The Johnson-Lindenstrauss Lemma

Rishav Mondal (190101072)
Siddharth Charan (190101085)

October 23, 2021

Table of contents

Table of Contents

Background and Motivation

- Many algorithms used in data analysis and machine learning try to model data in the form of data points with a certain number of features.
- The number of features in the dataset is often referred to it's **dimensionality**.
- Large dimensionality can cause poor performance in many algorithms. Low dimensionality, on the other hand, is good and means inexpensive computation.

Background and Motivation

- Therefore dimensionality reduction has been of particular interest over the years.
- The **Johnson-Lindenstrauss Lemma** (1984) is one of the most impressive and satisfying statements about dimensionality reduction in Euclidean spaces.
- Informally, the lemma says that, given N points in the d-dimensional Euclidean space, we can map these points into a much smaller Euclidean space of logarithmic dimension such that the pairwise Euclidean distances of the given points are preserved up to a multiplicative factor of $1 \pm \epsilon$.

Background and Motivation

- In other words, this lemma gives us a way to *compress* an input of size $n \times d$ (n points with d coordinates) to $O(n\epsilon^{-2} logn)$ while approximately preserving distances between points to a small multiplicative factor.
- This means a significant improvement on the original $O(n^2)$ space required to store the description of the n points, and brings it down to around $O(nlogn)$.

Table of Contents

Minkowski sum of two sets

Definition

We define the **Minkowski sum** of two sets A and B in $\mathbb{R}^d$ as $A + B = \{a + b | a \in A, b \in B\} = \cup_{p \in A}(p + B)$.

Remark

Consider A' and B' to be translated copies of the sets A and B; which are assumed to be closed and bounded. In other words, let $A' = A + p$ and $B' = B + q$ for points $p, q \in \mathbb{R}^d$. Then the set $A' + B' = A + B + p + q$ is also a translated copy of set $A + B$. Since volume(or any Lebesgue measure) is invariant under translation, we get $Vol(A + B) = Vol(A' + B')$.

Minkowski sum of two sets

The image below shows the Minkowski sum of two sets, a square (call it set A), and a circle (call it set B). Every point in resultant $A+B$ is obtained by putting a circle(set B) in place of every point in the square(set A); and then taking the union of all such circles.

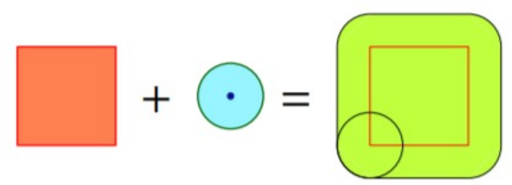

Figure: Minkowski sum of two sets

The Brunn Minkowski Inequality

Theorem (Brunn Minkowski Inequality)

Let A and B be two non empty compact sets in $\mathbb{R}^d$. Then the following inequality holds:
$Vol(A+B)^{\frac{1}{d}} \geq Vol(A)^{\frac{1}{d}} + Vol(B)^{\frac{1}{d}}$

Definition

We define a **brick set** in $\mathbb{R}^d$ as the disjoint intersection of finitely many **boxes** with disjoint interiors. Here by a **box** in $\mathbb{R}^d$, we mean a set of the form $[a_1, b_1] \times [a_2, b_2] \cdots \times [a_d, b_d]$ (a closed axis parallelo-piped). The volume of each such box would be $\prod_{i=1}^{d}(b_i - a_i)$ where $\alpha_i = b_i - a_i$ are the sides lengths of the box.

The Brunn Minkowski Inequality

Claim

If we can prove the Brunn Minkowski Inequality assuming A and B to be brick sets, it implies the truth of the inequality for all non empty compact sets A and B.

Proof.

The proof is intuitively clear by taking monotone sequences of brick sets, passing to a convergent subsequence and then application of limit arguments. $\square$

The Brunn Minkowski Inequality for brick sets

Brunn Minkowski Inequality for brick sets proof.

