CSE520 Computational Geometry
Lecture 26
Dimension Reduction

Antoine Vigneron

Ulsan National Institute of Science and Technology

June 15, 2020

Course Organization

- Today, I present a technique for dealing with high dimensional data.
- I will also review a few notions of linear algebra that will be needed.
- I will not give all the proofs.

References

- Reference: Sariel Har-Peled's book, chapter 19.
- Jiří Matoušek's textbook *Lectures on Discrete Geometry*. (Available online through UNIST library.)

Introduction

- So far we studied problems in 2D or in fixed dimension $d = O(1)$.
- In 2D, we saw various techniques (plane sweep, randomized incremental construction, duality).
- In fixed dimension, we studied range trees, segment trees, quadtrees.
- They yield algorithms whose time bounds have an exponential dependency on d, i.e. $\log^d n$ or 2^d.

- Problem: What can we do when d is large, for instance $d = 100$?
- In high dimension, algorithms often do not have good worst-case time bound.
- In this lecture, we will present one way around this problem: *dimension reduction*.
- But first we will study the geometry of high dimensional spaces.

Introduction

- In this lecture, n will denote the dimension of the space (instead of d).
- So we work in $\mathbb{R}^n$.
- High-dimensional data naturally arises in many applications.

Examples

- The space of 1000×1000 images has dimension 10^6. Each coordinate is the color of a pixel $p[i, j]$.
- DNA sequences have millions of pairs of bases, each of them can be seen as a dimension.
- A weighted graph, where each edge e has a weight $w(e)$, can be regarded as a space with dimension $|E|$.
- Medical records.

Volume of Balls

- Let $\mathbf{b}^n = \{p \in \mathbb{R}^n \mid \|p\| \leqslant 1\}$ be the unit ball centered at the origin.
- Its boundary is the *sphere* $\mathbb{S}^{n-1} = \{p \in \mathbb{R}^n \mid \|p\| = 1\}$.
- Then $r\mathbf{b}^n$ is the radius-r ball centered at the origin.

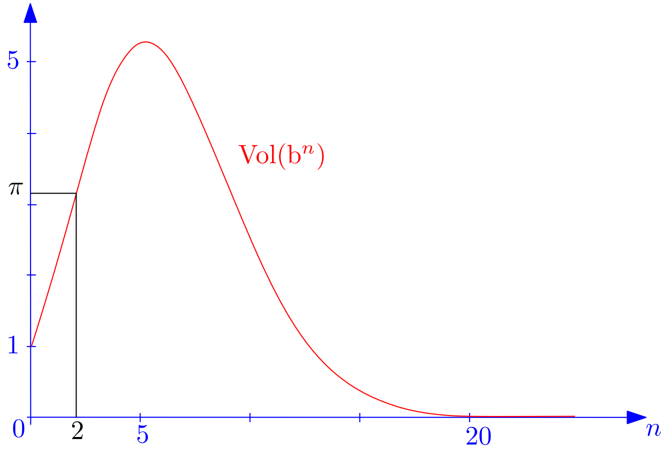
- Let $\mathrm{Vol}(r\mathbf{b}^n)$ denote the volume of this ball and let $\mathrm{S}(r\mathbf{b}^n)$ denote the surface area of $r\mathbb{S}^{n-1}$.
- For instance, $\mathrm{Vol}(r\mathbf{b}^3) = \frac{4}{3}\pi r^3$ and $\mathrm{S}(r\mathbf{b}^3) = 4\pi r^2$.

- Formula:

$$\mathrm{Vol}(r\mathbf{b}^n) = \frac{\pi^{n/2} r^n}{\Gamma(n/2+1)} \qquad \mathrm{S}(r\mathbf{b}^n) = \frac{2\pi^{n/2} r^{n-1}}{\Gamma(n/2)}$$

  where $\Gamma(n/2+1) = (n/2)!$ when n is even and $\Gamma(n/2+1) = \sqrt{\pi}(1 \cdot 3 \cdot 5 \cdot \cdots \cdot n)/2^{(n+1)/2}$ when n is odd.

Volume of Balls



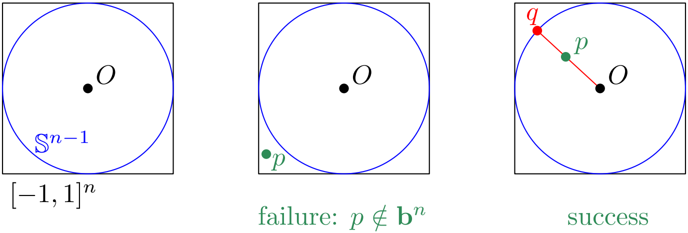
- The volume of the unit ball increases when the dimension goes from 0 to 5, and then it decreases.

