Dimensionality Reduction

Anil Maheshwari

anil@scs.carleton.ca
School of Computer Science
Carleton University
Canada

Outline

Metric Space

Isometric embedding

Universal Spaces

Distortion

L_∞ Norm

Corollaries

Normal Distribution

L_2 Norm - Johnson-Lindenstrauss Theorem

Metric Space

Metric Space $\langle X, d \rangle$

Let X be a set of n-points and let d be a distance measure associated with pairs of elements in X.

We say that $\langle X, d \rangle$ is a finite metric space if the function d satisfies metric properties, i.e.

(a) $\forall x \in X$, $d(x, x) = 0$,
(b) $\forall x, y \in X, x \neq y, d(x, y) > 0$,
(c) $\forall x, y \in X$, $d(x, y) = d(y, x)$ (symmetry), and
(d) $\forall x, y, z \in X$, $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

Isometric embedding

Embeddings

Let $\langle X, d \rangle$ and $\langle X', d' \rangle$ be two metric spaces.

Embedding: A map $f : X \to X'$ is called an embedding.

Isometric embedding (i.e., distance preserving) if for all $x, y \in X$, $d(x, y) = d'(f(x), f(y))$.

3-useful distance measures between a pair of points $p = (p_1, \ldots, p_k)$ and $q = (q_1, \ldots, q_k)$ in $\Re^k$.

1. L_2-norm (Euclidean): $||p - q||_2 = \sqrt{\sum_{i=1}^{k} (p_i - q_i)^2}$
2. L_1-norm (Manhattan): $||p - q||_1 = \sum_{i=1}^{k} |p_i - q_i|$
3. L_∞-norm: $||p - q||_\infty = \max\{|p_1 - q_1|, \ldots, |p_k - q_k|\}$

Motivating Problem

Input: X=Set of n-points in k-dimensional space, where $n >> 2^k$
Output: A pair of points that maximize L_1-distance.

Let $p = (p_1, \ldots, p_k)$ and $q = (q_1, \ldots, q_k)$ be two points in $\Re^k$,
$||p - q||_1 = \sum_{i=1}^{k} |p_i - q_i|$.
For example, $||(3,5) - (2,7)||_1 = |3-2| + |5-7| = 3$.

Naive Solution: Compute distance between every pair of points and find the pair with largest distance
Total Time $= O(k\binom{n}{2}) = O(kn^2)$.

Next: An algorithm using isometric embedding of $L_1^k \to L_\infty^{2^k}$ running in $O(2^k n)$ time.

Isometric embedding $f : L_1^k \to L_\infty^{2^k}$

Let $x = (x_1, \ldots, x_k) \in X$

Note that $||x||_1 = \sum_{i=1}^{k} |x_i| = \sum_{i=1}^{k} \text{sign}(x_i) x_i = \text{sign}(x) \cdot x$, where sign$(x)$ is the ± 1 vector of length k denoting the sign of each coordinate of x.

Claim 1

For any ± 1 vector $y = (y_1, \ldots, y_k)$ of length k $||x||_1 = \text{sign(x)} \cdot x \geq y \cdot x$.
Moreover, $||x||_1 = \max\{y \cdot x | y \in \{-1, 1\}^k\}$.

An Illustration

Claim 1

For any ± 1 vector $y = (y_1, \ldots, y_k)$ of length k $||x||_1 =$ sign(x) $\cdot x \geq y \cdot x$. Moreover, $||x||_1 = \max\{y \cdot x | y \in \{-1, 1\}^k\}$.

For $x = (-2, -3, 4)$, $||x||_1 = |-2| + |-3| + |4| = (-1, -1, 1) \cdot (-2, -3, 4) = 9$

$y \cdot x$		
$(\mathbf{-1, -1, 1}) \cdot (\mathbf{-2, -3, 4})$	=	$\mathbf{9}$
$(-1, 1, 1) \cdot (-2, -3, 4)$	=	3
$(1, -1, 1) \cdot (-2, -3, 4)$	=	5
$(1, 1, 1) \cdot (-2, -3, 4)$	=	-1
$(-1, -1, -1) \cdot (-2, -3, 4)$	=	1
$(-1, 1, -1) \cdot (-2, -3, 4)$	=	-5
$(1, -1, -1) \cdot (-2, -3, 4)$	=	-3
$(1, 1, -1) \cdot (-2, -3, 4)$	=	-9

Isometric embedding $f : L_1^k \to L_\infty^{2^k}$ (contd.)

For each ± 1 vector y, define $f_y : X \to \Re$ by $f_y(x) = y \cdot x$

For example, $f_{(1,-1,1)}((-2,-3,4)) = (1,-1,1) \cdot (-2,-3,4) = 5$

Isometric Embedding

Define $f : X \to \Re^{2^k}$ to be the concatenation of f_y's for all possible 2^k $y's$.

For our example, $f(x) = (9, 3, 5, -1, 1, -5, -3, -9)$ corresponding to $2^3 = 8$ possible values for 3-dimensional vector y.

Let $x = (-2, -3, 4)$ and $x' = (2, 3, -2)$.

$||x - x'||_1 = |-2-2| + |-3-3| + |4-(-2)| = 16$

$f(x') = (-7, -1, -3, 3, -3, 3, 1, 7)$.

