Dimension Reduction:

Last lecture we talked about structured data in high dimensional space, and algorithms to solve near-neighbor searches on it. The efficacy of the algorithms depend on the dimension of the space — so the natural question is whether we can reduce the dimension of the data (without affecting its other important properties).

In this lecture and the next:

- we will see two procedures to reduce the dimension.
  - today we will reduce dim and maintain distances approx.
  - next time we will maintain most of the "energy" or "variance" of the data (precise defn next time).

Today's menu:

- Recap of Gaussian random variables
- Some properties of high dimensional IRd
- Intuition and construction of projection
- Sketch of proof.
Gaussian / Normal Random Variables

Continuous random variable
denoted \( N(\mu, \sigma^2) \)
mean \( \mu \)
variance \( \sigma^2 \)

Density function
\[
f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

Most mass within \( \mu \pm \sigma(\sigma) \)

Useful facts:
(a) \( X_1 \sim N(\mu_1, \sigma^1_1) \Rightarrow X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma^1_1 + \sigma^2_2) \).

(b) \( X \sim N(\mu, \sigma^2) \Rightarrow cX \sim N(c\mu, c^2\sigma^2) \).

\[
\Rightarrow \sum a_i X_i \sim \sum N(a_i \mu, a_i^2 \sigma_i^2) = N(\sum a_i \mu, \sum a_i^2 \sigma_i^2)
\]

In particular if \( \mu_i = 0 \) \( \forall i \)
\( \sigma_i^2 = 1 \)

then \( \sum a_i X_i = N(0, \sum a_i^2) = N(0, \|a\|^2) \)

Euclidean length of \( a \).
High Dimensional Facts

Fact 1: In a Euclidean space $\mathbb{R}^d$, the volume of the ball is proportional to $(\text{radius})^d$ and surface area is prop. to $(\text{rad})^{d-1}$.

Analogous to $\mathbb{R}^2$, where volume of a ball of radius $r$ is $\pi r^2$, surface area is $2\pi r$.

---

Weird Fact 1: Most of volume is near the boundary.

Note: Volume of a shrink ball of radius $(1-\varepsilon)r$ is $\sim (1-\varepsilon)^d r^d$.

\[
\Rightarrow \text{ratio of volumes} = \frac{(1-\varepsilon)r^d}{r^d} = (1-\varepsilon)^d = \frac{1}{2},
\]

if $\varepsilon \approx \frac{1}{2}$.

So half the volume is within a shell of width $\frac{1}{2}d$ around the surface.

---

Weird Fact 2: Most of volume is near the equator.

Let $A(t)$ be "area" $(n-1)$ dimensional volume of slice at distance $t$ from equator.

Then \[
\frac{A(t)}{A(0)} = \left(\frac{1-t^2}{1-\varepsilon^2}\right)^{n-1} = (1-\varepsilon^2)^{n-1} \leq e^{-2\varepsilon^2/n}.
\]

So at distance $t = \frac{\sqrt{1-\varepsilon^2}}{\sqrt{n-1}}$, ratio $\leq \frac{1}{e}$ and falling rapidly.

In fact if $f_N(t)$ = density function of $N(0, \frac{1}{n-1})$ distribution $\Rightarrow e^{-t^2/2n}\leq f_N(t)$.

\[
\Rightarrow \frac{A(t)}{A(0)} \leq \frac{f_N(t)}{f_N(0)} \quad \text{And we know that most of the mass is within }\pm \frac{\text{rad}}{\sqrt{n-1}} \text{ of the origin.}
\]
Interestingly, most of mass of ball is close to surface and is close to the equator.

$\Rightarrow$ most of surface must be close to equator.

E.g.

Of course, we chose the equator arbitrarily. So most of mass is close to every central hyperplane.

Difficult to draw. Just means should be careful with intuition of what a ball is, high dimensional bodies are sometimes misleading.

However, have some good tools to argue about high dimensions.

1) Projections — project the body down to lower dimensions and talk about these projections.

Indeed — tool of choice today.
Main Result of Today: Dimension Reduction in Euclidean Space

Every set of \( n \) points can be mapped to \( O(\log n) \) dimensions without changing distances by more than \( O(1) \) factor.

Formally: \( O\left(\frac{\log n}{\varepsilon^2}\right) \) dimensions, \( \|f(x) - f(y)\| \leq (1 + \varepsilon) \|x - y\| \) for all original points.

Given \( 10^{10} = 2^{30} \) points, can map to \( O\left(\frac{\log n}{\varepsilon^2}\right) \) dimensions.

