Johnson - Lindenstrauss Lemma

Aniruddhan Ganesaraman, Rohan Shinde,
Sampurna Mondal

Large Sample Theory Project

24 April 2023

Overview

Overview

Motivation

- Most data (text, images, etc.) are high dimensional, which makes algorithms working on them very slow. JL Lemma is a classic (1984) "structure - preserving" dimension reduction result.

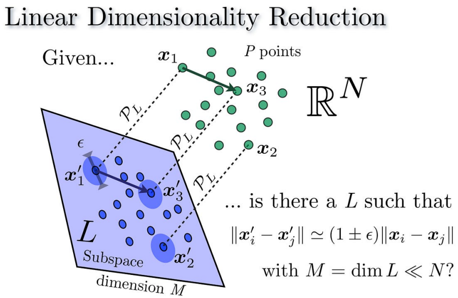
Motivation

- Most data (text, images, etc.) are high dimensional, which makes algorithms working on them very slow. JL Lemma is a classic (1984) “structure - preserving” dimension reduction result.
- It has its applications in applications in compressed sensing, manifold learning, dimensionality reduction, and graph embedding.

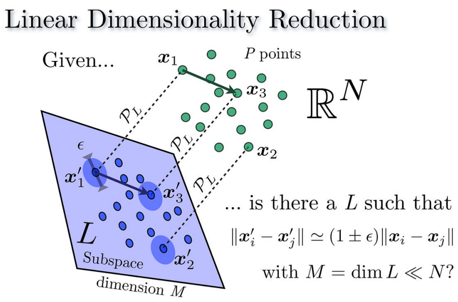
Motivation

- Most data (text, images, etc.) are high dimensional, which makes algorithms working on them very slow. JL Lemma is a classic (1984) "structure - preserving" dimension reduction result.
- It has its applications in applications in compressed sensing, manifold learning, dimensionality reduction, and graph embedding.
- **Idea:** A set of points in a high-dimensional space can be embedded into a space of much lower dimension in such a way that distances between the points are *nearly* preserved.

Immediate thoughts

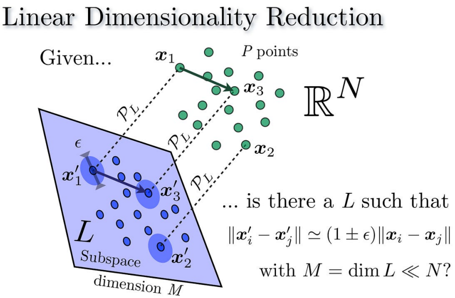


Immediate thoughts



Orthogonal projections reduce the average distance between points. JL Lemma deals with relative distances, which do not change under scaling.

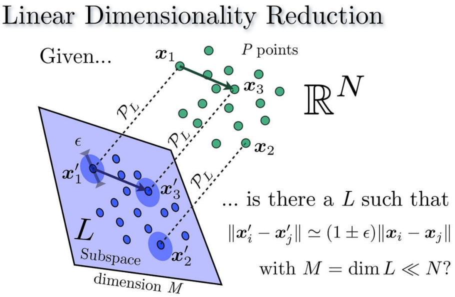
Immediate thoughts



Orthogonal projections reduce the average distance between points. JL Lemma deals with relative distances, which do not change under scaling.

Principal component analysis?

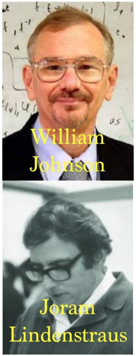
Immediate thoughts

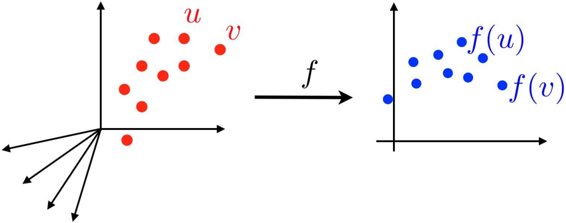


Orthogonal projections reduce the average distance between points. JL Lemma deals with relative distances, which do not change under scaling.

