The Johnson-Lindenstrauss Theorem

The object of this lecture is to introduce and prove the Johnson-Lindenstrauss Theorem. This theorem proves (under certain conditions) that a set of points in a high dimensional space can be embedded into a low dimensional subspace so that distances are preserved.

1. Preliminary Results

We begin with a simple result from probability theory, which we state without proof.

Lemma 1. *Let $Z \in \chi_k^2$ and $\epsilon \in (0, \frac{1}{2})$. Then*

$$\mathbb{P}(Z \geq (1+\epsilon)k) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)},$$

and

$$\mathbb{P}(z \leq (1-\epsilon)k) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}.$$

The following theorem comes from the theory of random matrices. Heuristically, it gives the probability that multiplying a vector by a random matrix preserves the vector's norm.

Theorem 2. *Let A be a $k \times d$ random matrix with i.i.d. entries $a_{ij} \in N(0,1)$. Set $y = \frac{1}{\sqrt{k}}Ax$ so $f : \mathbb{R}^d \to \mathbb{R}^k$. Fix $\epsilon \in (0, \frac{1}{2})$. Then $(1-\epsilon)\|x\|^2 \leq \|y\|^2 \leq (1+\epsilon)\|x\|^2$ with probability at least $1 - 2e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$.*

Proof. Set $Z = \frac{\sqrt{k}}{\|x\|}y$. Then $z_i = \frac{1}{\|x\|}\sum_{j=1}^d a_{ij}x_j$. Each variable z_i is a linear combination of Gaussian random variables. This tells us gives us three important pieces of information:

(1) z_i is a Gaussian random variable.
(2) $\mathbb{E}[a_{ij}] = 0 \; \forall i, j$.
(3) By Theorem 1 (from previous lecture), $Var(z_i) = \mathbb{E}[z_i^2] = 1$.

Combining these three facts, we see that $z_i \in N(0,1)$.

Next, observe that $\|z\|^2 = \sum_{j=1}^d z_i^2$, which implies that $\|z\|^2 \in \chi_k^2$.

Now, applying Lemma 1,

$$\begin{aligned}\mathbb{P}(\|y\|^2 \geq (1+\epsilon)\|x\|^2) &= \mathbb{P}\left(\frac{\|x\|^2\|z\|^2}{k} \geq (1+\epsilon)\|x\|^2\right) \\ &= \mathbb{P}(\|z\|^2 \geq (1+\epsilon)\|x\|^2) \\ &\leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}. \quad (4)\end{aligned}$$

Analogously,

$$\mathcal{P}(\|y\|^2 \leq (1-\epsilon)\|x\|^2) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}. \quad (5)$$

Combining (4) and (5), we have the desired result.

□

There is also an alternative proof for Theorem 2 that highlights the utility of Gaussian distributions. This proof is outlined below.

Proof. Let $y = \frac{1}{\sqrt{k}}Ax$, and let H be a unitary map so that

$$Hx = \begin{pmatrix} \|x\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

(One might recognize H as a Householder reflector.) We then have that $y = \frac{1}{\sqrt{k}}Ax = \frac{1}{\sqrt{k}}AH^*Hx$, due to the fact that H is unitary. Define $\widetilde{A} = AH^*$. Using the fact that Gaussian distributions are rotationarily invariant, $\widetilde{A}$ is a Gaussian matrix as well (with $\widetilde{a_{ij}} \in N(0,1)$).

Therefore,

$$\frac{1}{\sqrt{k}}\widetilde{A}\begin{pmatrix} \|x\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \|x\|\frac{1}{\sqrt{k}}\begin{pmatrix} \|x\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

This implies that $\|y\|^2 = \frac{\|x\|^2}{k}\|g\|^2$, where $g \in \chi_k^2$. Once we have result, we can apply Lemma 1 to retrieve the bounds stated in the theorem.

□

We are now prepared to state and prove the Johnson-Lindenstrauss Theorem.

2. The Johnson-Lindenstauss Theorem

The Johnson-Lindenstrauss Theorem is especially useful for data analysis in large dimensions because it allows us to project the data onto a low dimensional subspace while preserving the basic geometry of the data. What follows is a statement about the existence of a low dimensional embedding. However, the proof of the theorem is constructive, so provides a way to build such an embedding (see notes after the proof).

Theorem 3. ***(Johnson-Lindenstrauss):*** *Let Q be a collection of n points in $\mathbb{R}^d$. Let $\epsilon \in (0, \frac{1}{2})$. Pick an integer $k \geq \frac{20}{\epsilon^2}\log(n)$. Then there exists a Lipschitz map $f = \mathbb{R}^d \to \mathbb{R}^k$ so that $\forall\, u, v \in Q$:*

$$(1-\epsilon)\|u-v\|^2 \leq \|f(u) - f(v)\|^2 \leq (1+\epsilon)\|u-v\|^2. \quad (\star)$$

Proof. Set $y = f(x) = \frac{1}{\sqrt{k}}Ax$ where A is a $k \times d$ random matrix with a_{ij} drawn independently from an $N(0,1)$ distribution. Theorem 2 shows that for any pair $u, v \in Q$, the bound $(\star)$ holds with probability at least $1 - 2e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$. There are $\frac{n(n-1)}{2}$ unique pairs of points u, v. Use a simple union bound. Let F_{ij} be the event that pair $\{u_i, u_j\}$ fails. Theorem 2 implies that $\mathbb{P}(F_{ij}) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$. Therefore, the probability that no pair fails is

$$\begin{aligned} &\leq 1 - \sum_{\substack{\text{distinct pairs} \\ \{i,j\}}} \mathbb{P}(F_{ij}) \\ &\leq 1 - \frac{n(n-1)}{2}e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}. \end{aligned}$$

If we use the k given by the theorem statement, then we see that there is nonzero probability that $(\star)$ holds. This proves the existence of a low dimensional embedding that preserves distances. □

The following are some notes about this result:

(1) This proof provides a way to construct the map f. Namely, f could be a Gaussian random projection. Statements of the theorem that incorporate a construction of f give a probability that f preserves distances. As k is increased (above the necessary minimal value), the probability that a certain f preserves distances goes to 1 exponentially fast.

(2) Using Gaussian random projections is in some sense optimal, but other distributions work as well. For example, if the entires of A are drawn from Bournoulli distribution (the set $\{-1, 1\}$), then

$$\frac{1}{\sqrt{k}}\|Ax\| \leq (1+\epsilon)\|x\|, \quad \text{and} \quad \frac{1}{\sqrt{k}}\|Ax\| \geq (1-\epsilon)\|x\|$$

with probability bounded below by $1 - e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$.

(3) It is known that the Johnson-Lindenstrauss result is sharp up to a factor of $\log\left(\frac{1}{\epsilon}\right)$. This implies that we can build a set of points that require

$$\Omega\left(\frac{\log(n)}{\epsilon^2\log(\frac{1}{\epsilon})}\right)$$

dimensions to accurately represent the distances between the points.