Observe
$||f(x) - f(x')||_\infty = \max_y\{|f_y(x) - f_y(x')|\} = \max(|9-(-7)|, |3-(-1)|, |5-(-3)|, |-1-3|, |1-(-3)|, |-5-3|, |-3-1|, |-9-7|) = 16 = ||x - x'||_1$

Isometric embedding $f : L_1^k \to L_\infty^{2^k}$ (contd.)

Isometric Embedding Lemma

Under the mapping $f : X \to \Re^{2^k}$ given by the concatenation of f_y's for all possible 2^k $y's$, where $f_y(x) = y \cdot x$, we have that for any two points $x, x' \in X$, $||f(x) - f(x')||_\infty = ||x - x'||_1$

Proof Sketch:

$$\begin{aligned} ||f(x) - f(x')||_\infty &= \max_y\{|f_y(x) - f_y(x')|\} \\ &= \max_y\{|y \cdot x - y \cdot x'|\} \\ &= \max_y\{|y \cdot (x - x')|\} \\ &= ||x - x'||_1, \end{aligned}$$

(by Claim 1 $||x||_1 = \max\{y \cdot x | y \in \{-1, 1\}^k\}$)

$\square$

Isometric embedding $f : L_1^k \to L_\infty^{2^k}$ (contd.)

In place of finding the furthest pair of points in X with respect to L_1 metric we have the following:

New Problem: Given n points in 2^k dimensional space X', find the furthest pair in X' with respect to L_∞ metric.

$$\begin{aligned} \max_{u,v \in X'} ||u - v||_\infty &= \max_{u,v \in X'} \max_{i=1}^{2^k} |u_i - v_i| \\ &= \max_{i=1}^{2^k} \max_{u,v \in X'} |u_i - v_i| \end{aligned}$$

Fix a coordinate, find the pair of points that maximize the difference with respect to that coordinate. Among all the coordinates, pick the one that maximizes the difference.

Furthest pair using L_∞ metric

Observe that $\max_{u,v \in X'} |u_i - v_i|$, for a fixed i, can be computed in $O(n)$ time

$\implies \max_{i=1}^{2^k} \max_{u,v \in X'} |u_i - v_i|$ can be computed in $O(2^k n)$ time.

Theorem

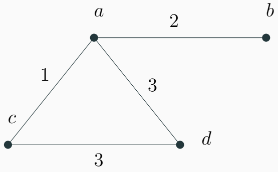
Given a set X of n points in $\Re^k$, by using the isometric embedding $f : L_1^k \to L_\infty^{2^k}$, we can compute the furthest pair of points in X with respect to L_1-metric by computing the furthest pair of points in the embedding with respect to L_∞-metric in $O(2^k n)$ time.

Universal Spaces

Universality of L_∞-metric space

Let $\langle X, d\rangle$ be any finite metric space, where $n = |X|$.
X can be isometrically embedded into L_∞-metric space of dimension n.

Proof Sketch: Let $X = \{x_1, \ldots, x_n\}$.
For each point $x \in X$, define $f(x) = (d(x, x_1), \ldots, d(x, x_n))$.



For example, let $X = \{a, b, c, d\}$, and we have

$$f(a) = (d(a,a), d(a,b), d(a,c), d(a,d)) = (0,2,1,2)$$
$$f(b) = (d(b,a), d(b,b), d(b,c), d(b,d)) = (2,0,3,5)$$
$$f(c) = (d(c,a), d(c,b), d(c,c), d(c,d)) = (1,3,0,3)$$
$$f(d) = (d(d,a), d(d,b), d(d,c), d(d,d)) = (3,5,3,0)$$

$$d(b,d) = ||f(b) - f(d)||_\infty = 5$$
$$d(a,d) = ||f(a) - f(d)||_\infty = 3$$

Universality of L_∞-metric (contd.)

Claim

For any pair of points $u, v \in X$, we have $d(u,v) = ||f(u) - f(v)||_\infty$

Proof:

$$\begin{aligned} ||f(u) - f(v)||_\infty &= \max_{x \in X} |d(u,x) - d(v,x)| \\ &\leq d(u,v) \text{ by triangle inequality} \end{aligned}$$

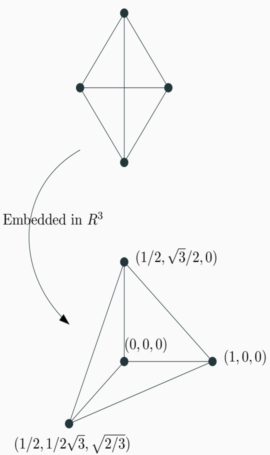
But, $\max_{x \in X} |d(u,x) - d(v,x)| \geq |d(u,u) - d(v,u)| = d(u,v)$
$\implies ||f(u) - f(v)||_\infty = d(u,v)$

□

$\implies$ the mapping of elements of $x \in X$ given by $f(x) = (d(x,x_1), \ldots, d(x,x_n))$ under L_∞-norm is universal.

Euclidean Metric

Input: Metric Space defined by K_4, C_4, and a star w.r.t. unweighted SP.
Question: Can one embed 4-points in Euclidean space (L_2) in any dimension isometrically?



Distortion

Distortion

Contraction: Is the maximum factor by which the distances shrink and it equals $\max\limits_{x,y \in X} \frac{d(x,y)}{d'(f(x),f(y))}$.

Expansion: Is the maximum factor by which the distances are stretched and it equals $\max\limits_{x,y \in X} \frac{d'(f(x),f(y))}{d(x,y)}$.