Principal component analysis? Speed and memory! (*reference*)

The improvement





The f so obtained is still linear (or Lipschitz).

JL Lemma

Theorem (1984)

Let $0 < \varepsilon < \frac{1}{2}$; *$Q \subset \mathbb{R}^d$ be a set of n points; and* $k = \frac{20 \log(n)}{\varepsilon^2}$. *There exists a Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in Q$,*

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

JL Lemma

Theorem (1984)

Let $0 < \varepsilon < \frac{1}{2}$; $Q \subset \mathbb{R}^d$ be a set of n points; and $k = \frac{20\log(n)}{\varepsilon^2}$. There exists a Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in Q$,

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

The dimension of the image space is only dependent on the error and the number of points. If the original dimension is very large, one can achieve significant dimension reduction.

Overview

Proof

Lemma (Norm preservation lemma)

Let $x \in \mathbb{R}^d$ *and* $A_{k \times d} = [[a_{ij}]]$ *where* $a_{ij} \overset{iid}{\sim} N(0,1)$. *Then*

$$\mathbb{P}\left(\underbrace{(1-\varepsilon)\|x\|^2 \leq \frac{1}{k}\|Ax\|^2 \leq (1+\varepsilon)\|x\|^2}_{(*)}\right) \geq 1 - 2e^{\frac{-(\varepsilon^2-\varepsilon^3)k}{4}}$$

Using "NP" Lemma

Let $f(x) = \frac{1}{\sqrt{k}} Ax$. By union bound over the $O(n^2)$ pairs of u and v,

$$\mathbb{P}(\exists u, v \text{ s.t. } (*)_{x=u-v} \text{ fails}) \leq \sum_{u,v} \mathbb{P}((*)_{x=u-v} \text{ fails})$$
$$\leq 2n^2 e^{\frac{-(\varepsilon^2-\varepsilon^3)k}{4}} < 1.$$

Using “NP” Lemma

Let $f(x) = \frac{1}{\sqrt{k}} Ax$. By union bound over the $O(n^2)$ pairs of u and v,

$$\mathbb{P}(\exists u, v \text{ s.t. } (*)_{x=u-v} \text{ fails}) \leq \sum_{u,v} \mathbb{P}((*)_{x=u-v} \text{ fails})$$

$$\leq 2n^2 e^{\frac{-(\varepsilon^2 - \varepsilon^3)k}{4}} < 1.$$

This completes the (deterministic probabilistic) proof, modulo NP lemma!

Preserving angles?

Corollary

If $\|u\|, \|v\| \leq 1$*, then* $\mathbb{P}(|\langle u, v\rangle - \langle f(u), f(v)\rangle| \geq \varepsilon) \leq 4e^{\frac{-(\varepsilon^2 - \varepsilon^3)k}{4}}$

Preserving angles?

Corollary

If $\|u\|, \|v\| \leq 1$, *then* $\mathbb{P}(|\langle u, v\rangle - \langle f(u), f(v)\rangle| \geq \varepsilon) \leq 4e^{\frac{-(\varepsilon^2-\varepsilon^3)k}{4}}$

Proof. With probability atleast $1 - 4e^{\frac{-(\varepsilon^2-\varepsilon^3)k}{4}}$,

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

Preserving angles?

