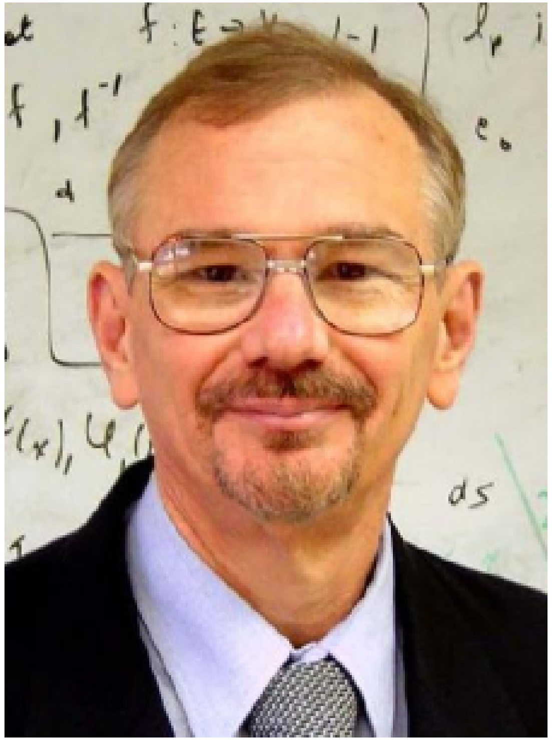
Johnson-Lindenstrauss Lemma and applications

EAI Math Reading Group

22/10/2023

Putting faces on the names



William B. Johnson (1944-)

Joram Lindenstrauss (1936-2012)

The Statement

Theorem:[1]

Given $\varepsilon > 0$ and an arbitrary set V of n points in $\mathbb{R}^d$, there is an integer $k = O\big(\varepsilon^{-2} \ln n\big)$ linear map $f : \mathbb{R}^d \to \mathbb{R}^k$ such that

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

for all $u, v \in V$.

Intuitively, we can project V into a lower dimensional subspace while keeping pairwise distances roughly the same.

[1] W. B. Johnson and J. Lindenstrauss ***Extensions of Lipschitz maps into a Hilbert Space***, Contemp. Math 26 (1984).

Practical uses[2]

Find (approximate) solutions of various problems faster

- Nearest-Neighbor Search
- Clustering (e.g. k-means)
- Outlier Detection
- Numerical Linear Algebra (Low rank approximation, Regression, ...)
- Convex Optimisation

Differential Privacy, Graph Embeddings, ...

C. B. Freksen, *An Introduction to Johnson–Lindenstrauss Transforms*, ArXiv 2103.00564, (2021).

A Proof[3] (1/3)

Lemma 1 Given some arbitrary vector v in $\mathbb{R}^d$, consider its projection v' onto a random k-dimensional subspace, then $L = \|v'\|^2$ has expected value

$$\mu = \mathbb{E}[L] = \frac{k}{d}.$$

Furthermore, by some standard probabilistic arguments, L is concentrated around its mean, i.e. for $k < d$,

- If $\beta < 1$, then

$$\mathbb{P}\left[L \leq \frac{\beta k}{d}\right] \leq \exp\left(\frac{k}{2}(1 - \beta + \ln \beta)\right)$$

- If $\beta > 1$, then

$$\mathbb{P}\left[L \geq \frac{\beta k}{d}\right] \leq \exp\left(\frac{k}{2}(1 - \beta + \ln \beta)\right)$$

---

[3] S. Dasgupta and A. Gupta, *An Elementary Proof of a Theorem of Johnson and Lindenstrauss*, (2003).

A Proof (2/3)

Proof of the Theorem If $d \leq k$, we can just trivially project to $\mathbb{R}^k$. If $k < d$, take a random k-dimensional subspace $S \subset \mathbb{R}^d$ and let v_i' be the projection of $v_i \in V$ into S. Applying the previous lemma to $L = \|v_i' - v_j'\|^2$ and $\mu = \left(\frac{k}{d}\right)\|v_i - v_j\|$, we get

$$\begin{aligned}\mathbb{P}[L \leq (1-\varepsilon)\mu] &\leq \exp\left(\frac{k}{2}(1-(1-\varepsilon)) + \ln(1-\varepsilon)\right)\\ &\leq \exp\left(\frac{k}{2}\left(\varepsilon - \left(\varepsilon + \frac{\varepsilon^2}{2}\right)\right)\right) = \exp\left(-\frac{k\varepsilon^2}{4}\right)\\ &\leq \exp(-2\ln n) = \frac{1}{n^2}\end{aligned}$$

Similarly,

$$\mathbb{P}[L \geq (1+\varepsilon)\mu] \leq \frac{1}{n^2}$$

A Proof (3/3)

Define $f(v_i) = \sqrt{\frac{d}{k}} v_i'$. For any pair i, j, the probability that

$$\|f(v_i) - f\big(v_j\big)\| \notin \big[(1-\varepsilon)\|v_i, v_j\|, (1+\varepsilon)\|v_i - v_j\|\big]$$

is at most $\frac{2}{n^2}$.

Therefore, the probability any pair of points in V has a large distortion is bounded by

$$\binom{n}{2} \frac{2}{n^2} = 1 - \frac{1}{n},$$

That is, f has the desired property with probability at least $\frac{1}{n}$ $\blacksquare$.

Projecting into a random subspace

What does it mean to sample a “random” subspace?

wlog, sample an orthonormal basis of $\mathbb{R}^k$ uniformly[4]. This is equivalent to sampling a random $d \times d$ orthogonal matrix and discarding the last $d - k$ rows.

Numerically, we can do this by sampling a $d \times k$ matrix A with iid $N(0, 1)$ entries and taking its QR decomposition:

$$A = QR$$

The rows of Q are an orthonormal basis for a random k-dimensional subspace of $\mathbb{R}^d$.

---

[4] This is known as the Haar measure on the Stiefel Manifold

Preserving inner products

Corollary Let d, ε, V and f be as defined in the main theorem. Then for every $u, v \in V$, if $-v \in V$, then

$$|\langle f(u), f(v)\rangle - \langle u, v\rangle| \leq \varepsilon \; \|u\|_2 \; \|v\|_2$$

In practice, we can just add the negations of all vectors in V before computing the transform.

Improvements

Improving the bound on k

- [Johnson-Lindenstrauss]: $k = O(\varepsilon^{-2} \ln n)$
- [Frankl-Maehara]: $k \geq \left\lceil 8\left(\varepsilon^2 - 2\frac{\varepsilon^3}{3}\right)^{-1} \ln n \right\rceil$

Constructing the transform

- [Indyk-Motswani]: Use matrices of iid Gaussians (no QR decomposition)
- [Arriaga-Vempala]: Use matrix with random entries in $\{-1, 1\}$
- [Achlioptas]: Use matrices with $\mathbb{P}\left[a_{ij} = 0\right] = \frac{2}{3}$ and $\mathbb{P}\left[a_{ij} = -1\right] = \mathbb{P}\left[a_{ij} = 1\right] = \frac{1}{6}$