Distortion: of an embedding is the product of its expansion and contraction factor.

L_∞ Norm

Input: A metric space $\langle X, d\rangle$, where X is a set of n-points and let d satisfies the metric properties.
Output: An embedding of X in a $k = O(Dn^{\frac{2}{D}}\log n)$ dimensional space such that the distances gets distorted (actually contracted) by a factor of at most D under L_∞ norm.

We denote this embedding by the following notation:

$$\langle X, d\rangle \overset{D}{\hookrightarrow} L_\infty^{k=O(Dn^{\frac{2}{D}}\log n)}$$

Note, when $D = O(\log n)$, we have

$$\langle X, d\rangle \overset{\log n}{\hookrightarrow} L_\infty^{k=O(\log^2 n)}$$

I.e. we can embed any metric space in $O(\log^2 n)$ dimensional L_∞-metric space and the distances are distorted by a factor of $O(\log n)$.

Let $x, y \in X$ and let $f(x), f(y)$ be their embedding in the k-dimensional space, respectively.

Property

The distances gets contracted by a factor of at most $D \geq 1$. Formally, $\max_{x,y\in X} \frac{d(x,y)}{||f(x)-f(y)||_{\infty}} \leq D$

Example: If $D = O(\log n)$, $k = O(\log^2 n)$, i.e. $\langle X, d\rangle \overset{O(\log n)}{\hookrightarrow} L_{\infty}^{O(\log^2 n)}$

Meaning: Any metric space $\langle X, d\rangle$ can be embedded in a $O(\log^2 n)$-dimensional space and the distances may distort (contract) by a factor of at most $O(\log n)$.

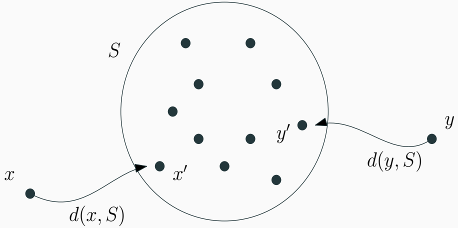
Space Saving Embedding: $\langle X, d\rangle$, where $n = |X|$, may require $O(n^2)$ space to capture distances between points. Whereas, in the mapped k-dimensional space, we only need to store $k = O(\log^2 n)$ coordinates for each point, thus requiring a total of $O(n \log^2 n)$ space.

Proof of $\langle X, d\rangle \overset{D}{\hookrightarrow} L_\infty^{k=O(Dn^{\frac{2}{D}}\log n)}$

Constructive proof via a randomized algorithm.

Definition

Let $S \subseteq X$. For $x \in X$, define the distance of x to the set S as $d(x, S) = \min_{z \in S} d(x, z)$



Claim 1

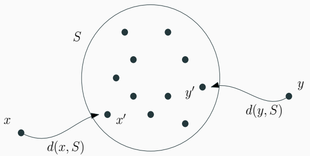
Let $x, y \in X$. For all $S \subseteq X$, $|d(x, S) - d(y, S)| \leq d(x, y)$.

Proof of Claim 1

Claim 1

Let $x, y \in X$. For all $S \subseteq X$, $|d(x, S) - d(y, S)| \leq d(x, y)$.

Proof:



Let $|d(x, S) - d(y, S)| = |d(x, x') - d(y, y')|$.

If $d(x, x') \geq d(y, y')$
$d(x, x') - d(y, y') \leq d(x, y') - d(y', y) \leq d(x, y)$ (by triangle inequality)
else $d(y, y') - d(x, x') \leq d(y, x') - d(x, x') \leq d(x, y)$.

Thus, $|d(x, S) - d(y, S)| = |d(x, x') - d(y, y')| \leq d(x, y)$.

□

$\Longrightarrow$ Distance to a subset amounts to contraction.

Proof Contd.

Definition

(**Mapping**) Let $x \in X$. Let $S_1, S_2, \cdots, S_k \subseteq X$. The mapping f maps x to the point

$$f(x) = \{d(x, S_1), d(x, S_2), \cdots, d(x, S_k)\}.$$

Claim 2

Let $S_1, S_2, \cdots, S_k \subseteq X$. For any pair of points $x, y \in X$, $||f(x) - f(y)||_\infty \leq d(x, y)$.

Proof: Follows from Claim 1, as for each $1 \leq i \leq k$, $|d(x, S_i) - d(y, S_i)| \leq d(x, y)$.

$\square$

Randomized Algorithm

Input: Metric space $\langle X, d \rangle$ and an integer parameter D.
Output: A set of $O(Dm)$ subsets of X.

1. $p \leftarrow \min(\frac{1}{2}, n^{-\frac{2}{D}})$
2. $m \leftarrow O(n^{\frac{2}{D}} \log n)$
3. For $j \leftarrow 1$ to $\lceil \frac{D}{2} \rceil$ and
   For $i \leftarrow 1$ to m:
   Choose set S_{ij} by sampling each element of X independently with probability p^j
4. For each $x \in X$ return $f(x) = [d(x, S_{11}), \cdots d(x, S_{m1}),$
   $d(x, S_{12}), \cdots, d(x, S_{m2}), \cdots\ d(x, S_{1\lceil \frac{D}{2} \rceil}), \cdots, d(x, S_{m\lceil \frac{D}{2} \rceil})]$

