MATHEMATICS OF DATA SCIENCE

December 5, 2023

Today: Johnson-Lindenstrauss lemma (continued)

Next session:
- Random matrices & covariance estimation
- Update on homework assignment (not due before end of January at the ear

JOHNSON-LINDENSTRAUSS (J-L) LEMMA :

Goal

Given $\{x_1, \dots, x_N\}$ N points in $\mathbb{R}^n$, find m such that

if $P = \frac{1}{\sqrt{m}} \begin{bmatrix} y_1^T \\ \vdots \\ y_m^T \end{bmatrix} \in \mathbb{R}^{m \times n}$ is formed by m random

vectors $y_1, \dots, y_m$ in $\mathbb{R}^n$ such that $\{y_i\}$'s are independent,

zero mean and isotropic ($\mathbb{E}[y_i y_i^T] = I_n \ \forall i$)

then $\mathbb{P}\left((1-\varepsilon)\|x_i - x_j\| \leq \|Px_i - Px_j\| \leq (1+\varepsilon)\|x_i - x_j\| \ \forall (i,j) \right)$ (1)

≥ 0.99

where $\varepsilon \in (0,1)$

↳ Interest : If we can guarantee (1) with $m < n$, this means we can project $\{x_1, \dots, x_N\}$ from $\mathbb{R}^n$ onto a subspace of $\mathbb{R}^n$ of dimension m and preserve the distances between points up to a tolerance ε

Proof of the JL lemma

① Reformulate (1) into a concentration inequality

$$\mathbb{P}\left((1-\varepsilon)\|x_i - x_j\| \leq \|Px_i - Px_j\| \leq (1+\varepsilon)\|x_i - x_j\| \ \forall (i,j) \in \{1, \dots, n\}^2 \right) \geq 0.99$$

always be true when $x_i = x_j$

$$\iff \mathbb{P}\left((1-\varepsilon) \leq \|P_z\| \leq (1+\varepsilon) \quad \forall z \in Z \right) \geq 0.99$$

where $Z = \left\{ \frac{x_i - x_j}{\|x_i - x_j\|} \;\middle|\; x_i \neq x_j \right\}$

$$P_z = \underbrace{\underbrace{\frac{1}{\sqrt{m}} \begin{bmatrix} y_1^T \\ \vdots \\ y_m^T \end{bmatrix}}_{\in \mathbb{R}^{m \times n}} \underbrace{z}_{\in \mathbb{R}^{m \times 1}}}_{\in \mathbb{R}^{m \times 1}} = \begin{bmatrix} \frac{1}{\sqrt{m}} y_1^T z \\ \vdots \\ \frac{1}{\sqrt{m}} y_m^T z \end{bmatrix}$$

$$\|P_z\| = \sqrt{\sum_{i=1}^{m} \left(\frac{1}{\sqrt{m}} y_i^T z \right)^2} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i^T z)^2}$$

$$\|P_z\|^2 = \frac{1}{m} \sum_{i=1}^{m} (y_i^T z)^2$$

$$(1-\varepsilon) \leq \|P_z\| \leq 1+\varepsilon$$

If $1-\varepsilon \leq \|P_z\|^2 \leq 1+\varepsilon$,

then using $\varepsilon \in (0,1)$, $\sqrt{1-\varepsilon} \leq \|P_z\| \leq \sqrt{1+\varepsilon}$

$$\sqrt{1-\varepsilon} \in (0,1) \iff \sqrt{1-\varepsilon} \geq 1-\varepsilon$$

$$\sqrt{1+\varepsilon} > 1 \iff \sqrt{1+\varepsilon} \leq 1+\varepsilon$$

$\forall z \in Z$, $\quad (1-\varepsilon) \leq \|P_z\|^2 \leq 1+\varepsilon \implies 1-\varepsilon \leq \|P_z\| \leq 1+\varepsilon$

$$\underline{\mathbb{P}\left((1-\varepsilon) \leq \|P_z\|^2 \leq 1+\varepsilon \;\; \forall z \in Z \right) \geq 0.99} \implies \mathbb{P}\left((1-\varepsilon) \leq \|P_z\| \leq 1+\varepsilon \;\; \forall z \in Z \right) \geq 0.99$$

(2)

