1 Overview

Motivation: computation in the context of data in a high dimensional space. For example:

\[ \langle u, v \rangle = \sum_{i=1}^{n} u_i v_i \]

This may be costly given the dimensionality. We run into even more problematic run times when we wish to transform vectors by a matrix, or multiply matrices:

\[
Mu \quad O(n^2) \\
MN \quad O(n^3)
\]

Many other matrix computations, most take \(O(n^3)\) time.

Goal: Reduce the dimension of data, while preserving useful properties.

Today: In particular, we will try to preserve the norms of and distance between data point vectors.

2 The Problem

Problem: given vectors \(x_1, x_2, \ldots, x_n \in \mathbb{R}^m\), find low dimensional vectors \(y_1, \ldots, y_n \in \mathbb{R}^d\), where \(d \ll n, m\), such that the following approximately hold:

\[ \|y_i\| \approx \|x_i\| \quad \|y_i - y_j\| \approx \|x_i - x_j\| \]

Remark 1. The norm is always assumed to be \(\ell_2\).

Remark 2. Without loss of generality, assume \(m \leq n\), since we can always reconsider the problem as taking place in the span of the relevant vectors \((x_i)_{i=1}^{n}\).

There are situations when preserving exact distance is impossible: for example consider data points that form the vertices of a regular tetrahedron, and hence are all mutually equidistant, in \(\mathbb{R}^3\). It is impossible, however, to project 4 points in \(\mathbb{R}^2\) with mutual equidistance. To see this, consider an equilateral triangle, and draw a circle around each point corresponding to the side-length: there are no points in the three way intersection of these circles. Thus we must settle for a close approximation instead of precise equality.

3 Classical Dimension Reduction

Classical dimension reduction: Principle Component Analysis (PCA).

Idea: given data, compute the covariance matrix
Then compute the top $d$ eigenvectors $v_1, \ldots, v_d$ of $M$, and project all data on to the span of these.

PCA is not good for our problem. It works by averaging data and hence cannot handle outliers well, thus destroying their distance from other vectors.

We are going to instead try a randomized algorithm.

## 4 Johnson-Lindenstrauss

**Theorem 1.** *(Johnson-Lindenstrauss)* For arbitrary points $x_1, \ldots, x_n \in \mathbb{R}^m$ and any $\varepsilon \in (0, \frac{1}{2})$, there exist $y_1, \ldots, y_n \in \mathbb{R}^d$, where $d = O\left(\frac{\log n}{\varepsilon^2}\right)$, such that

\[
\forall i : (1 - \varepsilon)\|x_i\|^2 \leq \|y_i\|^2 \leq (1 + \varepsilon)\|x_i\|^2
\]

\[
\forall i, j : (1 - \varepsilon)\|x_i - x_j\|^2 \leq \|y_i - y_j\|^2 \leq (1 + \varepsilon)\|x_i - x_j\|^2
\]

We note the following

- This is a multiplicative guarantee.
- The $\{y_i\}$ can be computed efficiently.
- The mapping $x_i \mapsto y_i$ can be linear and does not depend on $\{x_i\}$'s.

## 5 Probabilistic Method

Proof Idea: We will prove this via the so called probabilistic method, described as follows.

Suppose we want to prove that a subset $S$ consisting of vectors $y_i$ that satisfy some criteria—e.g. of the above theorem—is not empty. We can do so by designing a distribution $X$ and proving:

\[
\Pr[X \in S] > 0.
\]

## 6 Random Projection

To prove our theorem, we will use a random projection as the distribution with which we will apply the probabilistic method.

**Lemma 2.** Let $x \in \mathbb{R}^n$ and define $A \in \mathbb{R}^{d \times n}$ by $A_{i,j} \sim N(0,1)$ iid. Then

\[
\mathbb{E}[\|Ax\|^2] = d\|x\|^2
\]

\[
\Pr[(1 - \varepsilon)\|x\|^2 \leq \frac{1}{\sqrt{d}}\|Ax\|^2 \leq (1 + \varepsilon)\|x\|^2] \geq 1 - 2e^{-\frac{\varepsilon^2 d}{8}}
\]

We now show that this lemma implies the theorem.

