LECTURE 30

Last class: matrix concentration inequalities. We upgraded

**(Scalar) Koeffding Inequality**: let $a_i$ be numbers →

$$P \left( \left| \sum_{i=1}^{n} \varepsilon_i a_i \right| > t \right) \leq 2 \exp \left( -\frac{t^2}{2\sigma^2} \right), \quad \text{where} \quad \sigma^2 = \sum_{i=1}^{n} a_i^2$$

**Matrix Koeffding Inequality (MKI)**: let $A_i$ be $d \times d$ symmetric matrices →

$$P \left( \left\| \sum_{i=1}^{n} \varepsilon_i A_i \right\| > t \right) \leq 2d \cdot \exp \left( -\frac{t^2}{2\sigma^2} \right) \quad \text{where} \quad \sigma^2 = \left\| \sum_{i=1}^{n} A_i \right\|$$

• The cost of the upgrade = $d$. Is it big? Curse of K.D.?

$$P \left( \left\| S_n \right\| > t \right) \leq 2 \exp \left( \log d - \frac{t^2}{2\sigma^2} \right) = \frac{2}{d} \ll 1 \quad \text{in K.D.}$$

choose $t = 2\sigma \sqrt{\log d}$

$$\Rightarrow \quad \left\| S_n \right\| \leq 2\sigma \sqrt{\log d} \quad \text{with high probability}$$

The cost is logarithmic 🙃

• Expectation = ? Recall the integral identity for a r-variable $Z_0$:

$$E X = \int P\{ X \geq s \} ds$$

$$E \left\| S_n \right\| \leq \left( E \left\| S_n \right\|^2 \right)^{\frac{1}{2}} \quad (L_1 \text{ norm} \leq L_2 \text{ norm}). \quad \text{(*)}$$

$$E \left\| S_n \right\|^2 = \int P\{ \left\| S_n \right\| > s \} ds \quad \leq \int 2d \cdot \exp \left( -\frac{s}{2\sigma^2} \right) ds = 4\sigma^2 d$$

too large! not logarithmic 😞
• The loss occurs for small $t$.

For example, for $t = 0$, M.H.1 gives $P\{ \cdot \} \leq 2d$.

Useless! The trivial bound $P\{ -3 \leq 1 \}$ is better.

$\Rightarrow$ Use the trivial bound (1) for small $t$, for simplicity, let $\alpha = 1$.

M.H.1's bound for large $t$.

$E \| S_n \|^2 \leq \int_0^{s_0} P\{ \| S_n \|^2 \geq s \} \, ds + \int_{s_0}^\infty P\{ \| S_n \|^2 \geq s \} \, ds$

$\leq s_0 + 4d \exp(-s_0/2)$. Choose $s_0 = 2 \log d =)

\leq 2 \log d + 4 \leq 8 \log d \quad (d \geq 2)$

For general $\sigma$:

$E \| S_n \|^2 \leq 8 \sigma \log d \quad \text{(DIY)}$

Def. of $\sigma$ in M.H.1

Cor. For $d \times d$ symmetric matrices $A_i$,

$E \| \sum_{i=1}^n \xi_i A_i \|^2 \leq 8 \log (d) \| \sum_{i=1}^n A_i^2 \|$

• This is a matrix version of the scalar identity

$E \left( \sum_{i=1}^n \xi_i a_i \right)^2 = \sum_{i=1}^n a_i^2$ \quad $\forall a_i \in \mathbb{R}$

$\text{var (sum)} = \text{sum. (var)}$

• \((*)\) p.1 $\Rightarrow$

$E \| \mathbf{Z} \xi A_i \|^2 \leq 3 \sqrt{\log d} \| \sum_{i=1}^n A_i^2 \|^{1/2}$. 

-2 -
• Similarly, we can upgrade Chernoff inequality (HW) and Bernstein’s imp. from Oct. 3, 6

**Matrix Bernstein Inequality (MBI)**

Let $X_i$ be independent, mean zero, symmetric $d x d$ random matrices. Assume $\|X_i\| \leq 1$ with prob. 1. Then if $t \geq 2$:

$$P \left\{ \| \sum_{i=1}^{\tilde{\nu}} X_i \| \geq t \right\} \leq 2 d \cdot \exp \left[ -c \left( \frac{t^2}{\sigma^2} + \frac{t}{k} \right) \right]$$

where $\sigma^2 = \| \sum_{i=1}^{\tilde{\nu}} EX_i^2 \|$. 

**“matrix variance” absolute constant $> 0$ (see Thm 5.4.1 in book)**

- If we choose $t = C \sigma \sqrt{\log d} + CK \log d$ with a sufficiently large absolute constant $C$, we get $\| \sum_{i=1}^{\tilde{\nu}} X_i \| \leq \sqrt{\log d} \leq 1$. 

  $\Rightarrow \| \frac{\tilde{\nu}}{t} X_i \| \leq C \sigma \sqrt{\log d} + CK \log d$ with high probability $\heartsuit$

- Also, in expectation (argue similarly to p.1):

  $\cos \left( \mathbb{E} \| \sum_{i=1}^{\tilde{\nu}} X_i \| \right) \leq C \sigma \sqrt{\log d} + K \log d$. (DIY)

**A VERY USEFUL TOOL. No assumptions on random matrices $X_i$ — no independence of entries!**

**APPLICATION : COMMUNITY DETECTION IN NETWORKS**

lec.8 (Sep.14): a network is an (undirected) graph

$G = (V, E)$

• Examples:
Internet network

opte.org
Twitter network
Problem: find communities (clusters).

lec 8 (Sep 19): a mathematical model of a complex network = random graph.

- Erdős-Rényi model $G(n,p)$:
  - $n$ vertices; connect each pair $i \neq j$ prob. $p$ independently
  - $n=200$, $p=\frac{1}{40}$

  Pro of $G(n,p)$: elegant, simple model
  Con: too simple. No communities

How to model communities?

Answer: allow different probabilities of edges

$n/2$ BOYS $n/2$ GIRLS

| $p$ | $q \leq p$ | $p$ |

| $G(n, p, q)$ |

= Stochastic block model
COMMUNITY DETECTION PROBLEM:

Given a random graph drawn from a SBM, find the 2 communities.

- Unsupervised learning
- For what \( p,q \) is this possible? and how?

(eg for \( p=q \): impossible.)

Example: Adj. matrix of a SBM →

\[
\text{Adj} = \begin{pmatrix}
\_ & \_ & \_ & \_ & \_ & \_ & \_ \\
\_ & \_ & \_ & \_ & \_ & \_ & \_ \\
\_ & \_ & \_ & \_ & \_ & \_ & \_ \\
\_ & \_ & \_ & \_ & \_ & \_ & \_ \\
\_ & \_ & \_ & \_ & \_ & \_ & \_ \\
\_ & \_ & \_ & \_ & \_ & \_ & \_ \\
\_ & \_ & \_ & \_ & \_ & \_ & \_ \\
\end{pmatrix}
\]

Example: \( G(n, p, q) \) with \( n = 200, p = \frac{1}{20}, q = \frac{1}{200} \).

Key to success: random matrix theory

Def: The adjacency matrix \( A \) of a graph \( G = (V, E) \) is the \( n \times n \) matrix with entries:

\[
A_{ij} = \begin{cases} 
1 & \text{if } (i,j) \in E \\
0 & \text{if not}
\end{cases}
\]