Let us prove this via induction on the number k of bricks in sets A and B. If $k = 2$ then A and B consist of 1 brick. If we assume dimension of A to be $\{a_1, a_2, \dots a_d\}$ and that of B to be $\{b_1, b_2, \dots b_d\}$; then it is easy to observe that dimensions of $A + B$ to be $\{a_1 + b_1, a_2 + b_2, \dots a_d + b_d\}$. Thus we need to prove that $(\prod_{i=1}^{d}(a_i))^{\frac{1}{d}} + (\prod_{i=1}^{d}(b_i))^{\frac{1}{d}} \leq (\prod_{i=1}^{d}(a_i + b_i))^{\frac{1}{d}}$. If we divide the LHS by the RHS, we need to basically prove that $(\prod_{i=1}^{d} \frac{a_i}{a_i+b_i})^{\frac{1}{d}} + (\prod_{i=1}^{d} \frac{b_i}{a_i+b_i})^{\frac{1}{d}} \leq 1$.

$\square$

The Brunn Minkowski Inequality for brick sets

Brunn Minkowski Inequality for brick sets proof continued.

Notice that by the generalized AM-GM inequality we can write GM $\leq$ AM i.e.
$(\prod_{i=1}^{d} \frac{a_i}{a_i+b_i})^{\frac{1}{d}} + (\prod_{i=1}^{d} \frac{b_i}{a_i+b_i})^{\frac{1}{d}} \leq \frac{1}{d} \cdot (\sum_{i=1}^{d} \frac{a_i}{a_i+b_i} + \sum_{i=1}^{d} \frac{b_i}{a_i+b_i}) = 1.$
Thus we prove the inequality for the base case $k = 2$.
Now let $k > 2$ and let the Brunn Minkowski Inequality holds for sets A and B having a total of fewer than k sets together (induction hypothesis). In addition, assume that A has atleast 2 disjoint bricks. However due to **theorem on existence of separating hyperplane** between any two disjoint convex sets, we must have an axis parallel hyperplane h separating interior of any two disjoint bricks of A. $\square$

The Brunn Minkowski Inequality for brick sets

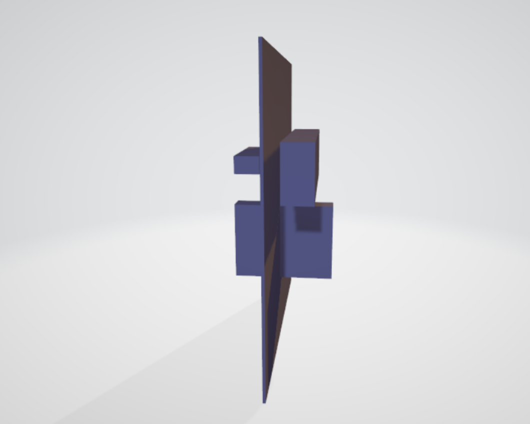

Figure: Dividing A into A^+ and A^- via h

The Brunn Minkowski Inequality for brick sets

Brunn Minkowski Inequality for brick sets proof continued.

Let $\bar{A}^{+} = A \cap h^{+}$ and $\bar{A}^{-} = A \cap h^{-}$ i.e. intersection with the positive and negative half-spaces of h. Let A^{+} and A^{-} be the closures(sets formed by including the boundary points) of sets $\bar{A}^{+}$ and $\bar{A}^{-}$ resp. Notice that A^{+} and A^{-} are also brick sets (both smaller than A by atleast one brick).
Now as volume is translation independent, we can freely translate B such that $Vol(B)$ is split in the same ratio as $Vol(A)$. Like in the previous case, we name two parts of B as B^{+} and B^{-}. Let $\rho = \frac{Vol(A^{+})}{Vol(A)} = \frac{Vol(B^{+})}{Vol(B)}$.
Also observe that since A^{+} and B^{+} lie on the same side of h (and same with A^{-} and B^{-}), we have $A^{+} + B^{+}$ lying on one side of h, and $A^{-} + B^{-}$ on the opposite side. Also, $A^{+} + B^{+} \subseteq A + B$ and $A^{-} + B^{-} \subseteq A + B$. $\square$

The Brunn Minkowski Inequality for brick sets

Brunn Minkowski Inequality for brick sets proof continued.

