v,\hat{v})) \leq 2||\hat{\Sigma}_S - \Sigma_S||_{op} \tag{16.3}$$

S is random, because it is the union of $supp(v)$ and $supp(\hat{v})$, so in order to bound Equation 16.3, we need to sup-out the randoms, and bound the larger term.

To be continued next week... ■

References