36-755: Advanced statistics **Fall 2016**

Lecture 13: October 12

Lecturer: Alessandro Rinaldo *Scribes: Jackie Mauro*

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

13.1 PCA

Even if we can estimate eigenvalues well, it can be very hard to estimate eigenvectors.

Let $\mathcal{E}$ and $\mathcal{F}$ be d-dimensional linear subspaces in $\mathbb{R}^p$ (if they have different dimensions, just take d to be the smaller one).

Let $P_{\mathcal{E}}$ and $P_{\mathcal{F}}$ be orthogonal projection matrices onto $\mathcal{E}$ and $\mathcal{F}$.

If $\mathcal{E}$ and $\mathcal{F}$ are 1-dimensional, spanned by $v_{\mathcal{E}}$ and $v_{\mathcal{F}} \in \mathbb{S}^{p-1}$, we can measure their distance by looking at the angle between $v_{\mathcal{E}}$ and $v_{\mathcal{F}}$:

$$\angle(v_{\mathcal{E}}, v_{\mathcal{F}}) = \cos^{-1}(|v_{\mathcal{E}}, v_{\mathcal{F}}|)$$

Notes: we need to normalize to the unit norm and the use of absolute value is because we are only looking at the acute angle.

For the subspaces $\mathcal{E}$ and $\mathcal{F}$, let E and F be $p \times d$ matrices with orthonormal columns and $range(E) = \mathcal{E}$, $range(F) = \mathcal{F}$.

$$\begin{aligned} P_{\mathcal{E}} &= EE^T \\ P_{\mathcal{F}} &= FF^T \end{aligned}$$

The k-th principal or canonical angle between $\mathcal{E}$ and $\mathcal{F}$, $k = 1, \ldots, d$ is defined as:

$$\cos^{-1}\left(\max_{\substack{x \in \mathcal{E} \\ ||x||=1}} \max_{\substack{y \in \mathcal{F} \\ ||y||=1}} |x^T y| \right) = \cos^{-1}\left(|x_k^T y_k|\right)$$

with $x_i^T x_i = y_i^T y_i = 0$, $i = 1, \cdots, k-1$.

Definition 13.1 *SVD for Principal/canonical angles. The principal or canonical angles between $\mathcal{E}$ and $\mathcal{F}$ are:*

$$\theta_1 = cos^{-1}(\sigma_1), \cdots, \theta_d = cos^{-1}(\sigma_d) \tag{13.1}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_d \in [0, 1]$ are the singular values of $E^T F$ or $F^T E$ (doesn't matter).

$$E^T F = U cos \Theta v^T \tag{13.2}$$

where

$$\Theta = \begin{bmatrix} \theta_1 & 0 & \dots \\ 0 & \theta_2 & \dots \\ \vdots & \vdots & \ddots \\ \dots & 0 & \theta_d \end{bmatrix}$$

We can equivalently define the principal or canonical angles as

$$\theta_k = sin^{-1}(s_k), k = 1, \dots, d$$

where s_k's are the singular values of $P_{\mathcal{E}}(I - P_{\mathcal{F}}) = P_{\mathcal{E}} P_{\mathcal{F}}^{\perp} = U sin \Theta v^T$. Follows from CS decomposition (Stewart & Jun 1990)

Definition 13.2 *The distance between $\mathcal{E}$ and $\mathcal{F}$ is:*

$$||sin\Theta(\mathcal{E}, \mathcal{F})||_F$$

This is actually a metric between d-dimensional subspaces.

In particular:

$$\begin{aligned} ||sin\Theta(\mathcal{E}\mathcal{F})||_F^2 &= ||P_{\mathcal{E}} P_{\mathcal{F}}^{\perp}||_F^2 \\ &= \frac{1}{2}||P_{\mathcal{E}} P_{\mathcal{F}}||_F^2 \end{aligned}$$

Recall: $P_{\mathcal{F}}^{\perp} = I_p - P_{\mathcal{F}}$

So this is one way to define the distance between two subspaces.

Theorem 13.3 *Davis - Kahan $sin\theta$ theorem*
See [YWS14]. Let Σ and $\hat{\Sigma}$ be $p \times p$ symmetric matrices with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ and $\hat{\lambda_1} \geq \hat{\lambda_2} \geq \dots \geq \hat{\lambda_p}$ respectively.

Let $1 \leq r < s \leq p$ and $d = s - r + 1$.

Let V and $\hat{V}$ be $p \times d$ matrices with columns given by eigenvectors of Σ and $\hat{\Sigma}$ corresponding to λ_j and $\hat{\lambda_j}$, $j = r, \cdots, s$. By construction, V and $\hat{V}$ have orthonormal columns.

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

$\hat{\lambda_0} = -\infty, \hat{\lambda}_{p+1} = \infty$ by convention.