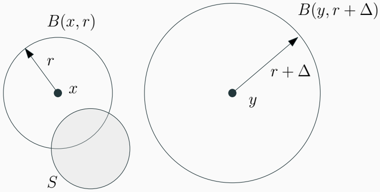
Intuition

- Each point $x \in X$ is embedded in $k = O(Dm)$ dimensional space via the mapping $f(x)$.
- By Claim 2, for any pair of points $x, y \in X$, $||f(x) - f(y)||_\infty \leq d(x, y)$, i.e. the distance shrinks.
- Fix a pair of points $x, y \in X$. We will prove a key lemma that states the following: *There exists an index $j \in \{1, \cdots, \lceil \frac{D}{2} \rceil\}$ such that if S_{ij} is as chosen in the Algorithm, than $Pr\left[||f(x) - f(y)||_\infty \geq \frac{d(x,y)}{D}\right] \geq \frac{p}{12}$.*
  In other words, under the L_∞-norm in the k-dimensional space, the distance doesn't shrink a lot!
- For index j we have m trials. So the probability that the above statement doesn't hold for all the m trials is $\leq (1 - \frac{p}{12})^m \leq e^{-\frac{pm}{12}} \leq \frac{1}{n^2}$. This follows from the choice of p and m as $p \leftarrow \min(\frac{1}{2}, n^{-\frac{2}{D}})$ and $m \leftarrow O(n^{\frac{2}{D}} \log n)$.
- We will apply the union bound to show that the above statement holds for all pairs of points with probability at least $1/2$.

An Observation

Observation 1

Let x, y be two distinct points of X. Let $B(x, r)$ be the set of points of X that are within a distance of r from x (think of $B(x, r)$ as a ball of radius r centred at x). Similarly, let $B(y, r + \Delta)$ be the set of points of X that are within a distance of $r + \Delta$ from y. Consider a subset $S \subset X$ such that $S \cap B(x, r) \neq \emptyset$ and $S \cap B(y, r + \Delta) = \emptyset$. Then $|d(x, S) - d(y, S)| \geq \Delta$.



Proof: $d(x, S) \leq r$ as $S \cap B(x, r) \neq \emptyset$

$d(y, S) \geq r + \Delta$ as $S \cap B(y, r + \Delta) = \emptyset$

$\implies |d(x, S) - d(y, S)| \geq \Delta$

$\square$

Ball Properties

Let $x, y \in X$. Set $\Delta = \frac{d(x,y)}{D}$.

Balls centred at x and y

For $i = 0, \cdots, \lceil \frac{D}{2} \rceil$, define balls of radius $i\Delta$ as follows.

Let $B_0 = \{x\}$.

B_1 be the ball of radius Δ centred at y.

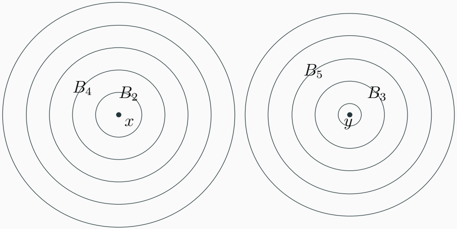
B_2 is the ball of radius 2Δ centred at x.

B_3 is the ball of radius 3Δ centred at y.

B_4 is the ball of radius 4Δ centred at x .

$\ldots$

$\ldots$



Properties of Balls

Property I

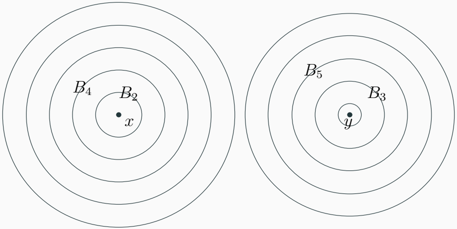
No balls centred at x overlaps with any of the balls centred at y.

Proof: Furthest point balls centred at x can reach is at distance $\leq \lceil \frac{D}{2} \rceil \Delta$.

Similarly, furthest point balls centred at y can reach is at distance $\leq (\lceil \frac{D}{2} \rceil - 1)\Delta$.

But $\lceil \frac{D}{2} \rceil \Delta + (\lceil \frac{D}{2} \rceil - 1)\Delta = 2\lceil \frac{D}{2} \rceil \Delta - \Delta < d(x, y)$, as $\Delta = \frac{d(x,y)}{D}$

□



Ball Properties (contd.)

For even (odd) i, let $|B_i|$ denote the number of points of X that are within a distance of at most $i\Delta$ from x (respectively, y).

Property II

There is an index $t \in \{0, \cdots, \lceil \frac{D}{2} \rceil - 1\}$, such that $|B_t| \geq n^{\frac{2t}{D}}$ and $|B_{t+1}| \leq n^{\frac{2(t+1)}{D}}$

Proof: Proof by contradiction.

$t = 0$: Since $|B_0| = 1 \implies |B_1| > n^{\frac{2}{D}}$

$t = 1$: If $|B_1| > n^{\frac{2}{D}} \implies |B_2| > n^{\frac{4}{D}}$

$t = 2$: If $|B_2| > n^{\frac{4}{D}} \implies |B_3| > n^{\frac{6}{D}}$

$\cdots$

$t = \lceil \frac{D}{2} \rceil - 1$: If $|B_t| > n^{\frac{2t}{D}} \implies |B_{\lceil \frac{D}{2} \rceil}| > n^{\frac{2\lceil \frac{D}{2} \rceil}{D}} \geq n$

But no ball can contain more than $|X| = n$ points. A contradiction.