**Proposition 3.** *Johnson-Lindenstrauss can be proven from Lemma 2.*
Proof. Let $A \in \mathbb{R}^{d \times n}$, $A_{i,j} \sim N(0, 1)$, let $y_i = \frac{1}{\sqrt{d}}Ax_i$. For any fixed pair $(i, j)$, we have

$$
\|y_i - y_j\| = \|\frac{1}{\sqrt{d}}Ax_i - \frac{1}{\sqrt{d}}Ax_j\|^2 = \|\frac{1}{\sqrt{d}}A(x_i - x_j)\|^2
$$

Apply Lemma 2 with $x = x_i - x_j$ and choose $d > \frac{16\log n}{\epsilon^2}$:

$$
\Pr[(1 - \epsilon)\|x_i - x_j\|^2 \leq \|\frac{1}{\sqrt{d}}A(x_i - x_j)\|^2 \leq (1 + \epsilon)\|x_i - x_j\|^2] \geq 1 - 2e^{-\frac{\epsilon^2 d}{8}} \geq 1 - \frac{2}{n^2}
$$

Since there are only $\binom{n}{2}$ pairs, we may apply a union bound:

$$
\Pr[\forall i, j : (1 - \epsilon)\|x_i - x_j\|^2 \leq \|y_i - y_j\|^2 \leq (1 + \epsilon)\|x_i - x_j\|^2] \geq (1 - \frac{2}{n^2})\binom{n}{2} > 0
$$

By the probabilistic proof method, this implies the theorem. \[\square\]

Remark 3. $d$ is tight up to additional log factors.

7 Proving the Lemma

Recall, if $x \sim N(0, \sigma^2), y \sim N(0, \tau^2)$:

$$
c \cdot x \sim N(0, c^2 \sigma^2)
$$

$$
x + y \sim N(0, \sigma^2 + \tau^2)
$$

We will use this to prove properties of $Ax \in \mathbb{R}^d$:

$$
v_1 = (Ax)_1 = \sum_{i=1}^{n} A_{1,i}x_i 
$$

$$
\sim N(0, \sum_{i=1}^{n} x_i^2) = N(0, \|x\|^2)
$$

Then

$$
\mathbb{E}[\|Ax\|^2] = \mathbb{E}[\sum_{i=1}^{d} (Ax)_i^2]
$$

$$
= \sum_{i=1}^{d} \mathbb{E}[(Ax)_i^2]
$$

$$
= \sum_{i=1}^{d} \|x\|^2
$$

$$
= d\|x\|^2
$$
We would like a concentration inequality for \( \sum_{i=1}^{d} v_i^2 \), where \( v_i \sim N(0, 1) \). Note that we cannot use Bernstein’s inequality, because it requires a fixed bound \( |v_i^2| < R \) for some \( R \). We solve this by first noting that \( v_i^2 \) is \( \chi^2 \) distributed, and then we employ a truncation argument. More formally, define

\[
  z_i = \begin{cases} 
  v_i^2 & v_i^2 \leq 20 \log n \\
  0 & \text{otherwise}
  \end{cases}
\]

We call \( z_i \) a truncation of \( v_i^2 \). We then have

\[
  |z_i| \leq R = 20 \log n \\
  \mathbb{E}z_i^2 \leq \mathbb{E}v_i^4 = 3.
\]

We may then apply Bernstein’s inequality to yield the following concentration bound.

\[
  \Pr \left[ \left| \sum_{i=1}^{d} z_i - \mathbb{E} \left( \sum_{i=1}^{d} z_i \right) \right| \geq t \right] \leq \exp \left( -\frac{t^2}{d + Rt} \right) = \exp \left( -\frac{\varepsilon^2 d}{d + Rd\varepsilon} \right).
\]

Note that the final equality is given by using \( t \approx \varepsilon d \).

We note that this is a good approximation because we have the following:

\[
  \Pr[z_i = v_i^2] \geq 1 - n^{-10},
\]

and hence:

\[
  \Pr[\forall i = 1, \ldots, d : z_i = v_i^2] \geq 1 - dn^{-10}.
\]

Combining this with our result proves Lemma 2.