Corollary

If $\|u\|, \|v\| \leq 1$, *then* $\mathbb{P}(|\langle u, v\rangle - \langle f(u), f(v)\rangle| \geq \varepsilon) \leq 4e^{\frac{-(\varepsilon^2-\varepsilon^3)k}{4}}$

Proof. With probability atleast $1 - 4e^{\frac{-(\varepsilon^2-\varepsilon^3)k}{4}}$,

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

But

$$\begin{aligned} 4\langle f(u), f(v)\rangle &= \|f(u+v)\|^2 - \|f(u-v)\|^2 \\ &\geq (1-\varepsilon)\|u+v\|^2 - (1+\varepsilon)\|u-v\|^2 \\ &= 4\langle u, v\rangle - 2\varepsilon\left(\|u\|^2 + \|v\|^2\right) \geq 4\langle u, v\rangle - 4\varepsilon. \end{aligned}$$

Similarly the other direction. ■

NP Lemma proof

For a fixed j,

$$\mathbb{E}\left[(Ax)_j^2\right] = \mathbb{E}\left[\left(\sum_i a_{ij}x_i\right)^2\right] = \mathbb{E}\left[\sum_{i,k} a_{ij}a_{kj}x_kx_i\right]$$
$$= \mathbb{E}\left[\sum_i a_{ii}^2x_i^2\right] = \sum_i x_i^2 = \|x\|^2.$$

NP Lemma proof

For a fixed j,

$$\mathbb{E}\left[(Ax)_j^2\right] = \mathbb{E}\left[\left(\sum_i a_{ij}x_i\right)^2\right] = \mathbb{E}\left[\sum_{i,k} a_{ij}a_{kj}x_kx_i\right]$$
$$= \mathbb{E}\left[\sum_i a_{ij}^2x_i^2\right] = \sum_i x_i^2 = \|x\|^2.$$

So,

$$\mathbb{E}\left[\frac{1}{k}\|Ax\|^2\right] = \frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[(Ax)_j^2\right] = \|x\|^2.$$

NP Lemma proof (contd.)

Note that $Y_j = \frac{(Ax)_j}{\|x\|} \overset{iid}{\sim} N(0,1)$. Also,

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{k}\|Ax\|^2 \geq (1+\varepsilon)\|x\|^2\right) &= \mathbb{P}\left(\sum_{j=1}^{k} Y_j^2 \geq (1+\varepsilon)k\right) \\
&= \mathbb{P}\left(\chi_k^2 \geq (1+\varepsilon)k\right)
\end{aligned}$$

A χ^2 concentration inequality

Lemma

$$\mathbb{P}(\chi_k^2 \geq (1+\varepsilon)k) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}} \quad \textit{and} \quad \mathbb{P}(\chi_k^2 \leq (1-\varepsilon)k) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}}$$

A χ^2 concentration inequality

Lemma

$$\mathbb{P}(\chi_k^2 \geq (1+\varepsilon)k) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}} \quad \textit{and} \quad \mathbb{P}(\chi_k^2 \leq (1-\varepsilon)k) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}}$$

Proof. Let $Z_1, \cdots, Z_k \overset{iid}{\sim} N(0,1)$. By Markov's inequality, for $0 < \lambda < \frac{1}{2}$,

$$\mathbb{P}(\chi_k^2 \geq (1+\varepsilon)k) = \mathbb{P}\left(\sum_{i=1}^{k} Z_i^2 \geq (1+\varepsilon)k\right)$$

$$\leq \frac{\mathbb{E}e^{\lambda \sum_{i=1}^{k} Z_i^2}}{e^{(1+\varepsilon)k\lambda}} = e^{-(1+\varepsilon)k\lambda}(1-2\lambda)^{-k/2}$$

A χ^2 concentration inequality

Lemma

$$\mathbb{P}(\chi_k^2 \geq (1+\varepsilon)k) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}} \quad \textit{and} \quad \mathbb{P}(\chi_k^2 \leq (1-\varepsilon)k) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}}$$

Proof. Let $Z_1, \cdots, Z_k \stackrel{iid}{\sim} N(0,1)$. By Markov's inequality, for $0 < \lambda < \frac{1}{2}$,

$$\mathbb{P}(\chi_k^2 \geq (1+\varepsilon)k) = \mathbb{P}\left(\sum_{i=1}^{k} Z_i^2 \geq (1+\varepsilon)k\right)$$

$$\leq \frac{\mathbb{E}e^{\lambda \sum_{i=1}^{k} Z_i^2}}{e^{(1+\varepsilon)k\lambda}} = e^{-(1+\varepsilon)k\lambda}(1-2\lambda)^{-k/2}$$

Choose the minimizer $\lambda = \frac{\varepsilon}{2(1+\varepsilon)}$ and use $1+\varepsilon \leq e^{\varepsilon - \frac{\varepsilon^2-\varepsilon^3}{2}}$.