Using the expression for $\|P_z\|^2$,

$$(2) \iff \mathbb{P}\left(1-\varepsilon \leq \frac{1}{m}\sum_{i=1}^{m} (y_i^T z)^2 \leq 1+\varepsilon \quad \forall z \in Z \right) \geq 0.99$$

$$\iff \mathbb{P}\Big(-\varepsilon \leq \underbrace{\frac{1}{m}\sum_{i=1}^{m}\left[(y_i^T z)^2 - 1\right]} \leq \varepsilon \quad \forall z \in Z \Big) \geq 0.99$$

$$\left(\frac{1}{m}\sum_{i=1}^{m}(y_i^T z)^2\right) - 1 = \frac{1}{m}\sum_{i=1}^{m}(y_i^T z)^2 - \frac{1}{m}\sum_{i=1}^{m} 1$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left((y_i^T z)^2 - 1\right)$$

$$\iff \mathbb{P}\left(\left| \frac{1}{m}\sum_{i=1}^{m}\left[(y_i^T z)^2 - 1\right] \right| \leq \varepsilon \quad \forall z \in Z \right) \geq 0.99$$

$$\iff \mathbb{P}\left(\max_{z \in Z} \left| \frac{1}{m}\sum_{i=1}^{m}\left[(y_i^T z)^2 - 1\right] \right| \leq \varepsilon \right) \geq 0.99$$

$$\iff 1 - \mathbb{P}\left(\max_{z \in Z} \left| \frac{1}{m}\sum_{i=1}^{m}\left[(y_i^T z)^2 - 1\right] \right| > \varepsilon \right) \geq 0.99$$

Union bound

$$\Longleftarrow \underline{1 - \sum_{z \in Z} \mathbb{P}\left(\left| \frac{1}{m}\sum_{i=1}^{m}\left[(y_i^T z)^2 - 1\right] \right| > \varepsilon \right) \geq 0.99}$$

(3)

probability that can be bounded using a concentration inequality

② Finding the right concentration inequality

Fix $z \in Z$. We wish to bound $\mathbb{P}\left(\left| \frac{1}{m}\sum_{i=1}^{m}\left[(y_i^T z)^2 - 1\right] \right| > \varepsilon \right)$

$y_1, \dots, y_m$ are independent, zero mean, isotropic and subgaussian random vectors

- $y_1, \dots, y_m$ independent $\Rightarrow (y_1^T z)^2 - 1, \dots, (y_m^T z)^2 - 1$ are independent

- $\forall i = 1..m$,

$$\mathbb{E}[(y_i^T z)^2] = \mathbb{E}[(y_i^T z)(y_i^T z)]$$

$a^T b = b^T a$

$$= \mathbb{E}[\underbrace{(z^T y_i)}_{\in \mathbb{R}} \underbrace{(y_i^T z)}_{\in \mathbb{R}}]$$

$$= \mathbb{E}[z^T y_i y_i^T z]$$

$$= \mathbb{E}[\underbrace{z^T}_{\in \mathbb{R}^{1\times m}} \underbrace{(y_i y_i^T)}_{\in \mathbb{R}^{m\times m}} \underbrace{z}_{\in \mathbb{R}^{m\times 1}}]$$

$$= z^T \mathbb{E}[y_i y_i^T] z$$

y_i isotropic
$\mathbb{E}[y_i y_i^T] = I_m$

$I_m z = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} z = z$

$$= z^T I_m z = z^T z = \underbrace{\|z\|^2 = 1}_{\text{by definition of } \mathcal{Z}}$$

We have shown

$$\mathbb{E}[(y_i^T z)^2] = 1$$

$$\Leftrightarrow \mathbb{E}[(y_i^T z)^2 - 1] = 0$$

$(y_i^T z)^2 - 1$ is a zero mean variable

- $y_1, \dots, y_m$ are subgaussian vectors

$\Rightarrow y_1^T z, \dots, y_m^T z$ are subgaussian random variables

$\Leftrightarrow (y_1^T z)^2, \dots, (y_m^T z)^2$ are subexponential random variables

$\Leftrightarrow (y_1^T z)^2 - 1, \dots, (y_m^T z)^2 - 1$ are subexponential random variables

Overall, $(y_1^T z)^2 - 1, \ldots, (y_m^T z)^2 - 1$ are independent, zero mean and subexponential random variables

Therefore, we can apply Bernstein's inequality (see the corollary from the past lecture):

Bernstein's
$$\left[\ \mathbb{P}\left(\left| \frac{1}{m} \sum_{i=1}^{m} \left[(y_i^T z)^2 - 1 \right] \right| \geq \varepsilon \right) \leq 2\exp\left(-\hat{c} \min(\varepsilon, \varepsilon^2) m\right) \right.$$

for some $\hat{c} > 0$

Since $\varepsilon \in (0,1)$, $\min(\varepsilon, \varepsilon^2) = \varepsilon^2$, we have

$$\mathbb{P}\left(\left| \frac{1}{m} \sum_{i=1}^{m} \left[(y_i^T z)^2 - 1 \right] \right| > \varepsilon \right)$$

$$\leq \mathbb{P}\left(\left| \frac{1}{m} \sum_{i=1}^{m} \left[(y_i^T z)^2 - 1 \right] \right| \geq \varepsilon \right) \leq 2\exp\left(-\hat{c}\, \varepsilon^2 m\right)$$