Thus by induction we have :
$Vol(A+B) \geq Vol(A^{+}+B^{+}) + Vol(A^{-}+B^{-})$
Now by induction hypothesis on smaller sets, we have
$\geq (Vol(A^{+})^{\frac{1}{d}} + Vol(B^{+})^{\frac{1}{d}})^{d} + (Vol(A^{-})^{\frac{1}{d}} + Vol(B^{-})^{\frac{1}{d}})^{d}$
$= (\rho^{\frac{1}{d}}(Vol(A)^{\frac{1}{d}} + Vol(B)^{\frac{1}{d}}))^{d} + ((1-\rho)^{\frac{1}{d}}(Vol(A)^{\frac{1}{d}} + Vol(B)^{\frac{1}{d}}))^{d}$
$= (\rho + (1-\rho)) \cdot (Vol(A)^{\frac{1}{d}} + Vol(B)^{\frac{1}{d}})^{d}$
$= (Vol(A)^{\frac{1}{d}} + Vol(B)^{\frac{1}{d}})^{d}$; and this completes the proof. $\square$

Brunn Minkowski for slice volumes

Theorem (Brunn Minkowski for slice volumes)

Let $\mathcal{P}$ be a convex set defined in $\mathbb{R}^{d+1}$, and now denote $A = \mathcal{P} \cap (x_1 = a), B = \mathcal{P} \cap (x_1 = b)$ and $C = \mathcal{P} \cap (x_1 = c)$ with $a < b < c$ (We call A, B and C to be three slices of $\mathcal{P}$). Then we must have $Vol(B) \geq min(Vol(A), Vol(C))$.

Brunn Minkowski for slice volumes proof.

Let us consider the function $v(t) = (Vol(\mathcal{P} \cap (x_1 = t)))^{\frac{1}{d}}$ and let $t \in [t_{min}, t_{max}] = \mathcal{I}$ be the interval of intersection with the hyperplane $x_1 = t$. **Let us prove that $v(t)$ is concave on $\mathcal{I}$ (i.e. it first increases then decreases)**. If a or c are outside our chosen interval, then minimum volume becomes 0, and so we are already done. Otherwise, let us denote $\alpha = (b - a)/(c - a)$. □

Brunn Minkowski for slice volumes

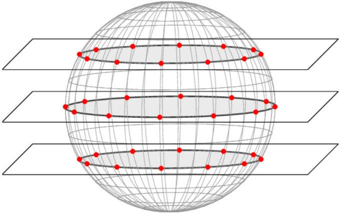

Figure: Slices made in a sphere in $\mathbb{R}^3$

Brunn Minkowski for slice volumes proof contd.

This gives us $b = (1-\alpha)\cdot a + \alpha \cdot c$. Now as $\mathcal{P}$ is a convex set, by definition of convexity we must have $(1-\alpha)\cdot A + \alpha \cdot C \subseteq B$. $\square$

Brunn Minkowski for slice volumes

Brunn Minkowski for slice volumes proof contd.

We have:

$v(b) = Vol(B)^{\frac{1}{d}} \geq Vol((1-\alpha)\cdot A + \alpha \cdot C)$ (by substitution).

$\geq Vol((1-\alpha)\cdot A)^{\frac{1}{d}} + Vol(\alpha \cdot C)^{\frac{1}{d}}$ (By Brunn Minkowski Ineq.)

$= (1-\alpha)\cdot Vol(A)^{\frac{1}{d}} + (\alpha)\cdot Vol(C)^{\frac{1}{d}}$

$= (1-\alpha)\cdot v(a) + (\alpha)\cdot v(c)$.

