CS 368 (Stanford, Spring 2022) Lecture 4

Adapted from CS 369G, Spring 2016. Scribes: Spencer Yee, Michael Xie, Paris Syminelakis.

The Johnson Lindenstrauss Lemma

The fact that we can estimate the second frequency moment very well with a small amount of space is perhaps less surprising given the Johnson-Lindenstrauss Lemma. The statement is as follows.

Lemma 1. *(Johnson-Lindenstrauss [JL84]) Let $G \subset (\mathbb{R}^d, l_2)$ be a set of n points. Then for any $0 < \epsilon < \frac{1}{2}$ and $k = O(\log(n)/\epsilon^2)$, there exists a mapping $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $v_i, v_j \in G$,*

$$(1-\epsilon)\|v_i - v_j\|_2 \le \|f(v_i) - f(v_j)\|_2 \le (1+\epsilon)\|v_i - v_j\|_2$$

For the l_2 norm, it is not hard to show that a set of n points can be mapped to a $(n-1)$-dimensional space without distance distortion and further, that $n-1$ dimensions are indeed needed to preserve distances exactly. What is amazing about the Johnson-Lindenstrauss Lemma is that we can reduce the dimensions to $O(\log(n)/\epsilon^2)$ if we are willing to tolerate a small $1+\epsilon$ distortion in pairwise distances. Many computational tasks on high dimensional data suffer from the so-called *curse of dimensionality*, i.e. the complexity of the best known algorithmic solutions scale very poorly with dimension, typically with an exponential (or worse) dependence. The Johnson-Lindenstrauss lemma provides a generic tool which reduces the dimensionality of the data at the cost of a slight distortion in pairwise distances.

The existence of such a mapping can be proven using a linear mapping of the form

$$f(v) = \frac{Mv}{\sqrt{k}}$$

where $M \in \mathbb{R}^{k\times d}$ and $M_{ij} = \mathcal{N}(0,1)$, a matrix with elements that are drawn i.i.d. from the unit normal distribution. Note that M is data-oblivious; we do not use information from the n points to define its elements.

We remark that the Johnson Lindenstrass Lemma was a helper geometric lemma in the Johnson Lindenstrauss paper [JL84] which was about the metric extension problem. However, this geometric lemma turned out to be the most influential contribution of their paper, at least outside mathematical circles. Progressively simpler proofs were given by Frankl and Maehara [FM88], Indyk and Motwani [IM98], and Gupta and Dasgupta [DG03]. The argument we outline here is closest to the presentation in [DG03].

We use the following tail bound on chi-squared distributions to show the JL lemma.

Lemma 2. *Let $Z_1, \ldots, Z_k$ be i.i.d. unit normal random variables. Let $Y = \sum_i Z_i^2$. Then*

$$\Pr\left[(1-\epsilon)^2 k \le Y \le (1+\epsilon)^2 k\right] \ge 1 - 2e^{-c\epsilon^2 k}$$

for some suitable constant c.

Consider the case where the input vector $v \in \mathbb{R}^d$ is a unit vector. Then for row i of M, we have

$$(Mv)_i = M_i^T v = \sum_j^d M_{ij} v_j = \left(\sum_j^d v_j \right)^{\frac{1}{2}} Y = Y$$

where Y is unit normal. Therefore in the case where the input is a unit vector, the output coordinates are distributed as i.i.d unit Gaussians. Then we can conclude that

$$\Pr\left[(1-\epsilon) \leq \left\| \frac{Mv}{\sqrt{k}} \right\|_2 \leq (1+\epsilon) \right] \geq 1 - 2e^{-c\epsilon^2 k}$$

or equivalently,

$$\Pr\left[(1-\epsilon)^2 k \leq \|Mv\|_2^2 \leq (1+\epsilon)^2 k \right] \geq 1 - 2e^{-c\epsilon^2 k}$$

since $\|Mv\|_2^2 = \sum_i^k (Mv)_i^2$ is a sum of squared unit normal distributions, on which we can invoke Lemma 2.