③ Combine the concentration inequality with (3)

$$(3) \iff 1 - \sum_{z \in Z} \mathbb{P}\left(\left| \frac{1}{m} \sum_{i=1}^{m} \left[(y_i^T z)^2 - 1 \right] \right| > \varepsilon \right) \geq 0.99$$

$$\iff \sum_{z \in Z} \mathbb{P}\left(\left| \frac{1}{m} \sum_{i=1}^{m} \left[(y_i^T z)^2 - 1 \right] \right| > \varepsilon \right) \leq 0.01$$

(4)

By concentration, for any $z \in Z$

$$\mathbb{P}\left(\left| \frac{1}{m} \sum_{i=1}^{m} \left[(y_i^T z)^2 - 1 \right] \right| > \varepsilon \right) \leq 2\exp\left(-\hat{c}\, \varepsilon^2 m\right)$$

$$\Rightarrow \sum_{z\in Z} \mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m}\left[(y_i^T z)^2-1\right]\right|>\varepsilon\right) \le \sum_{z\in Z} 2\exp(-\hat{c}\,\varepsilon^2 m)$$

$$= |Z| \times 2\exp(-\hat{c}\,\varepsilon^2 m)$$

$$|Z| = \{(x_i,x_j) \mid x_i \ne x_j,\ (i,j)\in\{1,\dots,N\}^2\} \le N^2 \qquad \le 2N^2\exp(-\hat{c}\,\varepsilon^2 m)$$

To satisfy $2N^2\exp(-\hat{c}\,\varepsilon^2 m) \le 0.01$,

we need $-\hat{c}\,\varepsilon^2 m \le \ln\left(\frac{0.01}{2N^2}\right)$

$$\Leftrightarrow \quad m \ge \frac{1}{\hat{c}\,\varepsilon^2} \times \left(-\ln\left(\frac{0.01}{2N^2}\right)\right)$$

$$= \frac{1}{\hat{c}\,\varepsilon^2}\ln\left(\frac{2N^2}{0.01}\right)$$

$$= \frac{2}{\hat{c}}\ln\left(\sqrt{\frac{2}{0.01}}\,N\right)\varepsilon^{-2}$$

Conclusion: If $m \ge \frac{2}{\hat{c}}\ln\left(\sqrt{\frac{2}{0.01}}\,N\right)\varepsilon^{-2}$,

then (4) holds

$\Leftrightarrow$ (3) holds

$\Rightarrow$ (2) holds

$\Rightarrow$ (1) holds, i.e.

$$\mathbb{P}\left((1-\varepsilon)\|x_i-x_j\| \le \|Px_i - Px_j\| \le (1+\varepsilon)\|x_i-x_j\| \ \forall\, i,j\right) \ge 0.99$$

Remark: The bound on m does not depend on n, and thus $m < n$ for n sufficiently large

Application: Linear least squares

$A \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^d$, seek $x \in \mathbb{R}^d$ such that

$\|Ax-b\| \simeq 0$

$\Rightarrow$ Not possible in general when $n >> d$

$\Rightarrow$ Instead we solve

$$\underset{x\in\mathbb{R}^d}{\text{minimize}} \; \|Ax-b\|^2$$

$$\hookrightarrow \begin{cases} x_{opt} \in \underset{x\in\mathbb{R}^d}{\text{argmin}} \|Ax-b\|^2 \\ f_{opt} = \|Ax_{opt}-b\|^2 \end{cases}$$

$\hookrightarrow$ Suppose that we apply a random projection matrix

$P = \frac{1}{\sqrt{m}} \begin{bmatrix} y_1^T \\ \vdots \\ y_m^T \end{bmatrix} \in \mathbb{R}^{m\times n}$ to the problem and we

solve

$$\underset{x\in\mathbb{R}^d}{\text{minimize}} \; \|PAx-Pb\|^2$$

$$\hookrightarrow \tilde{x}_{opt} \in \underset{x\in\mathbb{R}^d}{\text{argmin}} \|PAx-Pb\|^2$$

$$\hat{f}_{opt} = \|PA\tilde{x}_{opt}-Pb\|^2$$

Using J-L lemma, we can prove that

if $m \geqslant C\left(\frac{d}{\varepsilon} + d\ln(d)\right)$, then

$C > 0$

independent of n!

$$\mathbb{P}\left(\|Ax_{opt} - b\|^2 \leq \|A\tilde{x}_{opt} - b\|^2 \leq (1+\varepsilon)\|Ax_{opt} - b\|^2\right)$$

$\geqslant 0.99$