Now this is the condition required for a function to be concave on an interval and hence we can conclude that $v(t)$ is concave on $\mathcal{I}$.

In particular, this tells us that $v(b) \geq min(v(a), v(c))$; and by raising both sides to the power d we prove that

$Vol(B) \geq min(Vol(A), Vol(C))$.

$\square$

A useful corollary

Corollary

For non empty compact sets A and B in $\mathbb{R}^d$ we have the following result to be true: $Vol(\frac{A+B}{2}) \geq \sqrt{Vol(A) \cdot Vol(B)}$.

Proof.

$Vol(\frac{A+B}{2})^{\frac{1}{d}} = Vol(\frac{A}{2} + \frac{B}{2})^{\frac{1}{d}}$
$\geq Vol(\frac{A}{2})^{\frac{1}{d}} + Vol(\frac{B}{2})^{\frac{1}{d}}$ (By Brunn Minkowski Inequality)
$= \frac{(Vol(A)^{\frac{1}{d}} + Vol(B)^{\frac{1}{d}})}{2}$
(As $Vol(k \cdot A) = k^d \cdot Vol(A)$ so we pull out the constant $k = \frac{1}{2}$).
$\geq \sqrt{Vol(A)^{\frac{1}{d}} \cdot Vol(B)^{\frac{1}{d}}}$ (By using the AM-GM inequality)
Now raising both sides to the power d completes our proof. $\square$

Table of Contents

Measure Concentration on the Sphere

- In this section, our purpose is to find the **concentration of volume/mass** around the n dimensional unit sphere $\mathbb{S}^{n-1}$.
- We assume there is a uniform probability measure defined over the sphere and the total measure equals 1.
- A very surprising result holds true - most of the mass is **concentrated in the narrow strip around any equator!**
- We are going to show that a stronger result holds: The mass is concentrated close to the boundary of any set $A \subseteq \mathbb{S}^{n-1}$ such that $Pr[A] = 0.5$.

The strange and curious life of the hypersphere

- Here we try to get an intuition about the behaviour of the hypersphere in higher dimensions.
- Consider the ball of radius r denoted by rb^d, where b^d is the d dimensional unit ball centered at the origin. We can easily see that $Vol(rb^d) = r^d \cdot Vol(b^d)$.
- Notice that even if r is arbitrarily close to 1, r^d can be close to 0 if d is sufficiently large. For example, even if we take $r = 0.99$, then volume start to come down close to 0 as n takes on higher values ($n \approx 1000$). On the other hand if $r = 1$ then volume remains 1. This establishes the fact that **the volume is concentrated in a very thin shell close to the surface.**

The volume of a ball and the surface area of hypersphere

- Let $Vol(rb^d)$ and $Area(r\mathbb{S}^{d-1})$ denote the volume and surface area of the d dimensional unit sphere.
- It is a known fact that:

$$Vol(rb^d) = \frac{\pi^{d/2}}{\Gamma(d/2+1)} \; ; \qquad Area(r\mathbb{S}^{d-1}) = \frac{2\pi^{d/2}r^{d-1}}{\Gamma(d/2)}$$

- Here $\Gamma(.)$ is an extension of the factorial function used for dealing with real values.
- The most surprising implication of these two formulas is that the volume of the unit ball increases till dimension 5 and then starts decreasing to zero! (same with surface area as well)

The volume of a ball and the surface area of hypersphere

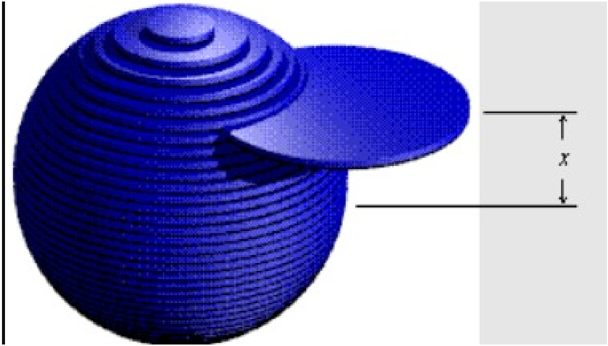

Figure: Slicing the sphere along parallel planes

The volume of a ball and the surface area of hypersphere

- To see this, let us compute the volume of the unit ball by integrating volumes of the $(d-1)$ dimensional slices obtained by chopping the ball along hyperplanes perpendicular to the d^{th} coordinate.