Random Point on a Sphere

Problem

We want to generate a point on $\mathbb{S}^{n-1}$ uniformly at random.

- Algorithm:



Random Point on a Sphere

Pseudocode

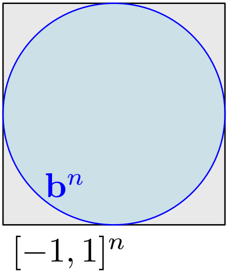
```
1: procedure RandomPointOnSphere(n)
2:     repeat
3:         for i ← 1, n do
4:             p_i ← random(−1, 1)          ▷ random number in [−1, 1]
5:     until 0 < ‖p‖ ⩽ 1
6:     return p/‖p‖
```

- Analysis: The condition at line 5 is satisfied with probability $\mathrm{Vol}(\mathbf{b}^n)/2^n$.
- So the loop is iterated on average $2^n/\mathrm{Vol}(\mathbf{b}^n)$ times.
- Each iteration takes $O(n)$ time.
- So this algorithm runs in $O(n2^n/\mathrm{Vol}(\mathbf{b}^n))$ time.

Random Point on a Sphere

- Is this a good algorithm?
- For small values of n, it is fast.
- Example: $n = 3$, $\quad \mathrm{Vol}(\mathbf{b}^3) \approx 4.19 \quad 3 \cdot 2^3/\mathrm{Vol}(\mathbf{b}^3) \approx 5.72$
- In higher dimension, however, it becomes extremely slow.
- Example: $n = 50$
  $\mathrm{Vol}(\mathbf{b}^{50}) \approx 1.73 \cdot 10^{-13} \quad 50 \cdot 2^{50}/\mathrm{Vol}(\mathbf{b}^{50}) \approx 3.25 \cdot 10^{29}$
- So in practice, this algorithm does not terminate in dimension 50.
- Conclusion: This algorithm is good in low dimension.
- But we need a better one for high dimension.

Random Point on a Sphere



- Geometric observation: In high dimension, the volume of the unit ball is much smaller than the volume of the hypercube $[-1, 1]^n$.
- This is not true in low dimension.
- So the figure above does *not* give the right intuition for high dimensional spaces

Random Point on a Sphere

- We now present a better approach.
- The *normal distribution* has density function

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

- We generate each p_i at random from this distribution.
- Then the distribution of $p = (p_1, \dots, p_n)$ is the multi-dimensional normal distribution

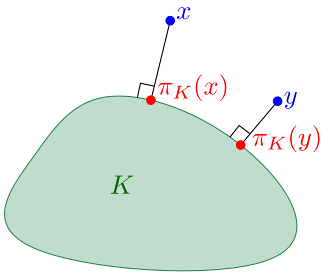
$$\frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{x_1^2}{2} - \dots - \frac{x_n^2}{2}\right).$$

- This distribution is spherically symmetric, so $p/\|p\|$ is a point chosen uniformly at random in $\mathbb{S}^{n-1}$.
- This approach takes $O(n)$ time.

Lipschitz Functions

Definition

Let $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^k$, and let $c > 0$. A function $f : A \to B$ is c-*Lipschitz* if $\|f(x) - f(y)\| \leqslant c\|x - y\|$ for every $x, y \in A$.



Definition

Let K be a closed convex subset of $\mathbb{R}^n$, and let $x \in \mathbb{R}^n$. The *projection* $\pi_K(x)$ of x onto K is the point $p \in K$ such that $\|xp\|$ is minimum.

Lipschitz Functions

Theorem

The projection onto a convex set is 1-Lipschitz.

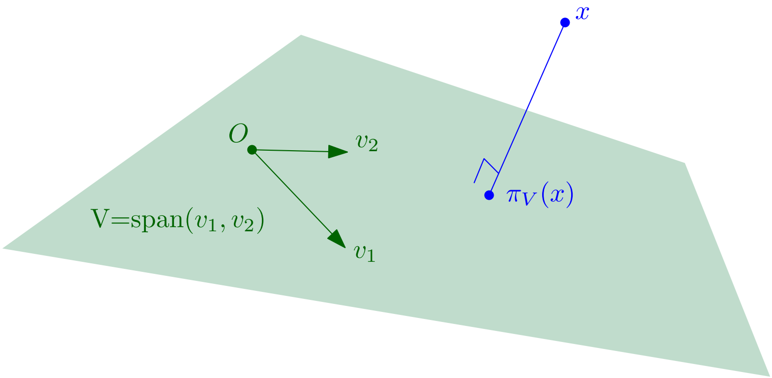
- In other words, $\|\pi_K(x) - \pi_K(y)\| \leqslant \|x - y\|$ for all $x, y \in \mathbb{R}^n$.