□

Ball Properties (contd.)

Let t be the index such that $|B_t| \geq n^{\frac{2t}{D}}$ and $|B_{t+1}| \leq n^{\frac{2(t+1)}{D}}$
Consider when $j = t + 1$ in the Algorithm.

Property III

The set S_{ij} chosen by the algorithm has non-empty intersection with B_t with probability at least $p/3$, and it avoids B_{t+1} with probability at least $1/4$.

Define two events:
Event E_1: $S_{ij} \cap B_t \neq \emptyset$.
Event E_2: $S_{ij} \cap B_{t+1} = \emptyset$.
We will show that $Pr(E_1) \geq p/3$ and $Pr(E_2) \geq 1/4$.

By Property I, the balls B_t and B_{t+1} are disjoint.
Thus, $Pr(E_1 \wedge E_2) = Pr(E_1)Pr(E_2)$.
$\implies Pr(E_1 \wedge E_2) \geq \frac{p}{12}$.

Event E_1

$Pr(S_{ij} \cap B_t \neq \emptyset) \geq p/3$

Proof:

$$
\begin{aligned}
Pr(E_1) &= 1 - Pr(S_{ij} \cap B_t = \emptyset) \\
&= 1 - (1 - p^j)^{|B_t|} \text{ (No element of } B_t \text{ is chosen in } S_{ij}) \\
&= 1 - (1 - p^j)^{n^{\frac{2(j-1)}{D}}} \\
&\geq 1 - e^{-p^j n^{\frac{2(j-1)}{D}}} \\
&= 1 - e^{-p^j n^{\frac{2}{D} j} n^{-\frac{2}{D}}} \\
&= 1 - e^{-n^{-\frac{2}{D}}} \text{ (As } p = n^{-\frac{2}{D}}) \\
&= 1 - e^{-p}
\end{aligned}
$$

If $p < \frac{1}{2}$, $1 - e^{-p} \geq p/3$.

□

Event E_2

$Pr(S_{ij} \cap B_{t+1} = \emptyset) \geq 1/4$

Proof:

$$
\begin{aligned}
Pr(E_2) &= Pr(S_{ij} \cap B_{t+1} = \emptyset) \\
&= (1 - p^j)^{|B_{t+1}|} \\
&\geq (1 - p^j)^{n^{\frac{2j}{D}}} \\
&= (1 - p^j)^{\frac{1}{p^j}}
\end{aligned}
$$

If $p^j < \frac{1}{2}$, $(1 - p^j)^{\frac{1}{p^j}} \geq \frac{1}{4}$.

The function $(1 - p^j)^{\frac{1}{p^j}}$ achieves minimum at $p^j = 0$ or $p^j = \frac{1}{2}$, and in both the cases it is $\geq \frac{1}{4}$.

□

Lemma

Let x, y be two distinct points of X. There exists an index $j \in \{1, \cdots, \lceil \frac{D}{2} \rceil\}$ such that if S_{ij} is as chosen in the Algorithm, than $Pr\left[||f(x) - f(y)||_\infty \geq \frac{d(x,y)}{D}\right] \geq \frac{p}{12}$

1. $p \leftarrow \min(\frac{1}{2}, n^{-\frac{2}{D}})$
2. $m \leftarrow O(n^{\frac{2}{D}} \log n)$
3. For $j \leftarrow 1$ to $\lceil \frac{D}{2} \rceil$ and
   For $i \leftarrow 1$ to m:
   Choose set S_{ij} by sampling each element of X independently with probability p^j
4. For each $x \in X$ return $f(x) = [d(x, S_{11}), \cdots d(x, S_{m1}),$
   $d(x, S_{12}), \cdots, d(x, S_{m2}), \cdots \; d(x, S_{1\lceil \frac{D}{2} \rceil}), \cdots, d(x, S_{m\lceil \frac{D}{2} \rceil})]$

Proof of Key Lemma

Fix a pair of points $x, y \in X$. We know that $\Delta = \frac{d(x,y)}{D}$.

By Property II, there is a value of $t \in \{0, \ldots, \lceil \frac{D}{2} \rceil - 1\}$, such that $|B_t|$ is sufficiently large and $|B_{t+1}|$ is not too big. Choose $j = t + 1$.

By Property III, the probability that S_{ij} chosen by the algorithm overlaps with B_t and avoids B_{t+1} completely is at least $p/12$.

What is the probability that none of the m trials are good for that value of j?

$$\leq \left(1 - \frac{p}{12}\right)^m \leq e^{-\frac{pm}{12}} \leq \frac{1}{n^2}$$

as $p = \min(\frac{1}{2}, n^{-\frac{2}{D}})$ and $m = O(n^{\frac{2}{D}} \log n)$.

□

Main Theorem

$$\langle X, d\rangle \overset{D}{\hookrightarrow} L_{\infty}^{k=O(Dn^{\frac{2}{D}}\log n)}$$

Proof: For a fix pair of points $x, y \in X$, by the key lemma ,we have that there exists an index $j \in \{1, \cdots, \lceil \frac{D}{2} \rceil\}$ such that if S_{ij} is as chosen in the Algorithm, than $Pr\left[||f(x) - f(y)||_{\infty} \geq \frac{d(x,y)}{D}\right] \geq \frac{p}{12}$.

