Lecture 12: Matrix Concentration.

Q: We receive iid samples of a centered random vector
X: $X^{(1)}, X^{(2)}, \ldots, X^{(n)} \in \mathbb{R}^d$; $E[X]=0$.

We cannot assume anything about the dependency structure of the coordinates of the same sample.

Goal:- Estimate the covariance matrix $E[XX^T]$, and in particular to study the estimator

$$\hat{\Sigma} := \frac{1}{n}\sum_{i=1}^{n} \underbrace{X^{(i)}X^{(i)T}}_{A^{(i)}}.$$

$A^{(i)}$ = random matrix $\in \mathbb{R}^{d\times d}$
entries are
not independent.

Assumption:- $\|X\| \leq K\sqrt{E[\|X\|^2]}$ for some constant K.

The vectors are absolutely bounded, but our parameter is not the bound intelf but the ratio between the bound and the 'standard deviation' [e.g. K does not change when X is scaled].

Setting is

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} A^{(i)}$$

where $A^{(i)}$ are independent random matrices.

$$E[A^{(i)}] = E\left[X^{(i)} X^{(i)T}\right] = \Sigma. \quad \text{— (1)}$$

$$A^{(i)^2} = X^{(i)} X^{(i)T} X^{(i)} X^{(i)T} = \|X^{(i)}\|^2 X^{(i)} X^{(i)T}$$

$$\precsim K^2 E[\|X\|^2] \, X^{(i)} X^{(i)T}$$

$$\Rightarrow \quad E[A^{(i)^2}] \precsim K^2 E(\|X\|^2) \Sigma. \quad \text{— (2)}$$

But what we would want to consider.

$$B^{(i)} := A^{(i)} - \Sigma \qquad \text{which are mean } 0 \text{ (by (1))}.$$

$$0 \preccurlyeq E[B^{(i)^2}] = E\left[A^{(i)^2}\right] - \Sigma^2 \preccurlyeq E[A^{(i)^2}]$$

$$\Rightarrow \quad 0 \preccurlyeq \sum_{i=1}^{n} E[B^{(i)^2}] \preccurlyeq \sum_{i=1}^{n} E[A^{(i)^2}].$$

$$\left\|\sum_{i=1}^{n} E[A^{(i)^2}]\right\| \leqslant n K^2 E[\|X\|^2] \, \|\Sigma\|$$

$$\Sigma = E[XX^T]$$

$$v^T \Sigma v = E[|v^T x|^2].$$

$$\leqslant E[\|v\|^2 \|x\|^2] \leqslant K^2 \|v\|^2 E[\|X\|^2].$$

$$\Rightarrow \quad \|\Sigma\| \leqslant K^2 E[\|X\|^2]$$

$$\Rightarrow \quad \left\|\sum_{i=1}^{n} E[A^{(i)^2}]\right\| \leqslant n K^4 E[\|X\|^2]^2.$$

Same calculation also gives, by the monotonicity argument above:-

$$\left\|\frac{1}{n}\sum_{i=1}^{n} E[B^{(i)^2}]\right\| \leqslant \underbrace{K^4 E[\|X\|^2]^2}_{\sigma^2}.$$

$$\|B^{(i)}\| \leqslant \|A^{(i)}\| + \|\Sigma\|$$

$\leq \underbrace{2K^2 E[\|X\|^2]}_{s}.$

So, the setting for our concentration inequality would be.

(1) Independent symmetric matrices $M_1 \ldots\ldots M_n$, mean 0.

(2) Bounds on $\|M_i\|$: $\|M_i\| \leq s$ for every i..

(3) Bound on $\left\|\frac{1}{n}\sum_{i=1}^n E[M_i^2]\right\|$: $\left\|\frac{1}{n}\sum_{i=1}^n E[M_i^2]\right\| \leq \sigma^2$.

THEN

Matrix Bernstein (Ahlswede - Winter, Tropp).

$$\Pr\left[\left\|\frac{1}{n}\sum_{i=1}^n M_i\right\| \geq t\right] \leq 2d \exp\left(-\frac{n t^2/2}{\sigma^2 + s t/3}\right)$$

under the above conditions. $M \in \mathbb{R}^{d\times d}$

[Ex: Is this better than taking a union bound over the d^2 individual entries?]

Remark: What about asymmetric (even rectangular) matrices? Suppose $S_1 \ldots S_n$, $d_1 \times d_2$, not necessarily symmetric. Consider.

$$X_i = \begin{pmatrix} 0 & S_i \\ S_i^T & 0 \end{pmatrix} \in \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}. \quad X_i^T = X_i$$

Now can apply Bernstein to X_i after translation

When $S = UDV^*$
$d_1\times d_1,\ d_1\times d_1,\ d_1 \times d_2$
S not symmetric, $d_1 \times d_2$

$$X = \begin{pmatrix} 0 & UDV^+ \\ VDU^+ & 0 \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}\begin{pmatrix} 0 & D \\ D & 0 \end{pmatrix}\begin{pmatrix} U^* & 0 \\ 0 & V^* \end{pmatrix}$$

So, X has same eigenvalues as $\begin{pmatrix} 0 & D \\ D & 0 \end{pmatrix}$ where $D = \text{diag}$(singular values of S).

$\Downarrow$

Eigenvalues are singular values and their negatives.

MATRIX CHERNOFF :- Let's consider the λ_{max} (largest eigenvalue) first. Consider a random Hermitian matrix M.

$$\Pr[\lambda_{max}(M) \geq t] = \Pr[\exp(\alpha\lambda_{max}(M)) \geq \exp(\alpha t)]$$
$$\forall \alpha > 0.$$

$$\leq \underbrace{E[\exp(\alpha\lambda_{max}(M))]} \exp(-\alpha t).$$
(Markov's inequality).

$$\exp(\alpha\lambda_{max}(M)) = \lambda_{max}(\exp(\alpha M)) \quad \Big| \begin{array}{l} M \text{ Hermitian} \\ \alpha > 0. \end{array}$$
$$\leq Tr(\exp(\alpha M))$$
$$\because \exp(\alpha M) \succeq 0 \quad \forall \ M \text{ Hermitian}.$$

Define: $\Psi_M(\alpha) := E[\exp(\alpha M)]$

$$\Pr[\lambda_{max}(M) \geq t] \leq \exp(-\alpha t) E[Tr(\exp(\alpha M))] \ \forall \alpha > 0$$
$$= \exp(-\alpha t) \cdot Tr\, E[\exp(\alpha M)] \ \forall \alpha > 0$$
$$= \exp(-\alpha t)\, Tr\, \Psi_M(\alpha) \ \forall \alpha > 0 \qquad (*).$$

Goal :- (1) 'Bound' $\Psi_M(\alpha)$ for M satisfying (2),(3) above.

(2) 'Bound' $\Psi_{\Sigma M_i}(\alpha)$ for M_i independent, each satisfying (2),(3) above.