- Special case: projection onto a linear subspace.
- Given a set V of m vectors $v_1, \dots, v_m \in \mathbb{R}^n$, the *linear subspace spanned* by V is the set

$$\mathrm{span}(V) = \{\lambda_1 v_1 + \cdots + \lambda_k v_m \mid \lambda_1, \dots, \lambda_k \in \mathbb{R}^n\}.$$

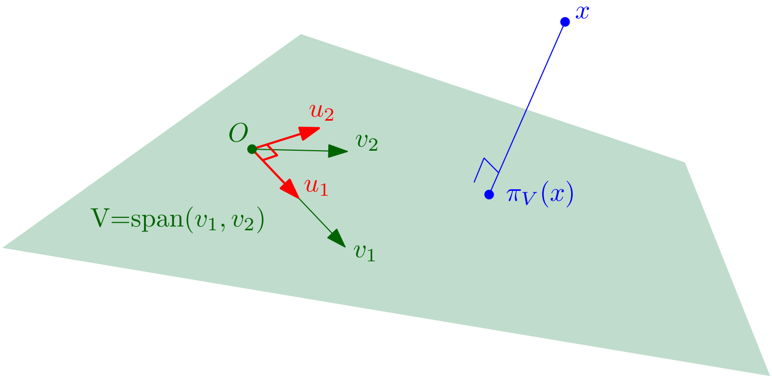
  (Remark: the terms "point" and "vector" mean the same: they are just elements of $\mathbb{R}^n$.)
- A linear subspace is convex, so we can apply the theorem above.

Projections



- So π_V is 1-Lipschitz when V is a linear subspace.
- How to compute $\pi_V(x)$?

Projections



- We use an *orthonormal basis*, that is, a set of vectors $u_1, \dots, u_k$ such that $\mathrm{span}(V) = \mathrm{span}(u_1, \dots, u_k)$, $\|u_i\| = 1$ for all i, and $\langle u_i, u_j \rangle = 0$ for all $i \neq j$.

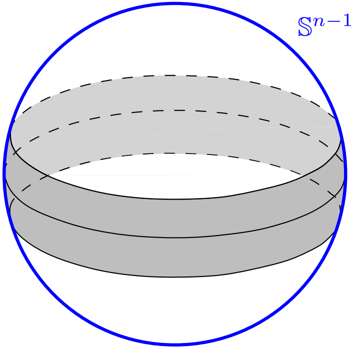
Projections

- Then the dimension of V is k, and for all $x \in \mathbb{R}^n$,

$$\pi_V(x) = \sum_{i=1}^{k} \langle u_i, x \rangle u_i.$$

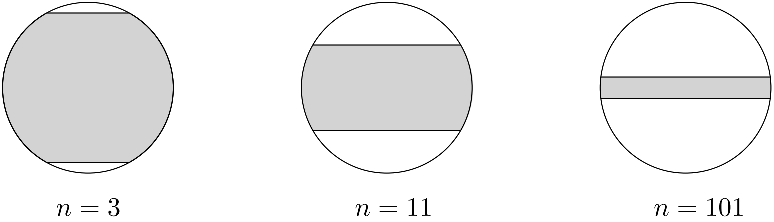
- How to find an orthonormal basis of $V = \mathrm{span}(v_1, \dots, v_m)$?
- We apply the Gram-Schmidt orthonormalization process. (See your linear algebra course.)
- It takes $O(nk^2)$ time in non-degenerate cases, i.e. $m = k$.

Measure Concentration on the Sphere



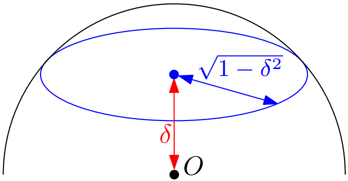
- When the dimension n is high, most of the surface measure on $\mathbb{S}^{n-1}$ is concentrated along a small band near the equator.

