1 Overview

Sometimes it is easier to solve a problem in a given $d$-dimensional space by transforming it into a problem in some other $k$-dimensional space where $k < d$ or $k > d$. In this lecture we explore dimension expansion and reduction techniques and prove the Johnson-Lindenstrauss theorem from real analysis about such transformations in Euclidean space and $\mathbb{R}^n$.

2 Motivating example

Find the diameter (largest distance between 2 points) for a given set of points in $\mathbb{R}^d$ space under $l_1$ norm. $l_1$ norm is given by 
\[ d(x, y) = |x_2 - x_1| + |y_2 - y_1| \]
For $n$ points in $\mathbb{R}^d$ the time required to compute the $l_1$ norm is $O(dn^2)$ (use a simple double loop to compute the distance between all pairs of points $(x, y)$ in the given space). If we transform the problem into one in $k$-dimension where $k > d$ (details of the transformation are given in Section 3.8), the time complexity is $O(nd2^d)$, which is a more efficient solution than the earlier one, for $d \ll n$.

3 Definitions

3.1 Metric space

Let $X$ be a set. Let $d : X \times X \rightarrow \mathbb{R}$.
$(X, d)$ is a metric space if for all $x, y, z \in X$
\begin{itemize}
  \item[i)] $d(x, y) \geq 0$
  \item[ii)] $d(x, y) = 0$ if and only if $X = Y$
  \item[iii)] $d(x, y) = d(y, x)$
\end{itemize}
iv) \( d(x, y) + d(y, z) \geq d(x, z) \)

If \( X \) is finite, \((X, d)\) is a finite metric space.

### 3.2 Normed space

A vector space on which a norm is defined is called a normed space.

\( V \) is a vector space over a field \( F \). A norm is a function from \( V \) to \( \mathbb{R} \) that satisfies for all \( \alpha \) in \( K \) and all \( \vec{u}, \vec{v} \in V \),

i) \( \|\vec{v}\| \geq 0 \) and equality if \( \vec{v} = 0 \)

ii) \( \|\alpha \vec{v}\| = |\alpha|\|\vec{v}\| \)

iii) \( \|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\| \)

### 3.3 Cauchy sequence

A sequence \( < x_n > \) is said to be a Cauchy sequence, if for any \( \varepsilon > 0 \), there exists an integer \( N \) s.t
\( d(x_n, x_m) < \varepsilon \), for all integers \( m, n > N \)

### 3.4 Complete space

A metric space \((X, d)\) is said to be complete if every Cauchy sequence converges to some point in that space.

### 3.5 Banach space

If a normed space is complete and \( \|x - y\| = d(x, y) \) then it is a Banach space.

### 3.6 Hilbert space

A Hilbert space is an inner product space which is complete. A inner product space is a vector space \( V \) with an inner product i.e. a function \( \langle \cdot, \cdot \rangle: X \times X \rightarrow \mathbb{R} \) satisfying

i) \( \langle \alpha_1 x_1 + \alpha_2 x_2, y \rangle = \alpha_1 \langle x_1, y \rangle + \alpha_2 \langle x_2, y \rangle \)

ii) \( \langle x, y \rangle = \langle y, x \rangle \)

iii) \( \langle x, x \rangle = \|x\|^2 \)

for all \( x, y, \alpha \in V \).

Therefore, a Hilbert space is a Banach space whose norm is determined by the inner product.
3.7 Embedding

Given two metric spaces \((X, d)\) and \((X', d')\), we can define an embedding, \(f : X_1 \to X_2\)

The embedding \(f\) is **isometric** if \(d(x, y) = d'(f(x), f(y))\).