This is theoretical limit, can often do better than worst-case bound.

Also, can combine with other dimension reduction techniques.

Theorem [Johnson–Lindenstrauss 1984]

Given \( S \subseteq \mathbb{R}^D \) of \( |S| = n \) size, \( S = \{v_1, v_2, \ldots, v_n\} \)

\( \exists \) a map \( f: \mathbb{R}^D \to \mathbb{R}^k \) with \( k = O\left(\frac{\log n}{\varepsilon^2}\right) \) such that

\[
1 - \varepsilon \leq \frac{\|f(v_i) - f(v_j)\|^2}{\|v_i - v_j\|^2} \leq 1 + \varepsilon \quad \forall i, j \in [n]
\]

In fact: Also is an algorithm, can be made efficient.

Here is the map.

Choose \( A \in \mathbb{R}^{k \times D} \) randomly: \( A = \frac{1}{\sqrt{k}} u \cdot \mathbb{N}(0,1) \) indep.

- Choose \( A \in \mathbb{R}^{k \times D} \) randomly: \( \quad A = \frac{1}{\sqrt{k}} u \cdot \mathbb{N}(0,1) \) indep.
- \( \forall i, j \in [n] \)
- \( \|f(v_i) - f(v_j)\|^2 \leq (1 + \varepsilon) \|v_i - v_j\|^2 \)

\( \|f(v_i) - f(v_j)\|^2 \leq (1 + \varepsilon) \|v_i - v_j\|^2 \) for all original points.
Very simple. (and can speed it up further, more later).

Why should it work?

- Let's show that distance between \( v_1 \) and \( v_2 \) maintained.

Define \( a = v_1 - v_2 \). Show that \( \|f(a)\| = \|f(v_1 - v_2)\| = \|f(v_1) - f(v_2)\| \) because linear map.

- Consider case when \( a = (a, 0, 0, \ldots, 0) \) axis aligned vector.

then \( f(a) = \frac{1}{\sqrt{k}} Aa = \frac{1}{\sqrt{k}} \begin{pmatrix} A_1 & A_2 & \cdots & A_D \end{pmatrix} \begin{pmatrix} a \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \frac{1}{\sqrt{k}} A_1 a \) first column of \( A \) all others are multiplied by 0.

So \( \|f(a)\|^2 = \frac{1}{k} \sum_{i=1} N(0, a^2) \) average of squares of \( N(0, a^2) \) random vars.

\[ E\left[ \text{square of } N(0, a^2) \right] \approx \text{Var}(N(0, a^2)) \]

\[ \approx \text{Var}(N(0, a^2)) \] with high prob. but \( \text{Var}(N) = E[N^2] - E[N]^2 \) \[ = E[N^2] \] \[ \Rightarrow \|f(a)\|^2 \approx \|a\|^2 \]
General case: $A = V_i - V_j$ is not axis aligned.

(a) Can rotate the world. == (Argo is spherically symmetric)

or (b) show still $Aa \sim \begin{pmatrix} N(0, \|a\|_2^2) \\ N(0, \|a\|_2^2) \end{pmatrix}$

using that $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

To recap: Build $A = \text{gaussian matrix}$ $f(x) = \frac{1}{\sqrt{k}} Ax$.

then for each vector $V_i - V_j$,

$\|f(V_i) - f(V_j)\| = \|f(V_i - V_j)\| \leq \|V_i - V_j\|$  

with high probability.

Do union bound over all $(\frac{n^2}{2})$ pairs.

Extensions: Can use $A = \Theta(\sqrt{k} \times D)$ instead of Gaussians.

: Matrix mult takes $O(kDn)$ time

Can make it faster — see Fast Johnson-Lindenstrauss Transform.

Uses: to preprocess data, bring it to smaller dimension.

For example: say we want to multiply matrices $M_1^T M_2$.

But OK with approx solutions.

then consider $(\frac{1}{\sqrt{k}} A M_1)^T (\frac{1}{\sqrt{k}} A M_2)$. 
Show in HW that gives good approximations.

- Takes time $KD^2$ to do dim reduction.
  - Matrix $k \times D$ x $(D \times D)$
  - and then multiply two matrices
    - $(D \times k)$ times $(k \times D) = O(KD^2)$
    - again.

Much better if $D \gg K = O\left(\frac{\log D}{\varepsilon^2}\right)$.

- General Idea of Dimension Reduction
  - useful in other contexts as well.

  E.g. next lecture: Compressive Sensing.
  - Want to learn sparse signal using few linear measurements.
  - Brings together dimension reduction and next topic: optimization LP.

---

$x$