Measure Concentration on the Sphere



- Band around the equator of $\mathbb{S}^{n-1}$ that contains 90% of the measure.

Measure Concentration on the Sphere: Intuition



- The radius of the horizontal section at height δ is $\sqrt{1-\delta^2} \approx 1 - \dfrac{\delta^2}{2}$.
- So its area is roughly proportional to

$$\left(1 - \frac{\delta^2}{2}\right)^{n-2} \leqslant \exp\left(-(n-2)\frac{\delta^2}{2}\right)$$

- When $\sqrt{n} \gg \delta$, this is very small.

Measure Concentration on the Sphere

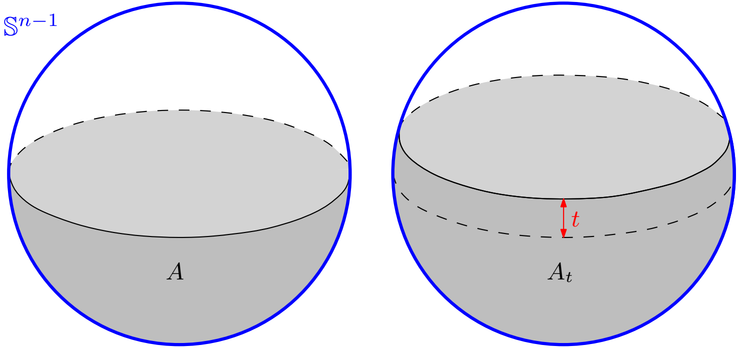
Theorem (Measure concentration on the sphere)

Let $A \subseteq \mathbb{S}^{n-1}$ be a measurable set with $\Pr[A] \geqslant 1/2$ and let A_t denote the set of points of $\mathbb{S}^{n-1}$ within distance at most t from A, where $t \leqslant 2$. Then $\Pr[A_t] \geqslant 1 - 2\exp(-nt^2/2)$.

- Here we assume that $\Pr[\mathbb{S}^{n-1}] = 1$, hence A has at least half the measure of $\mathbb{S}^{n-1}$.
- We have

$$A_t = \{x \in \mathbb{S}^{n-1} \mid \exists y \in A \text{ such that } \|x - y\| \leqslant t\}.$$

Measure Concentration on the Sphere: Example



- On the left, the bottom half-sphere with measure $\Pr[A] = 1/2$.
- On the right, the enlarged version A_t has measure almost 1. More precisely $\Pr[A_t] \geqslant 1 - 2\exp(-nt^2/2)$.

Measure Concentration of Lipschitz Functions

Definition

Let $f : \mathbb{S}^{n-1} \to \mathbb{R}$. The *median* of f is

$$\mathrm{med}(f) = \sup\{t \in \mathbb{R} \mid \Pr[f \leqslant t] \leqslant 1/2.\}$$

- Intuitively, it is the value of t that splits the sphere into two halves of measure $1/2$: the points corresponding to values at most t, and the points corresponding to values greater than t.

Measure Concentration of Lipschitz Functions

Lemma (Lévy)

Let $f : \mathbb{S}^{n-1}$ be 1-Lipschitz. Then for all $t \in [0,1]$, we have

$$\Pr[f > \mathrm{med}(f) + t] \leqslant 2\exp\left(-\frac{t^2 n}{2}\right)$$

and

$$\Pr[f < \mathrm{med}(f) - t] \leqslant 2\exp\left(-\frac{t^2 n}{2}\right).$$

- We now prove this lemma.
- Let $A = \{x \in \mathbb{S}^{n-1} \mid f(x) \leqslant \mathrm{med}(f)\}$. Then $\Pr[A] \geqslant 1/2$.
- Let A_t be the set defined above, and let $x \in A_t$.
- Let $\mathrm{nn}(x)$ be the nearest point to x in A, and thus $\|x - \mathrm{nn}(x)\| \leqslant t$.