Tying the loose ends

So far, $\mathbb{P}\left(\frac{1}{k}\|Ax\|^2 \geq (1+\varepsilon)\|x\|^2\right) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}}$ and similarly,

$$\mathbb{P}\left(\frac{1}{k}\|Ax\|^2 \leq (1-\varepsilon)\|x\|^2\right) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}}.$$

Tying the loose ends

So far, $\mathbb{P}\left(\frac{1}{k}\|Ax\|^2 \geq (1+\varepsilon)\|x\|^2\right) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}}$ and similarly,

$$\mathbb{P}\left(\frac{1}{k}\|Ax\|^2 \leq (1-\varepsilon)\|x\|^2\right) \leq e^{\frac{-k(\varepsilon^2-\varepsilon^3)}{4}}.$$

Thus,

$$\mathbb{P}\left((1-\varepsilon)\|x\|^2 \leq \frac{1}{k}\|Ax\|^2 \leq (1+\varepsilon)\|x\|^2\right) \geq 1 - 2e^{\frac{-(\varepsilon^2-\varepsilon^3)k}{4}}.$$

■

Overview

Simulating for Norm preservation lemma

- We simulate NP Lemma (which holds for any k) for $k = 100, 200, \cdots, 5000$; $d = 10000$ and $\epsilon = 0.1$.

Simulating for Norm preservation lemma

- We simulate NP Lemma (which holds for any k) for $k = 100, 200, \cdots, 5000$; $d = 10000$ and $\epsilon = 0.1$.
- Generate A and generate x randomly, say $x \sim t_4$.

Simulating for Norm preservation lemma

- We simulate NP Lemma (which holds for any k) for $k = 100, 200, \cdots, 5000$; $d = 10000$ and $\epsilon = 0.1$.
- Generate A and generate x randomly, say $x \sim t_4$.
- Calculate $\frac{\left|\frac{1}{k}||Ax||^2 - ||x||^2\right|}{||x||^2}$

Simulating for Norm preservation lemma

- We simulate NP Lemma (which holds for any k) for $k = 100, 200, \cdots, 5000$; $d = 10000$ and $\epsilon = 0.1$.
- Generate A and generate x randomly, say $x \sim t_4$.
- Calculate $\frac{\left|\frac{1}{k}||Ax||^2 - ||x||^2\right|}{||x||^2}$
- For a fixed k repeat the previous two steps 500 times

Simulating for Norm preservation lemma

- We simulate NP Lemma (which holds for any k) for $k = 100, 200, \cdots, 5000$; $d = 10000$ and $\epsilon = 0.1$.
- Generate A and generate x randomly, say $x \sim t_4$.
- Calculate $\frac{\left|\frac{1}{k}||Ax||^2 - ||x||^2\right|}{||x||^2}$
- For a fixed k repeat the previous two steps 500 times
- Calculate the proportion of times the above ratio is less than ϵ to get the empirical probability

Simulating for Norm preservation lemma (Contd.)

