DS-563/CS-543 Lecture 11/2023-02-28

Today: Johnson-Linderstrauss Lemma

Goal: ~~Take~~ Map high-dimensional data to low dimension, while preserving its "properties"

Theorem (JL Lemma)

$S =$ set of n points in $\mathbb{R}^k$, $\varepsilon \in (0, 1/2)$

$d = \left\lceil \frac{24 \ln n}{\varepsilon^2} \right\rceil$

There is a linear embedding $f: \mathbb{R}^k \to \mathbb{R}^d$ s.t. for all $u, v \in S$:

$$(1-\varepsilon)\|u-v\|^2 \leq \|f(u)-f(v)\| \leq (1+\varepsilon)\|u-v\|^2$$

Informal/notes:

- Can map n points into $O(\log n)$ dimensions, while ~~preservr~~ preserving their distances.
- Allows for more efficient processing of data in lower dimension: useful for some nearest neighbor search algorithms.

Our construction

$N(\alpha, \sigma^2)$ = Gaussian distribution with mean α and variance σ^2

$$f(x) = \frac{1}{\sqrt{d}} \begin{bmatrix} \text{each entry independent} \\ \text{from } N(0,1) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$$

$f(x)$: d rows, 1 column

matrix: d rows, k columns

x: k rows, 1 column

Lemma 1

For any $x \in \mathbb{R}^k$,

$$(1-\varepsilon)\|x\|^2 \leq \|f(x)\|^2 \leq (1+\varepsilon)\|x\|^2$$

with probability $\geq 1 - 2e^{-\frac{\varepsilon^2 d}{12}}$.

Proof: Lemma 1 $\Rightarrow$ JL Lemma

Any $u, v \in S$. Let $x = u - v$.

Since f linear, $f(u) - f(v) = f(u-v) = f(x)$.

By Lemma 1, w.p. $\geq 1 - 2e^{-\frac{\varepsilon^2 d}{12}}$,

$$(1-\varepsilon)\|x\|^2 \leq \|f(x)\|^2 \leq (1+\varepsilon)\|x\|^2$$

$$(1-\varepsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2$$

Via union bound, this holds for all pairs $u, v \in S$, w.p. $\geq 1 - \binom{n}{2} \cdot 2e^{-\frac{\varepsilon^2 d}{12}} > 1 - n^2 e^{-\frac{\varepsilon^2 d}{12}}$

$$\geq 1 - n^2 e^{-\left(\frac{\varepsilon^2}{12} \cdot \frac{24 \ln n}{\varepsilon^2}\right)} = 1 - n^2 \cdot \underbrace{e^{-2\ln n}}_{=1/n^2} = 0.$$

So with positive probability f has the desired properties. ∎

[Still have to prove Lemma 1]

Review of properties of the Gaussian distribution

Notation: $X \sim D \equiv$ random variable X distributed according to D

Properties:

- X, Y independent & $X \sim \mathcal{N}(\alpha, \sigma_1^2)$ & $Y \sim \mathcal{N}(\beta, \sigma_2^2)$
$$\Rightarrow X + Y \sim \mathcal{N}(\alpha + \beta, \sigma_1^2 + \sigma_2^2)$$
- $X \sim \mathcal{N}(0, \sigma^2) \Rightarrow \underset{\in \mathbb{R}}{\alpha} X \sim \mathcal{N}(0, \alpha^2 \sigma^2)$
- $X \sim \mathcal{N}(0, \sigma^2) \Rightarrow \mathbb{E}[X^2] = \sigma^2$
- $\lambda < 1/2$ & $X \sim \mathcal{N}(0, \sigma^2) \Rightarrow \mathbb{E}[e^{\lambda X^2}] = (1 - 2\lambda)^{-1/2}$

Useful properties

Recall $f(x) = \frac{1}{\sqrt{d}} A x$

$\uparrow$

d row × k column matrix

A_i = i-th row of A

$A_{i,j}$ = j-th entry in A_i

$$f(x) = \frac{1}{\sqrt{d}} \begin{bmatrix} A_1 x \\ A_2 x \\ \vdots \\ A_d x \end{bmatrix} = \frac{1}{\sqrt{d}} \begin{bmatrix} \sum_{j=1}^{k} A_{1,j} x_j \\ \vdots \\ \sum_{j=1}^{k} A_{d,j} x_j \end{bmatrix}$$

<u>Claim</u>: For any $i \in [d]$, $A_i x \sim \mathcal{N}(0, \|x\|^2)$.

<u>Proof</u>: Each $A_{i,j} \sim \mathcal{N}(0,1) \Rightarrow A_{i,j} x_j \sim \mathcal{N}(0, x_j^2)$.

$A_{i,j}$'s independent $\Rightarrow A_i x = \sum_{j=1}^{k} A_{i,j} x_j \sim \mathcal{N}(0, \sum_{j=1}^{k} x_j^2)$

$$\mathcal{N}(0, \sum_{j=1}^{k} x_j^2) \equiv \mathcal{N}(0, \|x\|^2)$$

$\blacksquare$

Lemma 2 (concentration bound)

$X_1, \dots, X_d$ - independent, distributed according to $\mathcal{N}(0,1)$

For each $i \in d$, let $Z_i = X_i^2$.