3.8 Example

We now demonstrate a dimension expansion technique for solving example of Section 2. We define an embedding from the given set of points to a hypercube in \(d -\)dimensions and compute the diameter under \(L_1\) norm in the hypercube. The embedding is given by:

\[
(f(p) = < \bar{s}_0, \bar{s}_1, \ldots, \bar{s}_{2^d - 1}, \bar{p} > \text{ where } \bar{p} \in L_1 \text{ and } \bar{s}_i = \{-1, +1\}^d.
\]

We compute the diameter under \(L_\infty\) norm as shown below:

\[
\max_{x, y \in S} ||x - y||_\infty = \max_{x, y \in S} \max_{i=1}^n |x_i - y_i| = \max_{i=1}^n \max_{x, y \in S} |x_i - y_i|
\]

The embedding is also isometric i.e. diameter in \(L_1\) norm = diameter in \(L_\infty\) norm as shown below:

\[
\|f(p) - f(q)\|_\infty = \|f(p - q)\|_\infty
\]

\[
= \max_s |s \cdot (p - q)|
\]

by defn. of \(f\) and \(L_\infty\)

\[
= \max_s |\sum_{j=1}^{k'} s_j \cdot (p - q)_j|
\]

\[
= \sum_j |(p - q)_j|
\]

consider only positive \(s_j\)

\[
= \|p - q\|_1
\]

Complexity: \(L_\infty\) computation takes time \(O(n.2^d.d)\) as computation of \(f(p)\) for each point takes \(d2^d\) time and there are \(n\) such points. This transformation performs better compared to directly computing \(L_1\) norm (\(O(d.n^2)\)) when \(n\) is very large.

3.9 Distorted embedding

The embedding \(f : (X, d) \to (X', d')\) is distorted if

\[
r d(x, y) \leq d'(f(x), f(y)) \leq \alpha r d(x, y) \text{ where } \alpha \geq 1, r > 0
\]

where \(\alpha\) is distortion factor, \(r\) is scaling factor.
4 Bourgain’s Theorem

For any metric space \((X, d)\), there exists an embedding \(f : (X, d) \rightarrow l_2^{O(\log n)}\) with distortion \(= O(\log n)\).

5 Johnson-Lindenstrauss Theorem

Let \((X, d) = \mathbb{R}^n = L_2^n\). Then there exists an embedding \(\phi : \mathbb{R}^n \rightarrow \mathbb{R}^k\) with \(k = O\left(\frac{\log n}{\epsilon^2}\right)\) such that for some \(p\),

\[
(1 - \epsilon) \leq \frac{\|\phi(x) - \phi(y)\|_2^2}{\rho\|x - y\|_2^2} \leq (1 + \epsilon)
\]

Proof: Choose \(\tilde{r}^{(1)}, \ldots, \tilde{r}^{(n)} \in \mathbb{R}^n\) with \(\tilde{r}^{(i)} \sim N(0, 1)\). We claim that \(\tilde{\phi}(\tilde{v}) = \langle \tilde{v}, \tilde{r}^{(1)} \rangle, \ldots, \langle \tilde{v}, \tilde{r}^{(n)} \rangle\) is an embedding satisfying 1.

Given a vector \(v = (\tilde{v}_1, \ldots, \tilde{v}_n) \in \mathbb{R}^{(n)}\) and a random vector \(r = (\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_n) \sim N(0, 1)\), let \(X = (\tilde{v}, \tilde{r})\). Since a linear combination of normal random variables is also normal, \(X\) is a normal random variable with variance \(l^2\) where \(l = \|\tilde{v}\|_2\).

\[
E[X] = E[\sum_{i=1}^n v_i r_i] = \sum_{i=1}^n v_i E[r_i] = 0 \text{ and } \ V ar[X] = \sum_{i=1}^n V ar[v_i r_i] = \sum_{i=1}^n v_i^2 V ar[r_i] = \sum_{i=1}^n v_i^2 = l^2
\]

\[
E[\|\tilde{\phi}(\tilde{v})\|_2^2] = E[\sum_{j=1}^k \langle \tilde{v}, \tilde{r}^{(j)} \rangle^2] = \sum_{j=1}^k E[\|< \tilde{v}, \tilde{r}^{(j)} >\|^2] = \sum_{j=1}^k E[< \tilde{v}, \tilde{r}^{(j)} >^2] = \sum_{j=1}^k E[\sum_{i=1}^n v_i^2 r_i^{(j)} + 2 \sum_{i \neq i'} v_i v_i' r_i^{(j)} r_i'^{(j)}] = \sum_{j=1}^k (l^2 + \sum_{i \neq i'} v_i v_i' E[r_i] E[r_i']) = kl^2
\]

by independence of \(r_i\) and \(r_{i'}\)

\[
\text{as mean for each } r_i = 0
\]
We show that if \( k = O\left(\frac{\log n}{\epsilon^2}\right) \) then \( \|\phi(v)\|_2^2 \simeq (1 + \epsilon)k\|
\) with probability \( \geq 1 - \frac{1}{n^2} \).

