• Problem Set 2 is due this upcoming Sunday 3/8 at 8pm.
• Midterm is next Thursday, 3/12. See webpage for study guide/practice questions.
• I will hold office hours after class today.
• Next week office hours will be at the usual time after class Tuesday and also before class at 10:00am.
Last Class: Finished Up Johnson-Lindenstrauss Lemma

- Completed the proof of the Distributional JL lemma.
- Showed two applications of random projection: faster support vector machines and $k$-means clustering.
- Started discussion of high-dimensional geometry.

This Class: High-Dimensional Geometry

- Bizarre phenomena in high-dimensional space.
- Connections to JL lemma and random projection.
What is the largest set of mutually orthogonal unit vectors in $d$-dimensional space? Answer: $d$.

What is the largest set of unit vectors in $d$-dimensional space that have all pairwise dot products $|\langle \vec{x}, \vec{y} \rangle| \leq \epsilon$? (think $\epsilon = .01$) Answer: $2^{\Theta(\epsilon^2 d)}$.

In fact, an exponentially large set of random vectors will be nearly pairwise orthogonal with high probability!
Claim: $2^{\Theta(e^2d)}$ random $d$-dimensional unit vectors will have all pairwise dot products $|\langle \vec{x}, \vec{y} \rangle| \leq \epsilon$ (be nearly orthogonal).

Proof: Let $\vec{x}_1, \ldots, \vec{x}_t$ each have independent random entries set to $\pm 1/\sqrt{d}$.

- What is $||\vec{x}_i||_2$? Every $\vec{x}_i$ is always a unit vector.
- What is $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle]$? $\mathbb{E}[\langle \vec{x}_i, \vec{x}_j \rangle] = 0$
- By a Chernoff bound, $\Pr[|\langle \vec{x}_i, \vec{x}_j \rangle| \geq \epsilon] \leq 2e^{-\epsilon^2d/6}$.
- If we chose $t = \frac{1}{2}e^{\epsilon^2d/12}$, using a union bound over all $\binom{t}{2} \leq \frac{1}{8}e^{\epsilon^2d/6}$ possible pairs, with probability $\geq 3/4$ all will be nearly orthogonal.
Up Shot: In $d$-dimensional space, a set of $2^{\Theta(\epsilon^2 d)}$ random unit vectors have all pairwise dot products at most $\epsilon$ (think $\epsilon = .01$)

$$\|\vec{x}_i - \vec{x}_j\|_2^2 = \|\vec{x}_i\|_2^2 + \|\vec{x}_j\|_2^2 - 2\vec{x}_i^T\vec{x}_j \geq 1.98.$$ 

Even with an exponential number of random vector samples, we don’t see any nearby vectors.

• Can make methods like nearest neighbor classification or clustering useless.

Curse of dimensionality for sampling/learning functions in high-dimensional space – samples are very ‘sparse’ unless we have a huge amount of data.

• Only hope is if we lots of structure (which we typically do...)
CURSE OF DIMENSIONALITY

Distances for MNIST Digits:

Distances for Random Images:

Another Interpretation: Tells us that random data can be a very bad model for actual input data.
Recall: The Johnson Lindenstrauss lemma states that if \( \mathbf{\Pi} \in \mathbb{R}^{m \times d} \) is a random matrix (linear map) with \( m = O\left(\frac{\log n}{\epsilon^2}\right) \), for \( \vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^d \) with high probability, for all \( i, j \):

\[
(1 - \epsilon)\|\vec{x}_i - \vec{x}_j\|_2^2 \leq \|\mathbf{\Pi}\vec{x}_i - \mathbf{\Pi}\vec{x}_j\|_2^2 \leq (1 + \epsilon)\|\vec{x}_i - \vec{x}_j\|_2^2.
\]

Implies: If \( \vec{x}_1, \ldots, \vec{x}_n \) are nearly orthogonal unit vectors in \( d \)-dimensions (with pairwise dot products bounded by \( \epsilon/8 \)), then \( \frac{\mathbf{\Pi}\vec{x}_1}{\|\mathbf{\Pi}\vec{x}_1\|_2}, \ldots, \frac{\mathbf{\Pi}\vec{x}_n}{\|\mathbf{\Pi}\vec{x}_n\|_2} \) are nearly orthogonal unit vectors in \( m \)-dimensions (with pairwise dot products bounded by \( \epsilon \)).

- Similar to SVM analysis. Algebra is a bit messy but a good exercise to partially work through.
Claim 1: $n$ nearly orthogonal unit vectors can be projected to $m = O \left( \frac{\log n}{\epsilon^2} \right)$ dimensions and still be nearly orthogonal.

Claim 2: In $m$ dimensions, there are at most $2^{O(\epsilon^2 m)}$ nearly orthogonal vectors.

- For both these to hold it might be that $n \leq 2^{O(\epsilon^2 m)}$.
- $2^{O(\epsilon^2 m)} = 2^{O(\log n)} \geq n$. Tells us that the JL lemma is optimal up to constants.
- $m$ is chosen just large enough so that the odd geometry of $d$-dimensional space still holds on the $n$ points in question after projection to a much lower dimensional space.
Let $\mathcal{B}_d$ be the unit ball in $d$ dimensions. $\mathcal{B}_d = \{ x \in \mathbb{R}^d : \|x\|_2 \leq 1 \}$.