Moreover, as stated above, that this doesn't hold for all the m choices of S_{ij} is with probability at most $\frac{1}{n^2}$.

Since in all we have $\binom{n}{2}$ pairs of points in X, the probability of failure (for any pair) by the union bound is at most $\frac{1}{2}$.

$\implies$ probability of succeeding is $\geq \frac{1}{2}$

$\square$

Corollaries

Corollary 1

$\langle X, d \rangle \overset{\Theta(\log n)}{\hookrightarrow} L_{\infty}^{O(\log^2 n)}$

Proof: Set $D = \Theta(\log n)$, in the Theorem $\langle X, d \rangle \overset{D}{\hookrightarrow} L_{\infty}^{k=O(Dn^{\frac{2}{D}} \log n)}$ and we obtain $\langle X, d \rangle \overset{\Theta(\log n)}{\hookrightarrow} L_{\infty}^{O(\log^2 n)}$.

$\square$

Corollary 2

$\langle X, d\rangle \overset{\log^2 n}{\hookrightarrow} L_1^{O(\log^2 n)}$

Proof: Let $k = O(\log^2 n)$ be the dimension of embedding.
For a pair of points $x, y \in X$, we have $||f(x) - f(y)||_1 \leq kd(x, y)$ (it holds for each coordinate).

In the Theorem, for a pair $x, y \in X$, we know that there is at least one set which is good, i.e., with probability $\geq 1 - 1/n^2$, $||f(x) - f(y)||_\infty \geq \frac{d(x,y)}{\Theta(\log n)}$.

Extend the machinery in the Theorem to show that with high probability there are $\log n$ sets that are good by choosing slightly larger value for m (but still of order of $O(\log n)$). If this is the case, then
$||f(x) - f(y)||_1 \geq \log n \frac{d(x,y)}{\Theta(\log n)} = \Theta(d(x, y))$
Thus we have $\Theta(d(x, y)) \leq ||f(x) - f(y)||_1 \leq kd(x, y)$, and hence we have a mapping with distortion $O(\log^2 n)$.

$\square$

Corollary 3

$\langle X, d \rangle \overset{\log^{1.5} n}{\hookrightarrow} L_2^{O(\log^2 n)}$

Proof: Let $k = O(\log^2 n)$ be the dimension of embedding. Observe that for the same embedding as in Corollary 1, for a pair of points $x, y \in X$, we have

$$||f(x) - f(y)||_2 = \sqrt{\sum (d(x, S_{ij}) - d(y, S_{ij}))^2} \leq \sqrt{k} d(x, y)$$

We can show,

$$\begin{aligned} ||f(x) - f(y)||_2 &= \sqrt{\sum (d(x, S_{ij}) - d(y, S_{ij}))^2} \\ &\geq \sqrt{\log n \left(\frac{d(x, y)}{\Theta(\log n)} \right)^2} \\ &\geq \frac{d(x, y)}{\Theta(\sqrt{\log n})} \end{aligned}$$

This results in a total distortion of $O(\log^{1.5} n)$.

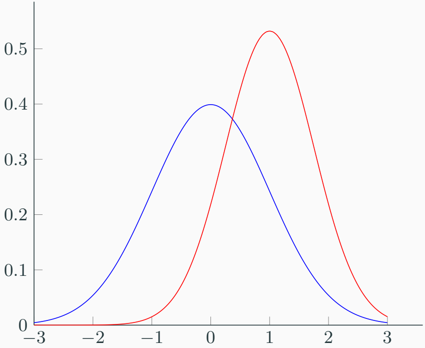
□

Normal Distribution

Normal Distribution

Random variable X has a *Normal Distribution* $\mathcal{N}(\mu, \sigma^2)$, with mean μ and standard deviation $\sigma > 0$, if its probability density function is of the form $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty$

Example: Plot of $\mathcal{N}(0, 1)$ and $\mathcal{N}(1, 0.75)$



Normal Distribution (contd.)

If X has a Normal distribution $\mathcal{N}(\mu, \sigma^2)$, than $aX + b$ has a Normal distribution $\mathcal{N}(a\mu + b, a^2\sigma^2)$, for constants a, b.

The distribution $\mathcal{N}(0, 1)$, with pdf $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, is referred to as the *standardized normal distribution*.

Sum of Normal Distributions

Let X and Y be independent r.v. with Normal distributions $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$. Let r.v. $Z = X + Y$.
Z has a Normal distribution $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
The sum of two independent Normal distributions is a Normal distribution.

L_2 Norm - Johnson-Lindenstrauss Theorem

Johnson-Lindenstrauss Theorem

Let V be a set of n points in d-dimensions. A mapping $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ can be computed, in randomized polynomial time, so that for all pairs of points $u, v \in V$,

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

where $0 < \epsilon < 1$ and n, d, and $k \geq 4(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})^{-1} \ln n$ are positive integers.

Comments:

- The function f maps points of V to a $O(\frac{\ln n}{\epsilon^2})$-dimensional space from a d-dimensional space such that the distortion is within a factor of $1 \pm \epsilon$.
- $|| \cdot ||$ is with respect to Euclidean distance
- Function f is defined in terms of a matrix $A_{k \times d}$ with entries from Normal distribution $\mathcal{N}(0, \frac{1}{k})$.
- A point $v \in \Re^d$ is mapped to the point $v' = Av$. Note that $v' \in \Re^k$.

