Lecture 12: PCA

Isabella Zhu

18 March 2025

Assume that $X_1, \dots X_n \sim N_d(0, \Sigma)$. The data is assumed to be somewhat low-dimensional, close to span(v). Our goal is to estimate v.

§1 Spiked Covariance Model

Let $Y_1, \dots Y_n$ be i.i.d $N(0, \theta)$. Let $Z_1, \dots Z_n$ be i.i.d $N_d(0, I_d)$. We observe $X_i = Y_i v + Z_i$ for $1 \le i \le n$. Assume $|v|_2 = 1$. Clearly, $X_i \sim N_d(0, \Sigma)$, where

$$\Sigma = \mathbb{E}[XX^T] = \mathbb{E}[(Yv + Z)(Yv + Z^T)] = \theta v v^T + I_d$$

This is known as the **spiked covariance model**.

The eigenvalues of Σ are $1 + \theta$ and remaining are 1s. Define sample covariance matrix

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T$$

§2 PCA

Sample covariance matrix $\hat{\Sigma}$. Estimator $\hat{v}$ is the largest eigenvector of $\hat{\Sigma}$.

Theorem 2.1

Estimator $\hat{v}$ is at most

$$|\sin(v, \hat{v})| \lesssim \frac{1 + \theta}{\theta} \sqrt{\frac{d}{n}}$$

away from v, assuming d is much smaller than n.

Proof. We use Davis-Kahan to get

$$|\sin\langle(\hat{v}, v)| \le \frac{2}{\theta} ||\hat{\Sigma} - \Sigma||_{op}$$

To bound $||\hat{\Sigma} - \Sigma||_{op}$, we use the ϵ-net argument. Let N be a quarter-net of B_2. We have $|N| \le 9^d$. We get

$$||\hat{\Sigma} - \Sigma||_{op} \le \max_{x,y \in B_2} x^T(\hat{\Sigma} - \Sigma)y \le 2 \max_{x,y \in N} x^T(\hat{\Sigma} - \Sigma)y$$

where the last step was achieved by typical rearrangement argument for ϵ-net.

$$\begin{aligned} x^T(\hat{\Sigma}-\Sigma)y &= x^T\left(\frac{1}{n}\sum X_iX_i^T - \mathbb{E}[XX^T]\right)y \\ &= \frac{1}{n}\sum_{I=1}^{n}\left((x^TX_i)(y^TX_i) - \mathbb{E}[(x^TX_i)(y^TX_i)]\right) \end{aligned}$$

We have $x^TX_i \sim N(0, x^T\Sigma x)$ and $y^TX_i \sim N(0, y^T\Sigma y)$. We have

$$\begin{aligned} \left((x^TX_i)(y^TX_i) - \mathbb{E}[(x^TX_i)(y^TX_i)]\right) &= \frac{((x+y)^TX_i)^2 - ((x-y)^TX_i)^2}{4} \\ &\quad - \frac{\mathbb{E}[((x+y)^TX_i)^2] - \mathbb{E}[((x-y)^TX_i)^2]}{4} \\ &= \frac{(x+y)^T\Sigma(x+y)}{4}(\psi^2(x,y)-1) \\ &\quad + \frac{(x-y)^T\Sigma(x-y)}{4}(\zeta^2(x,y)-1) \end{aligned}$$

where $\psi^2(x,y) \sim \chi_1^2$ and $\zeta^2(x,y) \sim \chi_1^2$. Therefore, we get

$$x^T(\hat{\Sigma}-\Sigma)y \sim \frac{(x+y)^T\Sigma(x+y)}{4}\left(\frac{1}{n}\chi_n^2\right) + \frac{(x-y)^T\Sigma(x-y)}{4}\left(\frac{1}{n}\chi_n^2\right)$$

which is less than

$$\frac{\lambda_{max}(\Sigma)}{n}\max_{x,y}|\sum_{i=1}^{n}(\psi^2(x,y)-1)| + \frac{\lambda_{max}(\Sigma)}{n}\max_{x,y}|\sum_{i=1}^{n}(\zeta^2(x,y)-1)|$$

Therefore, we get

$$||\hat{\Sigma}-\Sigma||_{op} \lesssim \frac{1+\theta}{n}\left(\sqrt{nd\log(1/\delta)} + d\log(1/\delta)\right)$$

with probability at least $1-\delta$. Thus,

$$|\sin(v,\hat{v})| \lesssim \frac{1+\theta}{\theta}\sqrt{\frac{d}{n}}$$

where d is much less than n.

§3 Sparse PCA

Sparse spiked covariance model is $X_1, \dots X_n \sim N(0,\Sigma)$ and $\Sigma = I_d + \theta vv^T$ with $|v|_2 = 1$ and $|v|_0 = k < d$. We find

$$\max_{x\in B_2, |x|_0\le k} x^T\hat{\Sigma}x$$

We can apply Cauchy to get $|x|_1 \le \sqrt{k}$ so we do L_1 relaxation

$$\max_{x\in B_2, |x|_1\le\sqrt{k}} x^T\hat{\Sigma}x$$

We cannot apply Davis-Kahan directly!! We get something like

$$\langle \hat{A} - A, \hat{u}\hat{u}^T - uu^T\rangle$$

Note that $\hat{u}\hat{u}^T - uu^T$ is sparse, in that it basically has a $2k$ by $2k$ matrix and all else are 0s. We can rewrite as

$$\langle \hat{\Sigma} - \Sigma, \hat{v}\hat{v}^T - vv^T \rangle = \langle \hat{\Sigma}_S - \Sigma_S, \hat{v}\hat{v}^T - vv^T \rangle \leq ||\hat{\Sigma}_S - \Sigma_S||_{op} ||\hat{v}\hat{v}^T - vv^T||_1$$

so this is on the scale of $\sqrt{2k/n}$, but since we also take max over all supports, we pick up another factor of $\sqrt{\log(d/2k)/n}$ and we are done.