Measure Concentration of Lipschitz Functions

- As f is 1-Lipschitz, it follows that

$$f(x) \leqslant f(\mathrm{nn}(x)) + \|x - \mathrm{nn}(x)\| \leqslant \mathrm{med}(f) + t$$

  and thus

$$\Pr[A_t] \leqslant \Pr[f \leqslant \mathrm{med}(f) + t].$$

- So by the theorem of measure concentration on the sphere

$$\Pr[f > \mathrm{med}(f) + t] \leqslant 1 - \Pr[A_t] \leqslant 2\exp(-t^2 n/2).$$

The Johnson-Lindenstrauss Lemma

Lemma

For any $x \in \mathbb{S}^{n-1}$, *let* $f(x) = \sqrt{x_1^2 + \cdots + x_k^2}$ *be the length of the vector formed by its first* k *coordinates. Then there exists* $m = m(n,k)$ *such that*

$$\Pr[f(x) \geqslant m + t] \leqslant 2\exp\left(-\frac{t^2 n}{2}\right)$$

and

$$\Pr[f(x) \leqslant m - t] \leqslant 2\exp\left(-\frac{t^2 n}{2}\right)$$

for any $t \in [0,1]$. *Furthermore, we have* $m \geqslant \frac{1}{2}\sqrt{\frac{k}{n}}$ *whenever* $k \geqslant 10 \ln n$.

The Johnson-Lindenstrauss Lemma

- This shows that if we project a unit vector chosen uniformly at random onto the linear space spanned by the first k coordinate vectors, the length of the projected vector is sharply concentrated.

- The proof of the first part follows directly from Lévy's lemma.
- The bound on k needs an additional argument. We do not present it here.

- The lemma above applies to a *random* vector x and the *fixed* subspace of dimension k spanned by the first k vectors of the basis.
- Alternatively, it applies to a *fixed* vector x, and a subspace of dimension k *chosen uniformly at random*.
- How to generate this random subspace? Pick k unit vectors at random, and construct a base of their spanning subspace by Gram-Schmidt orthonormalization.

The Johnson-Lindenstrauss Lemma

Definition

Let $X \subseteq \mathbb{R}^n$ and $\varepsilon > 0$. A function $f : \mathbb{R}^n \to \mathbb{R}^k$ is a $(1+\varepsilon)$-embedding for X if

$$\frac{1}{1+\varepsilon}\|p-q\| \leqslant \|f(p)-f(q)\| \leqslant (1+\varepsilon)\|p-q\|$$

for all $p, q \in X$.

- In other words, distances are distorted by a factor at most $1+\varepsilon$ after applying this function.
- If we can find such a mapping, then we can work in lower dimension k instead of n, and obtain a good approximation.
- So we would have performed *dimension reduction*.

The Johnson-Lindenstrauss Lemma

Theorem (Johnson-Lindenstrauss)

Let X be a set of n points in a Euclidean space, and let $0 < \varepsilon \leqslant 1$. Then there exists a $(1+\varepsilon)$-embedding of X into $\mathbb{R}^k$, where $k = O(\log(n)/\varepsilon^2)$.

- Here we just sketch the proof.
- We may have $X \subseteq \mathbb{R}^m$ with $m > n$. In this case, just work within $\mathrm{span}(X)$, which has dimension at most n.
- We consider $X \subseteq \mathbb{R}^n$, and we generate a random subspace V of dimension k.
- We project X onto V_k.
- Let x, y be two points in X.
- By the lemma on Slide 25, with high probability, the projected vector $\pi_V\left(\frac{x-y}{\|x-y\|}\right)$ has length close to the constant $m = m(n,k)$.

The Johnson-Lindenstrauss Lemma

- So we have $\|\pi_V(x-y)\| \approx m\|x-y\|$.
- Since π_V is linear, it means that

$$\left\|\pi_V\left(\frac{x}{m}\right) - \pi_V\left(\frac{y}{m}\right)\right\| \approx \|x-y\|.$$

- By choosing $k = O(\log(n)/\varepsilon^2)$ appropriately, we can ensure that $x \to \pi_V(x/m)$ is a $(1+\varepsilon)$-embedding with non-zero probability.
- So there exists a suitable subspace V.

- In fact, if we increase k by a constant factor, it happens with high probability.

The Johnson-Lindenstrauss Lemma

- The Johnson-Lindenstrauss lemma is a theoretical tool for designing geometric approximation algorithms.

- The idea behind its proof is also useful in practice:
- Given high-dimensional data, project it onto a random subspace of lower dimension, and work with the projected points.
- It may give a good approximation of the desired result.

- We also saw an analogous idea in the lecture on approximating the diameter of a point set:
- We projected the point set onto several lines, which are a 1-dimensional subspaces.
- The Johnson-Lindenstrauss lemma projects onto a subspace of dimension $k = O(\log(n)/\varepsilon^2)$ instead of 1.

The Johnson-Lindenstrauss Lemma

- It can also be shown that the embedding we constructed also approximately preserves angles, not just distances.