↖ chi-squared distribution

For any $\varepsilon \in (0,1)$,

$$\Pr\left[\left|d - \sum_{i=1}^{d} Z_i\right| > \varepsilon d\right] \leq 2e^{-\frac{\varepsilon^2 d}{12}}$$

Proof: Lemma 2 ⇒ Lemma 1

Want to show that with good probability:

$$(1-\varepsilon)\|x\|^2 \leq \left\|\frac{1}{\sqrt{d}} Ax\right\|^2 \leq (1+\varepsilon)\|x\|^2$$

$$\Big\Updownarrow \text{ multiply by } \frac{d}{\|x\|^2}$$

$$(1-\varepsilon) d \leq \left\|\frac{Ax}{\|x\|}\right\|^2 \leq (1+\varepsilon) d$$

$$\Big\Updownarrow$$

$$\left| d - \left\|\frac{Ax}{\|x\|}\right\|^2 \right| \leq \varepsilon d$$

Each entry $A_i x$ of Ax independent & $\sim N(0, \|x\|^2)$

⇒ Each entry of $\frac{Ax}{\|x\|}$ independent & $\sim \mathcal{N}(0,1)$.

$\boxed{11-5}$

$$\left\| \frac{Ax}{\|x\|} \right\|^2 = \text{sum of squares of } d \underbrace{\text{Gaussians}}_{\text{independent}} \mathcal{N}(0,1).$$

From Lemma 2:

$$\Pr\left[\left| d - \left\| \frac{Ax}{\|x\|} \right\|^2 \right| \leq \varepsilon d \right] \geq 1 - 2e^{-\frac{\varepsilon^2 d}{12}}$$

Proof of Lemma 2 (the concentration bound)

[note: similar to proof of Chernoff bound]

Part 1: $\sum_{i=1}^{d} Z_i$ is usually not too high

$\lambda \in (0,1) \leftarrow$ additional variable, set later

$$p_{HIGH} = \Pr\left[\sum_{i=1}^{d} Z_i > d(1+\varepsilon) \right] = \Pr\left[e^{\lambda \sum_i Z_i} > e^{d\lambda(1+\varepsilon)} \right]$$

Via Markov's inequality:

$$p_{HIGH} \leq \frac{\mathbb{E}\left[e^{\lambda \sum_i Z_i}\right]}{e^{d\lambda(1+\varepsilon)}} \underset{\text{independent } Z_i\text{'s}}{=} \frac{\prod_i \mathbb{E}\left[e^{\lambda Z_i}\right]}{e^{d\lambda(1+\varepsilon)}} \underset{\text{same distribution}}{=} \frac{\left(\mathbb{E}\left[e^{\lambda Z_1}\right]\right)^d}{e^{d\lambda(1+\varepsilon)}}$$

$$\underset{(*)}{=} \frac{\left((1-2\lambda)^{-1/2}\right)^d}{e^{d\lambda(1+\varepsilon)}} = \left((1-2\lambda)e^{2\lambda(1+\varepsilon)}\right)^{-d/2}$$

Set $\lambda = \frac{\varepsilon}{2(1+\varepsilon)} \in (0, 1/2)$, so ⊛ holds.

$$p_{HIGH} \leq \left(\left(1 - \frac{\varepsilon}{(1+\varepsilon)}\right) e^{\varepsilon} \right)^{-d/2} = \left((1+\varepsilon) e^{-\varepsilon} \right)^{d/2}$$

$$\boxed{\text{Via Taylor series: } \ln(1+t) \leq t - \frac{t^2}{2} + \frac{t^3}{3} \text{ for } t \in [0,1]}$$

$$\leq \left(e^{\varepsilon - \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3} - \varepsilon} \right)^{d/2} \leq \left(e^{-\frac{\varepsilon^2}{6}} \right)^{d/2} = e^{-\frac{\varepsilon^2 d}{12}}$$

<u>Part 2:</u> $\sum_{i=1}^{d}$ is usually not too low

[very similar proof]

$$p_{LOW} = \Pr\left[\sum_{i=1}^{d} Z_i < d(1-\varepsilon) \right] = \Pr\left[e^{-\lambda \sum Z_i} > e^{-d\lambda(1-\varepsilon)} \right]$$

$$\leq \frac{\mathbb{E}\left[e^{-\lambda \sum_i Z_i} \right]}{e^{-d\lambda(1-\varepsilon)}} = \frac{\prod_{i=1}^{d} \mathbb{E}\left[e^{-\lambda Z_i} \right]}{e^{-d\lambda(1-\varepsilon)}}$$

$$= \frac{\left((1+2\lambda)^{-1/2} \right)^{d}}{e^{-d\lambda(1-\varepsilon)}} = \left((1+2\lambda) e^{-2\lambda(1-\varepsilon)} \right)^{-d/2}$$

Set $\lambda = \frac{\varepsilon}{2(1-\varepsilon)} > 0$.

$$p_{LOW} \leq \left((1-\varepsilon) e^{\varepsilon} \right)^{d/2}$$

$$\boxed{\text{Via Taylor series: } \ln(1-t) \leq -t - \frac{t^2}{2} \text{ for } t \in [0,1]}$$

$$p_{LOW} \leq \left(e^{-\varepsilon - \frac{\varepsilon^2}{2} + \varepsilon} \right)^{d/2} = e^{-\frac{\varepsilon^2 d}{4}} \leq e^{-\frac{\varepsilon^2 d}{12}}$$

<u>Overall</u>: $\sum_i z_i$ neither too high nor too low with probability $1 - p_{HIGH} - p_{LOW} \geq 1 - 2e^{-\frac{\varepsilon^2 d}{12}}$.

Notes

- We showed existence. Increase the dimension to get ~~a~~ the claim with high probability.
- Used Gaussians. But $\{-1, +1\}$ works as well.
- Dense matrix. Can be replaced by a sparse matrix. This leads to a faster transformation.

- k-means / k-median clustering:
  - ~~O(log~~ $O\left(\frac{1}{\varepsilon^2} \log n\right)$ dimensions to preserve all pairwise distances
  - $O\left(\frac{1}{\varepsilon^2} \log (k/\varepsilon)\right)$ suffices to preserve the cost of optimal k-means / k-median clustering