What percentage of the volume of $\mathcal{B}_d$ falls within $\epsilon$ distance of its surface? Answer: all but a $(1 - \epsilon)^d \leq e^{-\epsilon d}$ fraction. Exponentially small in the dimension $d$!

Volume of a radius $R$ ball is $\frac{\pi^\frac{d}{2}}{(d/2)!} \cdot R^d$. 
All but an $e^{-\epsilon d}$ fraction of a unit ball’s volume is within $\epsilon$ of its surface. If we randomly sample points with $\|x\|_2 \leq 1$, nearly all will have $\|x\|_2 \geq 1 - \epsilon$.

- **Isoperimetric inequality:** the ball has the maximum surface area/volume ratio of any shape.

- If we randomly sample points from any high-dimensional shape, nearly all will fall near its surface.

- ‘All points are outliers.’
What fraction of the cubes are visible on the surface of the cube?

\[
\frac{10^3 - 8^3}{10^3} = \frac{1000 - 512}{1000} = .488.
\]
What percentage of the volume of $\mathcal{B}_d$ falls within $\epsilon$ distance of its equator? Answer: all but a $2^{\Theta(-\epsilon^2d)}$ fraction.

Formally: volume of set $S = \{x \in \mathcal{B}_d : |x(1)| \leq \epsilon\}$.

By symmetry, all but a $2^{\Theta(-\epsilon^2d)}$ fraction of the volume falls within $\epsilon$ of any equator! $S = \{x \in \mathcal{B}_d : |\langle x, t \rangle| \leq \epsilon\}$
Claim 1: All but a $2^{\Theta(-\epsilon^2 d)}$ fraction of the volume of a ball falls within $\epsilon$ of any equator.

Claim 2: All but a $2^{\Theta(-\epsilon d)}$ fraction falls within $\epsilon$ of its surface.

How is this possible? High-dimensional space looks nothing like this picture.
Claim: All but a \(2^{\Theta(-\epsilon^2 d)}\) fraction of the volume of a ball falls within \(\epsilon\) of its equator. I.e., in \(S = \{x \in B_d : |x(1)| \leq \epsilon\}\).

Proof Sketch:

• Let \(x\) have independent Gaussian \(\mathcal{N}(0, 1)\) entries and let \(\tilde{x} = \frac{x}{\|x\|_2}\). \(\tilde{x}\) is selected uniformly at random from the surface of the ball.

• Suffices to show that \(\Pr[|\tilde{x}(1)| > \epsilon] \leq 2^{\Theta(-\epsilon^2 d)}\). Why?

• \(\tilde{x}(1) = \frac{x(1)}{\|x\|_2}\). What is \(\mathbb{E}[\|x\|_2^2]?\mathbb{E}[\|x\|_2^2] = \sum_{i=1}^{d} \mathbb{E}[x(i)^2] = d.\)

\(\Pr[\|x\|_2^2 \leq d/2] \leq 2^{-\Theta(d)}\)

• Conditioning on \(\|x\|_2^2 \geq d/2\), since \(x(1)\) is normally distributed,

\[
\Pr[|\tilde{x}(1)| > \epsilon] = \Pr[|x(1)| > \epsilon \cdot \|x\|_2] \\
\leq \Pr[|x(1)| > \epsilon \cdot \sqrt{d/2}] = 2^{\Theta(-(\epsilon \sqrt{d/2})^2)} = 2^{\Theta(-\epsilon^2 d)}.
\]
Let $C_d$ be the $d$-dimensional cube: $C_d = \{x \in \mathbb{R}^d : |x(i)| \leq 1 \forall i\}$.

In low-dimensions, the cube is not that different from the ball.

But volume of $C_d$ is $2^d$ while volume of $B_d$ is $\frac{\pi^{d/2}}{(d/2)!} = \frac{1}{d^{\Theta(d)}}$. A huge gap! So something is very different about these shapes...
Corners of cube are $\sqrt{d}$ times further away from the origin than the surface of the ball.
Data generated from the ball $B_d$ will behave very differently than data generated from the cube $C_d$.

- $x \sim B_d$ has $\|x\|_2^2 \leq 1$.
- $x \sim C_d$ has $\mathbb{E}[\|x\|_2^2] = ?d/3$, and $\Pr[\|x\|_2^2 \leq d/6] \leq 2^{-\Theta(d)}$.
- Almost all the volume of the unit cube falls in its corners, and these corners lie far outside the unit ball.
• High-dimensional space behaves very differently from low-dimensional space.
• Random projection (i.e., the JL Lemma) reduces to a much lower-dimensional space that is still large enough to capture this behavior on a subset of $n$ points.
• Need to be careful when using low-dimensional intuition for high-dimensional vectors.
• Need to be careful when modeling data as random vectors in high-dimensions.