Matrix with entries from Normal distribution

- Let A be $k \times d$ dimensional matrix, where its entries are chosen independently from $\mathcal{N}(0, \frac{1}{k})$.

- Let x be a vector in R^d.

- Consider the k-dimensional vector Ax

- Next we show that the expected squared length of the vector $||Ax||^2$ is $||x||^2$.

Expected squared length

Lemma 1: $E[||Ax||^2] = ||x||^2$

Proof: Assume $z = Ax$, where $z = (z_1, \ldots, z_k) \in \Re^k$. We want to show that $E[||z||^2] = ||x||^2$.

Note that $||z||^2 = \sum_{i=1}^{k} z_i^2$.

Consider the first coordinate z_1 of z.

Note that $z_1 = \sum_{i=1}^{d} A_{1i} x_i$. What is the distribution of r.v. z_1?

Proof of $E[||Ax||^2] = ||x||^2$ (contd.)

1. Recall that if X has a Normal distribution $\mathcal{N}(0, \sigma^2)$, aX has a Normal distribution $\mathcal{N}(0, a^2\sigma^2)$, for a constant a. Moreover, the sum of two independent r.v. with Normal distributions $\mathcal{N}(0, \sigma_1^2)$ and $\mathcal{N}(0, \sigma_2^2)$ has a Normal distribution $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.
2. Since each A_{1i} is distributed independently by $\mathcal{N}(0, \frac{1}{k})$. The distribution of $z_1 = \sum_{i=1}^{d} A_{1i}x_i$ is the same as the sum of d independent Normal distributions (where each of them have an associated scalar x_i).
3. Thus, z_1 has $\mathcal{N}\left(0, \frac{\sum_{i=1}^{d} x_i^2}{k}\right) = \mathcal{N}\left(0, \frac{||x||^2}{k}\right)$ distribution.
4. Consider $||z||^2 = ||Ax||^2 = z_1^2 + \ldots + z_k^2$, where z_i has $\mathcal{N}(0, \frac{||x||^2}{k})$ distribution.
5. What is $E[||z||^2]$?

Proof of $E[||Ax||^2] = ||x||^2$

1. $E[||z||^2] = E[z_1^2 + \ldots + z_k^2] = kE[z_1^2]$
2. By definition: $Var[z_1] = E[z_1^2] - E[z_1]^2$.
   But z_1 has $\mathcal{N}\left(0, \frac{||x||^2}{k}\right)$ distribution
   $\implies Var[z_1] = \frac{||x||^2}{k}$ and $E[z_1] = 0$.
   $\implies E[z_1^2] = Var[z_1] = \frac{||x||^2}{k}$
3. Therefore, $E[||z||^2] = E[z_1^2 + \ldots + z_k^2] = kE[z_1^2] = ||x||^2$

$\square$

How good is the estimate $E[||Ax||^2] = ||x||^2$?

Is $E[||Ax||^2] = ||x||^2$ a good bound?

Estimate $Pr(||Ax||^2 \geq (1+\epsilon)||x||^2)$ and $Pr(||Ax||^2 \leq (1-\epsilon)||x||^2)$, for $\epsilon \in (0,1)$.

We know that $Pr(||Ax||^2 \geq (1+\epsilon)||x||^2) = Pr(\sum_{i=1}^{k} z_i^2 \geq (1+\epsilon)||x||^2)$, where z_i is a random variable with distribution $\mathcal{N}(0, \frac{||x||^2}{k})$.

Set $Y_i = \frac{\sqrt{k}}{||x||} z_i$.
Since z_i has distribution $\mathcal{N}(0, \frac{||x||^2}{k})$, Y_i has distribution $\mathcal{N}(0,1)$

In the expression $Pr(\sum_{i=1}^{k} z_i^2 \geq (1+\epsilon)||x||^2)$, divide by $\frac{||x||^2}{k}$, and we obtain $Pr(\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k)$.

New Problem

Estimate $Pr(\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k)$, where Y_i has a $\mathcal{N}(0,1)$ distribution.

Estimating $Pr(\sum_{i=1}^{k} Y_i^2)$

Lemma 2

1. $Pr(\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$
2. $Pr(\sum_{i=1}^{k} Y_i^2 \leq (1-\epsilon)k) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$

Proof of 1:

$$
\begin{aligned}
Pr(\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k) &= Pr(e^{\lambda \sum_{i=1}^{k} Y_i^2} \geq e^{(1+\epsilon)\lambda k}) \text{ (for } \lambda > 0\text{)} \\
&\leq \frac{E\left[e^{\lambda \sum_{i=1}^{k} Y_i^2}\right]}{e^{(1+\epsilon)\lambda k}} \text{ (applying Markov's Inequality)} \\
&= \frac{E\left[e^{\lambda Y_1^2}\right]^k}{e^{(1+\epsilon)\lambda k}} \text{ (Independence of } Y_i\text{'s)}
\end{aligned}
$$

A useful identity

An Identity

Let X be a random variable distributed $\mathcal{N}(0,1)$ and $\lambda < \frac{1}{2}$ be a constant. Then, $E\left[e^{\lambda X^2}\right] = \frac{1}{\sqrt{1-2\lambda}}$

Proof: PDF of standard normal distribution is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

By definition, $E[H(x)] = \int\limits_{-\infty}^{+\infty} H(x) f(x) dx$

Thus, $E\left[e^{\lambda X^2}\right] = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{\lambda x^2} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-(1-2\lambda)\frac{x^2}{2}} dx$

Substitute $y = x\sqrt{1-2\lambda}$, and we obtain

$$E\left[e^{\lambda X^2}\right] = \frac{1}{\sqrt{1-2\lambda}} \left[\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy \right]$$

But, $\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = 1$, as this is the area under the Normal distribution curve.