$$\begin{aligned} Vol(b^d) &= \int_{x_d=-1}^{1} Vol(\sqrt{1-x_d^2} \cdot b^{d-1}) dx_d \\ &= Vol(b^{d-1}) \int_{x_d=-1}^{1} (1-x_d^2)^{(d-1)/2} dx_d \end{aligned}$$

- Notice that for large values of d, the term $(1-x_d^2)^{(d-1)/2}$ goes to 0 except for a small interval where $x_d \to 0$. This implies that the main contribution of the volume of the ball happens when we consider slices of the ball by hyperplanes of the form $x_d = \delta$, where δ is very small.

Measure Concentration on the Sphere theorem

Theorem (Measure Concentration on the Sphere theorem)

Let $A \subset \mathbb{S}^{n-1}$ be a measurable set. Given that $Pr[A] \geq 0.5$; and let A_t denote the set of points of $\mathbb{S}^{n-1}$ which have distance atmost t from A, where $t \leq 2$. Then we must have $Pr[A_t] \geq 1 - 2 \cdot exp(\frac{-nt^2}{2})$.

Measure Concentration on the Sphere theorem proof.

- The proof is a consequence of the Brunn-Minkowski inequality. Let A be any measurable surface on the unit sphere with $Pr[A] \geq \frac{1}{2}$. Let B be the surface on the unit sphere $\mathbb{S}^{n-1}$ containing points at least t far from A. In other words, $B = \mathbb{S}^{(n-1)} \backslash A_t$.

□

Measure Concentration on the Sphere theorem

Measure Concentration on the Sphere theorem proof contd.

- We consider the sets $\tilde{A}, \tilde{B} \subset b^n$ as follows: $\tilde{A} = \{\alpha x | x \in A; \alpha \in [0,1]\}$ and $\tilde{B} = \{\alpha x | x \in B; \alpha \in [0,1]\}$
- We observe that $Vol(\tilde{A}) = Pr[A] \cdot v_n = v_n/2$ and similarly $Vol(\tilde{B}) = Pr[B] \cdot v_n$.
- We will bound the distance of any point $\frac{\tilde{a}+\tilde{b}}{2}$ from the origin.The distance is maximized when $\tilde{a} \in A$ and $\tilde{b} \in B$ and also they are as close to each other as possible. However by definition we must have $||a-b|| \geq t$.

$\square$

Measure Concentration on the Sphere theorem

Measure Concentration on the Sphere theorem proof contd.

- This implies we must have $||\frac{\tilde{a}+\tilde{b}}{2}|| \leq \sqrt{1-t^2/4} \leq 1-t^2/8$. However this also implies we must have all points in the body $\frac{\tilde{A}+\tilde{B}}{2}$ contained in a sphere of radius $1-t^2/8$.
- By discussed corollary of Brunn-Minkowski inequality, we have: $Vol(\frac{A+B}{2}) \geq \sqrt{Vol(A) \cdot Vol(B)}$
$\implies \sqrt{\frac{v_n}{2} \cdot Pr[B]v_n} \leq (1-\frac{t^2}{8})^n v_n$
$\implies Pr[B] \leq 2(1-\frac{t^2}{8})^{2n} \leq 2exp(-nt^2/4)$
- This proves that $Pr[A_t] \geq 1-2\cdot exp(\frac{-nt^2}{4})$. which is slightly weaker than what we set out to prove (i.e. $1-2\cdot exp(\frac{-nt^2}{2})$).