If $\delta > 0$ (meaning there is an eigengap), then:

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

Where $\mathcal{E} = range(v)$, and $\mathcal{F} = range(\hat{V})$

The same inequality holds for the operator norm $||\cdot||_{op}$ and for any unitarialy innvariant matrix norm:

$$|||A||| = |||OAU^T|||$$

How do we use this in practice? Assume that $||\Sigma - \hat{\Sigma}||_{op} \leq \gamma n$ with high probability and:

$$\begin{aligned} |\hat{\lambda_{s+1}} - \lambda_s| &\geq \lambda_s - \lambda_{s+1} - \gamma n > 0 \\ |\hat{\lambda_{r-1}} - \lambda_r| &\geq \lambda_{r-1} - \lambda_r - \gamma n > 0 \end{aligned}$$

Then Davis - Kahan gives:

$$||sin\Theta|| \leq \frac{||\Sigma - \hat{\Sigma}||}{\delta^* - \gamma n}$$

where $\delta^* = min\{\lambda_s - \lambda_{s+1}, \lambda_{r-1} - \lambda_r\}$

Typically $r = 1$, $s = d < p$ which gives $\delta* = \lambda_d - \lambda_{d+1}$ the eigengap.

λ_{d+1} is the first eigenvalue we are not interested in. If it is too close to the ones we are interested in, it will contaminate them and we can't tell them apart.

If $\gamma n \to 0$, $\delta^* - \gamma n \geq \frac{\delta}{2}$ for n large. The improvement by Yu, Wang and Samworth [YWS14] is:

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

Application: Spiked Covariance Model

$\Sigma = \theta v v^T + I_p$, $\theta > 0, v \in \mathbb{S}^{p-1}$.

The eigenvalues of Σ_i are $(1 + \theta, 1, \cdots, 1)$ (p-1 coordinates equal to 1).

θ is the eigengap.

If $X_1, \cdots, X_n \overset{iid}{\sim} (0, \Sigma_i), X_i \in SG_p(||\Sigma_i||_{op})$.

Let $\hat{v}$ be the largest eigenvector of $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^T$.

Let v be the leading eigenvector of Σ. Then:

$$\begin{aligned} \min_{\epsilon \in \{-1,1\}} ||\epsilon\hat{v} - v||^2 &= 2 - 2|v^T v| \\ &\leq 2 - 2(\hat{v}^T v)^2 \\ &= 2sin^2(\angle(\hat{v}, v)) \\ &= ||\hat{v}v^T - vv^T||_F^2 \end{aligned}$$

Then by Davis-Kahan (modified):

$$\begin{aligned} \min_{\epsilon \in \{-1,1\}} ||\epsilon\hat{v} - v||^2 &\leq \sqrt{\frac{8}{\theta^2}||\Sigma - \hat{\Sigma}||_{op}^2} \\ &\lesssim \frac{1+\theta}{\theta} max\{\frac{\sqrt{p + log(1/\delta)}}{n}, \frac{d + log(1/\delta)}{n}\} \end{aligned}$$

Comes from our previous result on $||\Sigma - \hat{\Sigma}||_{op}$

13.2 Sparse PCA

$\Sigma = \theta v v^T + I_p$ with $\theta > 0$, $v \in \mathbb{S}^{p-1}$, $||v||_0 = k \leq d/2$. Task: estimate v using $\hat{v}$, the solution to:

$$\max_{\substack{u \in \mathbb{S}^{p-1} \\ ||u||_0 \leq k' \leq d/2}} \quad u^T \hat{\Sigma} u$$

Note: this is not computationally feasible.

Theorem 13.4 *Assume* $X_1, \cdots, X_n \overset{iid}{\sim} (0, \Sigma)$, $X_i \in SG(||\Sigma||_{op})$. *Then:*

$$min_{\epsilon \in \{-1,1\}} ||\hat{v}\epsilon - v||^2 \lesssim \frac{1+\theta}{\theta} max\{A, \sqrt{A}\}$$

For $A = [(k+k') log \frac{ep}{k+k'} + log(\frac{1}{\delta})] \frac{1}{n}$

Proof: we have that:

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

Next:

$$\begin{aligned} v^T \Sigma v - \hat{v}^T \Sigma \hat{v} &= v^T \hat{\Sigma} v - \hat{v} \Sigma \hat{v} - v^T (\hat{\Sigma} - \Sigma) v \\ &\leq \hat{v^T} \hat{\Sigma} \hat{v} - \hat{v} \Sigma \hat{v} - v^T (\hat{\Sigma} - \Sigma) v \\ &= \hat{v}^T (\hat{\Sigma} - \Sigma) \hat{v} - v^T (\hat{\Sigma} - \Sigma) v \end{aligned}$$

■

References

[YWS14] Y. Yu, T. Wang and R.J. Samworth "A useful variant of the DavisKahan theorem for statisticians," *Biometrika 102.2*, 2015, pp. 315–323.