Our goal is to see whether the empirical probability is above the lower bound of the NP Lemma for every k

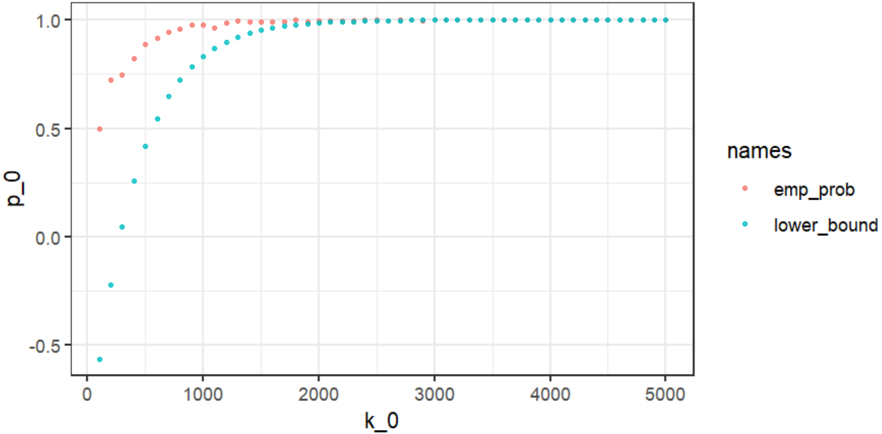


Figure: Empirical Probability vs k

JL Lemma verification

- Let $X_{n \times d} = [[x_{ij}]]$, $x_{ij} \overset{iid}{\sim} Exp(1)$; $n = 5, d = 10000$. Take $\varepsilon = 0.1$.

JL Lemma verification

- Let $X_{n \times d} = [[x_{ij}]]$, $x_{ij} \overset{iid}{\sim} Exp(1); n = 5, d = 10000$. Take $\varepsilon = 0.1$.
- Then $k \approx 3218$.

JL Lemma verification

- Let $X_{n \times d} = [[x_{ij}]]$, $x_{ij} \overset{iid}{\sim} Exp(1)$; $n = 5, d = 10000$. Take $\varepsilon = 0.1$.
- Then $k \approx 3218$.
- Generate $A_{k \times d}$

JL Lemma verification

- Let $X_{n \times d} = [[x_{ij}]]$, $x_{ij} \overset{iid}{\sim} Exp(1)$; $n = 5, d = 10000$. Take $\varepsilon = 0.1$.
- Then $k \approx 3218$.
- Generate $A_{k \times d}$
- Calculate $X_{proj} = (AX^T)^T$.

JL Lemma verification

- Let $X_{n\times d} = [[x_{ij}]]$, $x_{ij} \overset{iid}{\sim} Exp(1)$; $n = 5, d = 10000$. Take $\varepsilon = 0.1$.
- Then $k \approx 3218$.
- Generate $A_{k\times d}$
- Calculate $X_{proj} = (AX^T)^T$.
- For any x_i and x_j, check if