$\square$

Table of Contents

The Isoperimetric Inequality

- This item is additional and not used anywhere in derivations concerning the JL Lemma, but still provided because of its mathematical elegance.
- The **isoperimetric inequality** states that for convex bodies, if you fix the surface area to be the same, then it will be the ball (solid sphere) which will have the largest volume.
- In particular, if we consider all convex bodies of surface area 4π, then it will be the unit sphere which will have the largest volume $(= \frac{4\pi}{3})$.
- The proof of the general d dimensional case is an easy consequence of the **Brunn Minkowski Inequality.**

The Isoperimetric Inequality

The isoperimetric inequality has practical consequences as well. For example, a water droplet will try to reduce its surface tension, which is proportional to its surface area. So out of all possible shapes, it will acquire a symmetrical solid spherical shape.

Figure: Raindrop acquiring spherical shape

The Isoperimetric Inequality

Theorem (The Isoperimetric Inequality)

Let K be any convex body in $\mathbb{R}^d$ and also let b be the n dimensional unit sphere centered at the origin. Let $S(X)$ denote the surface area of any compact set $X \subseteq \mathbb{R}^d$. Then we must have:

$$\left(\frac{Vol(K)}{Vol(b)}\right)^{\frac{1}{d}} \leq \left(\frac{S(K)}{S(b)}\right)^{\frac{1}{d-1}}$$

In other words, if we scale K to have the same surface area as b, we must have $Vol(K) \leq Vol(b)$. It also equivalently means that if we scale K to have the same volume as b, we must have $S(K) \geq S(b)$.

The Isoperimetric Inequality

The Isoperimetric Inequality proof.

Let K be some convex body defined in $\mathbb{R}^d$. Now consider the very small region of width ϵ around it. This region will have volume approximately equal to $\epsilon \cdot S(K)$. Also we shall be using the fact that for the d dimensional ball, $Vol(b) = \frac{S(b)}{d}$ (can be proved by considering the d-ball as a union of concentric $(d-1)$-balls).

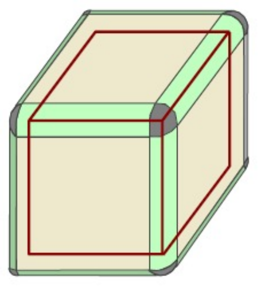

Figure: K with very small region of width ϵ around it(in green)

The Isoperimetric Inequality

The Isoperimetric Inequality proof contd.

We define surface area $S(K)$ as follows:

$$S(K) = lim_{\epsilon\to 0^+} \frac{Vol(K+\epsilon b) - Vol(K)}{\epsilon}$$

$$\geq lim_{\epsilon\to 0^+} \frac{(Vol(K)^{\frac{1}{d}} + Vol(\epsilon b)^{\frac{1}{d}})^d - Vol(K)}{\epsilon} \qquad \text{(By Brunn Mink. Ineq)}$$

$$= lim_{\epsilon\to 0^+} \frac{(Vol(K) + \binom{d}{1}\epsilon Vol(K)^{\frac{d-1}{d}} Vol(b)^{\frac{1}{d}}) + (\text{higher order terms}) - Vol(K)}{\epsilon}$$

We neglect higher order terms as their value $\to 0$ anyway.

$$= lim_{\epsilon\to 0^+} \frac{d\epsilon Vol(K)^{(d-1)/d} Vol(b)^{\frac{1}{d}}}{\epsilon} = dVol(K)^{(d-1)/d} Vol(b)^{\frac{1}{d}}$$

Dividing both sides by $S(b) = dVol(b)$ and raising both sides to power d we get our desired result:

$$\left(\frac{Vol(K)}{Vol(b)}\right)^{\frac{1}{d}} \leq \left(\frac{S(K)}{S(b)}\right)^{\frac{1}{d-1}} \qquad ; \text{hence proved.}$$

□

Thank you

Thank you for attending! Any questions?