Since \( \phi \) is linear (as its a projection onto a \( k \)-dimensional subspace spanned by the \( r^\dagger \)'s), we have \( \frac{\|\phi(x) - \phi(y)\|_2^2}{k\|x - y\|_2^2} = \frac{\|\phi(x) - \phi(y)\|_2^2}{k\|x - y\|_2^2} \).

Let \( X = \sum_{i=1}^{k} X_i^2 \) where \( X_i = \langle \hat{v}, r^{\dagger}_i \rangle \) and \( X_i \sim N(0,1) \) and \( \hat{v} \) is original vector but whose length has been rescaled by dividing it by \( l^2 \). Therefore, \( E[X] = k \) from Equation 2.

Let \( \alpha = (1 + \epsilon)k \)

\[
Pr[X \geq \alpha] = Pr[e^{sX} \geq e^{s\alpha}, \text{for all } s \geq 0] = Pr[e^{sX} e^{-s\alpha} \geq 1] \leq E[e^{sX} e^{-s\alpha}]
\]

Using Markov’s inequality \( Pr[X \geq t] \leq E[X]/t, X \geq 0 \)

\[
= e^{-s\alpha} E[e^{s\sum_i X_i^2}] = e^{-s\alpha} E[\prod_i e^{sX_i^2}] = e^{-s\alpha} \prod_i E[e^{sX_i^2}] \tag{3}
\]

Now we compute \( E[e^{sX_i^2}] \)

Since \( X_i \sim N(0,1), Pr[|X_i| \geq x] = 2 \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-(t^2/2)} dt \)

So \( |X_i| \sim \frac{2}{\sqrt{2\pi}} e^{x^2/2} \)

Let \( Y = |X_i|^2 \). By change of variable formula, we have \( Y \sim \frac{1}{\sqrt{2\pi}} e^{-(y/2)}y^{-1/2} \)

Computing the moment generating function of \( Y \)

\[
= E[e^{sY}] = \int_0^\infty e^{sy} \frac{1}{\sqrt{2\pi}} e^{-(y/2)}y^{-1/2} dy = \frac{1}{\sqrt{1 - 2s}} \text{ (defined for } -\infty < s < \frac{1}{2} )
\]

Integrating by parts twice

Substituting in Equation 3

\[
Pr[X \geq \alpha] \leq e^{-s\alpha}(1 - 2s)^{-\frac{k}{2}} \tag{4}
\]

Minimizing rhs as a function of \( s \) (to get tightest bound) we get

\[
s = \frac{1}{2} \left(\frac{1}{\alpha} - \frac{k}{2}\right) = \frac{1}{2} \left(1 - \frac{k}{(1 + \epsilon)k}\right) \tag{5}
\]
From 4 and 5, we have

\[
Pr[X \geq \alpha] \leq e^{-s\alpha}(1 - 2s)^{-\frac{k}{2}} \\
= e^{-\frac{1}{2}(a-k)}(\frac{\alpha}{k})^{\frac{k}{2}} \\
= e^{-\frac{k}{2}}(1 + \epsilon)^{k/2} \\
= \left(\frac{e^\epsilon}{1+\epsilon}\right)^{(-k/2)} \\
= \exp(-\frac{k}{2}(\epsilon - \log(1 + \epsilon))) \\
\leq \exp(-\frac{k}{2}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}))
\]

using Taylor’s series expansion.

Hence if \( k = \frac{4\log n}{(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})} \), we have \( Pr[X \geq (1 + \epsilon)k] \leq \frac{1}{n^2} \).

Similarly, we can prove \( Pr[X \leq (1 - \epsilon)k] \leq \frac{1}{n^2} \).