$\square$

Proof of $Pr(\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$ (contd.)

We have

$Pr(\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k) \leq \frac{E\left[e^{\lambda Y_1^2}\right]^k}{e^{(1+\epsilon)\lambda k}} = e^{-(1+\epsilon)k\lambda}\left(\frac{1}{\sqrt{1-2\lambda}}\right)^k$ (using the identity)

Set $\lambda = \frac{\epsilon}{2(1+\epsilon)}$ and we have

$$
\begin{aligned}
Pr(\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k) &\leq e^{-(1+\epsilon)k\lambda}\left(\frac{1}{\sqrt{1-2\lambda}}\right)^k \\
&= e^{-\frac{\epsilon}{2}k}(1+\epsilon)^{\frac{k}{2}} \\
&= \left((1+\epsilon)e^{-\epsilon}\right)^{\frac{k}{2}} \\
&\leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)} \text{ (as } 1+\epsilon \leq e^{\epsilon-\frac{\epsilon^2-\epsilon^3}{2}}\text{)}
\end{aligned}
$$

This finishes the proof of the 1st part of Lemma 2. The proof of 2nd part is similar and is left as an exercise.

□

Estimating $Pr(\sum_{i=1}^{k} Y_i^2)$

Corollary 1

If $k = c\frac{\ln n}{\epsilon^2}$, for some constant $c > 4$,

$Pr((1-\epsilon)k \leq \sum_{i=1}^{k} Y_i^2 \leq (1+\epsilon)k) \geq 1 - \frac{1}{n^3}$

Proof: From Lemma 2 we have that

$Pr(\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$ and $Pr(\sum_{i=1}^{k} Y_i^2 \leq (1-\epsilon)k) \leq e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$.

Hence $Pr\left((\sum_{i=1}^{k} Y_i^2 \geq (1+\epsilon)k) \vee (\sum_{i=1}^{k} Y_i^2 \leq (1-\epsilon)k)\right) \leq 2e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$ (by Union Bound)

Thus, $Pr((1-\epsilon)k \leq \sum_{i=1}^{k} Y_i^2 \leq (1+\epsilon)k) \geq 1 - 2e^{-\frac{k}{4}(\epsilon^2-\epsilon^3)}$

Substituting, $k = c\frac{\ln n}{\epsilon^2}$ we have that

$Pr((1-\epsilon)k \leq \sum_{i=1}^{k} Y_i^2 \leq (1+\epsilon)k) \geq 1 - \frac{1}{n^3}$ (bit sloppy computation)

$\square$

J-L Theorem

Let V be a set of n points in d-dimensions. A mapping $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ can be computed, in randomized polynomial time, so that for all pairs of points $u, v \in V$,

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

where $0 < \epsilon < 1$ and n, d, and $k \geq 4(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})^{-1} \ln n$ are positive integers.

By choosing matrix $A_{k\times d}$ consisting of independent values from $\mathcal{N}(0, \frac{1}{k})$, we show that $\forall u, v \in V$
$Pr((1-\epsilon)||u-v||^2 \leq ||Au - Av||^2 \leq (1+\epsilon)||u-v||^2) \geq 1 - \frac{1}{n}$

Proof of J-L Theorem

Proof: By Corollary 1, we know that for any vector $x \in R^d$,
$Pr((1-\epsilon)||x||^2 \leq ||Ax||^2 \leq (1+\epsilon)||x||^2) \geq 1 - \frac{1}{n^3}$

Consider any pair of points $u, v \in V$. Set $x = u - v$. Then
$Pr((1-\epsilon)||u-v||^2 \leq ||A(u-v)||^2 \leq (1+\epsilon)||u-v||^2) \geq 1 - \frac{1}{n^3}$

There are in all $\binom{n}{2}$ pairs of points in V.
By union bound, we have that $\forall u, v \in V$
$Pr((1-\epsilon)||u-v||^2 \leq ||Au - Av||^2 \leq (1+\epsilon)||u-v||^2) \geq 1 - \frac{1}{n}$

$\square$

Comments

1. Choice of matrix A doesn't depend on points in V
2. What properties A needed to satisfy?
3. $E[||Ax||^2] = ||x||^2$
4. A is dense $\implies$ Av takes more computation time
5. Can we find sparse matrix A?
   Choose entries of A from $\{-1, 1, 0\}$ with probabilities $1/6$,$1/6$, and $2/3$, respectively and normalize.
6. $\ldots$

References

1. Johnson and Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, Contemporary Mathematics 26:189-206, 1984.
2. Achlioptas, Database-friendly random projections, JCSS 66(4): 671-687, 2003.
3. Dasgupta, and Gupta, An elementary proof of a theorem of Johnson and Lindenstrauss" Random Structures & Algorithms, 22 (1): 60-65, 2003.
4. Dubhashi and Panconesi, Concentration of Measure for the Analysis of Randomized Algorithms, Cambridge University Press, 2009.
5. Matousek, Lectures on Discrete Geometry, Volume 212 of Graduate Texts in Mathematics. Springer, New York, 2002.
6. Ankush Moitra Notes