Geometric Foundations of Data Analysis I

Joshua Maglione

October 30, 2025

Contents

3.6 Random Projections and Johnson–Lindenstrauss

We now cover a topic somewhat related to **Principal Component Analysis (PCA)** but also very different. The main question we are concerned with is whether we can accomplish dimension reduction without PCA?

Suppose we have n data points in $\mathbb{R}^m$ and we want to project it to $\mathbb{R}^k$ with k much smaller than m while preserving the geometry as much as we can. In other words, we want a map $\varphi : \mathbb{R}^m \to \mathbb{R}^k$ such that all distances are preserved. In symbols, we want, for a small $\varepsilon > 0$ and a subset $S \subseteq \mathbb{R}^m$, for all $u, v \in S$:

$$(1-\varepsilon)\|u-v\|^2 \leqslant \|\varphi(u)-\varphi(v)\|^2 \leqslant (1+\varepsilon)\|u-v\|^2 \tag{3.1}$$

So that ε gives us an **approximation threshold**.

This is different from what happens with PCA. In that case, φ is given by matrix multiplication: using only the first k principal components. There, the **average error** is small, so that some distances can have large errors, and most would be small. In our context right now, φ is much more controlled.

The idea to create $\varphi : \mathbb{R}^m \to \mathbb{R}^k$ so that (3.1) is satisfied is simple: project onto a random k-dimensional subspace. See Algorithm 1.

The following theorem states the probability of this approach yielding the desired outcome.

Theorem 3.17 (Johnson–Lindenstrauss (1984)). *There exists $\varphi : \mathbb{R}^m \to \mathbb{R}^k$ satisfying the inequalities in* (3.1) *with probability at least $1 - \frac{1}{n}$ as long as*

$$k \geqslant \frac{8\ln(n)}{\varepsilon^2}$$

where φ is of the form

$$\varphi(x) = \frac{1}{\sqrt{k}}Ax$$

*where A is a matrix with independent and identically distributed **Gaussian entries** with zero mean and unit variance.*

Algorithm 1 Random Projection

```
Input: $X \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ and $k \in \mathbb{N}$.
Output: $Y \in \mathrm{Mat}_{k\times n}(\mathbb{R})$.
  for $i \in \{1, \dots, k\}$ do
      Choose random vector $v_i$ from a Gaussian distribution.
      Rescale $v_i$: $v_i = \sqrt{\frac{m}{\|v_i\|^2}} v_i$                # Useful for proof.
  end for
  for each column of $X$, $x_i$ do
      $y_i = (x_i \cdot v_1, x_i \cdot v_2, \dots, x_i \cdot v_k)^T$
  end for
  $Y = [y_1, y_2, \dots, y_n]$
  return $Y$.
```

We will not prove Theorem 3.17.

Remark 3.18. Although this is called the JL Lemma (or Theorem), this formulation benefits from recent research on these problems. One can get better bounds for k; see Frankl–Maehara [1]. The conditions on the independent and identically distributed Gaussian entries is an improvement due to Har-Peled–Indyk–Motwani; see [2].

Now that we have two ways to perform dimension reduction, when should we use one over the other?

- Generally, PCA is the safest. It preserves largest distances, but may not preserve the smaller ones. This is essentially down to PCA caring about **high variance** and disregarding **low variance**. However, PCA might be computationally expensive.

- If you want more control over the distances (i.e. want them all essentially the same) or if you need to optimize time or memory, then a random projection would be better, provided the hypotheses of Theorem 3.17 hold.

It is possible to use both with good success; see [3].

References

[1] Peter Frankl and Hiroshi Maehara, *Some geometric applications of the beta distribution*, Ann. Inst. Statist. Math. **42** (1990), no. 3, 463–474.

[2] Sariel Har-Peled, Piotr Indyk, and Rajeev Motwani, *Approximate nearest neighbor: towards removing the curse of dimensionality*, Theory Comput. **8** (2012), 321–350, DOI 10.4086/toc.2012.v008a014. MR2948494

[3] Fan Yang and Sifan Liu and Edgar Dobriban and David P. Woodruff, *How to reduce dimension with PCA and random projections?*, 2021, arXiv:2005.00511.