$$\frac{\left| ||x_{proj_i} - x_{proj_j}||^2 - ||x_i - x_j||^2 \right|}{||x_i - x_j||^2} < \varepsilon.$$

```
(abs(((dist(X_new)^2)/k)-(dist(X)^2)))/(dist(X)^2)<=eps
 [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Overview

Applications of JL lemma

- **Nearest-neighbour search**:
  - 1998, Kushilevitz et al used JL to randomly partition space rather than reduce the dimension (The algorithm proposed in the paper is based on locality-sensitive hashing (LSH) and involves mapping the points in the high-dimensional space to a low-dimensional space using a hash function)
  - Finding nearest neighbours without false negatives (2017, Sankowski et al): Based on LSH; The algorithm guarantees that it will not miss the true nearest neighbor and will not return false positives

Applications of JL lemma

- **Nearest-neighbour search**:
  - 1998, Kushilevitz et al used JL to randomly partition space rather than reduce the dimension (The algorithm proposed in the paper is based on locality-sensitive hashing (LSH) and involves mapping the points in the high-dimensional space to a low-dimensional space using a hash function)
  - Finding nearest neighbours without false negatives (2017, Sankowski et al): Based on LSH; The algorithm guarantees that it will not miss the true nearest neighbor and will not return false positives
- **Clustering**:
  - Subspace clustering (2017, Reinhard Heckel et al)
  - Graph clustering (2020, Xiao Guo et al, Randomized Spectral Co-Clustering for Large-Scale Directed Networks)
  - K- means clustering (2019, Luca Becchetti et al ; 2017, Michael B. Cohen et al; 2014,Christos Boutsidis et al)

Applications of JL lemma (Contd.)

- **Several Machine Learning algorithms**: Johnson–Lindenstrauss has been used together with
  - Support Vector Machines (2014, Saurabh Paul et al; 2020, Zijian Lei)
  - Fisher's linear discriminant (2010, Robert Durant et al)
  - Neural networks (2018, Benjamin Schmidt et al)

Applications of JL lemma (Contd.)

- **Several Machine Learning algorithms**: Johnson–Lindenstrauss has been used together with
  - Support Vector Machines (2014, Saurabh Paul et al; 2020, Zijian Lei)
  - Fisher's linear discriminant (2010, Robert Durant et al)
  - Neural networks (2018, Benjamin Schmidt et al)
- **Image data**:
  - Usually images contain $\sim 20,00,000$ dimensions (depending on the resolution of the image)
  - JL lemma can be useful to reduce these dimensions and further use this for classification, clustering, etc.

Example of Application of JL to Image data

Figure: Original grayscale image (1080 px× 1920 px)

Figure: JL reduced grayscale image (1080 px× 1920 px)

Example of Application of JL to Image data (Contd.)

Figure: Original image
(1600 px$\times$ 2560 px)

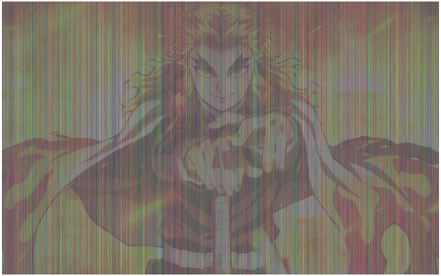


Figure: JL reduced image
(1600 px$\times$ 2560 px)

Overview

Practical implications

A JL map can be found in a randomized polynomial time. Repeating the projection $O(n)$ times, we can boost the success probability to as high as we like, giving a randomized polynomial time algorithm.

Practical implications

A JL map can be found in a randomized polynomial time. Repeating the projection $O(n)$ times, we can boost the success probability to as high as we like, giving a randomized polynomial time algorithm.

Lemma (Distributional JL Lemma)

For $0 < \varepsilon, \delta < \frac{1}{2}$ *and* $d \in \mathbb{N}$, *there exists a distribution over* $\mathbb{R}^{k \times d}$ *from which the matrix* A *is drawn such that for* $k = O(-\log(\delta)/\varepsilon^2)$ *and for* $x \in S^{d-1} \subset \mathbb{R}^d$, *we have* $\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| > \varepsilon\right) < \delta$.

Practical implications

A JL map can be found in a randomized polynomial time. Repeating the projection $O(n)$ times, we can boost the success probability to as high as we like, giving a randomized polynomial time algorithm.

Lemma (Distributional JL Lemma)

For $0 < \varepsilon, \delta < \frac{1}{2}$ and $d \in \mathbb{N}$, there exists a distribution over $\mathbb{R}^{k \times d}$ from which the matrix A is drawn such that for $k = O(-\log(\delta)/\varepsilon^2)$ and for $x \in S^{d-1} \subset \mathbb{R}^d$, we have $\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| > \varepsilon\right) < \delta$.

Taking $x = \frac{u-v}{\|u-v\|_2}$ and $\delta < \frac{1}{n^2}$, the "original" JL lemma follows by taking a union bound over all such pairs.

References

1. *en.wikipedia.org/wiki/Johnson%E2%80%93Lindenstrauss_lemma*
2. *home.ttic.edu/ gregory/courses/LargeScaleLearning/lectures/jl.pdf*
3. *www.math.toronto.edu/undergrad/projects-undergrad/Project03.pdf*
4. *cs.stanford.edu/people/mmahoney/cs369m/Lectures/lecture1.pdf*
5. *arxiv.org/pdf/2103.00564.pdf*