For general inputs v, we can write $\tilde{v} = v / \|v\|_2$ so that, using the result above,

$$\Pr\left[(1-\epsilon) \leq \left\| \frac{M\tilde{v}}{\sqrt{k}} \right\|_2 \leq (1+\epsilon) \right] \geq 1 - 2e^{-c\epsilon^2 k}$$

Then it follows that

$$\Pr\left[(1-\epsilon)\|v\|_2 \leq \left\| \frac{Mv}{\sqrt{k}} \right\|_2 \leq (1+\epsilon)\|v\|_2 \right] \geq 1 - 2e^{-c\epsilon^2 k}.$$

Taking the inputs v to be the $\binom{n}{2}$ differences between pairs of points in the original space, we use the union bound to show that we can make the failure probability small. The failure probability for a particular difference $v_{ij} = v_i - v_j$ is at most $2e^{-c\epsilon^2 k}$. Choosing $k = (c' \log(n))/\epsilon^2$ for some suitable constant c', we have that the failure probability is at most $2e^{-cc' \log n}$. Choosing sufficiently large constants so that this failure probability is less than $(1/n^3)$, by union bound over the $O(n^2)$ pairs, the failure probability of the random scheme for finding the linear mapping M is bounded by $(1/n)$.

In terms of improvements to this basic scheme, Achlioptas [Ach01] showed that the random projection can be obtained from a random matrix M with iid uniform random $\{\pm 1\}$ entries. The fast JL transform of Ailon and Chazelle[AC09] constructs M with structure such that the projection of a d-dimensional point can be computed in time roughly $O(d \log d + \log^3 n/\epsilon^2))$ can be done. For more general versions of the Johnson-Lindenstrauss Lemma, see [KM05; IN07; Mat08].

Another note is that the F_2 estimator in some sense also preserves the l_2 distances between data streams using a linear mapping; using the Johnson-Lindenstrauss argument with the relaxation that each pair of distances have low distortion with high probability, we can find that we can reduce to a dimension $k = O\left(\frac{1}{\epsilon^2}\log\left(\frac{1}{\delta}\right)\right)$, similar to the space requirement in the F_2 estimator. For dimension reduction schemes where the norm in the original and reduced space is the same, it is not possible to prove results similar to Johnson-Lindenstrauss (i.e. there are strong lower bounds on the number of dimensions needed) for l_∞ [AR92; Mat96] and for l_1 [BC05].

References

[AC09] Nir Ailon and Bernard Chazelle. "The fast Johnson–Lindenstrauss transform and approximate nearest neighbors". In: *SIAM Journal on computing* 39.1 (2009), pp. 302–322.

[Ach01] Dimitris Achlioptas. "Database-friendly random projections". In: *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 2001, pp. 274–281.

[AR92] Juan Arias-de-Reyna and Luis Rodrguez-Piazza. "Finite metric spaces needing high dimension for Lipschitz embeddings in Banach spaces". In: *Israel Journal of Mathematics* 79.1 (1992), pp. 103–111.

[BC05] Bo Brinkman and Moses Charikar. "On the impossibility of dimension reduction in l1". In: *Journal of the ACM (JACM)* 52.5 (2005), pp. 766–788.

[DG03] Sanjoy Dasgupta and Anupam Gupta. "An elementary proof of a theorem of Johnson and Lindenstrauss". In: *Random Structures & Algorithms* 22.1 (2003), pp. 60–65.

[FM88] Peter Frankl and Hiroshi Maehara. "The Johnson-Lindenstrauss lemma and the sphericity of some graphs". In: *Journal of Combinatorial Theory, Series B* 44.3 (1988), pp. 355–362.

[IM98] Piotr Indyk and Rajeev Motwani. "Approximate nearest neighbors: towards removing the curse of dimensionality". In: *Proceedings of the thirtieth annual ACM symposium on Theory of computing*. 1998, pp. 604–613.

[IN07] Piotr Indyk and Assaf Naor. "Nearest-neighbor-preserving embeddings". In: *ACM Transactions on Algorithms (TALG)* 3.3 (2007), 31–es.

[JL84] William B Johnson and Joram Lindenstrauss. "Extensions of Lipschitz mappings into a Hilbert space 26". In: *Contemporary mathematics* 26 (1984), p. 28.

[KM05] B Klartag and Shahar Mendelson. "Empirical processes and random projections". In: *Journal of Functional Analysis* 225.1 (2005), pp. 229–245.

[Mat08] Ji Matouek. "On variants of the Johnson–Lindenstrauss lemma". In: *Random Structures & Algorithms* 33.2 (2008), pp. 142–156.

[Mat96] Ji Matouek. "On the distortion required for embedding finite metric spaces into normed spaces". In: *Israel Journal of Mathematics* 93.1 (1996), pp. 333–344.