CS-GY 6763: Lecture 4
High Dimensional Geometry

NYU, Prof. Ainesh Bakshi

Unifying Theme of the Course

How do we deal with data (vectors) in high-dimensions?

- High-dimensional similarity search.

Unifying Theme of the Course

How do we deal with data (vectors) in high-dimensions?

- High-dimensional similarity search.
- Iterative methods for optimizing functions in high-dimensions.

Unifying Theme of the Course

How do we deal with data (vectors) in high-dimensions?

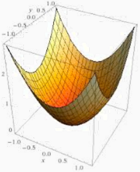
- High-dimensional similarity search.
- Iterative methods for optimizing functions in high-dimensions.
- SVD + low-rank approximation to find and visualize low-dimensional structure.

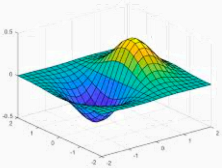
Unifying Theme of the Course

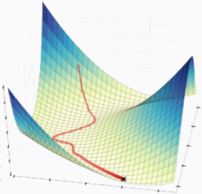
How do we deal with data (vectors) in high-dimensions?

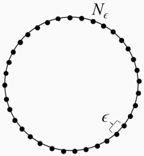
- High-dimensional similarity search.
- Iterative methods for optimizing functions in high-dimensions.
- SVD + low-rank approximation to find and visualize low-dimensional structure.
- Convert large graphs to high-dimensional vector data to uncover interesting things.

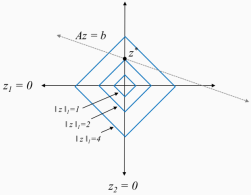
Often visualize data and algorithms in 1,2, or 3 dimensions.

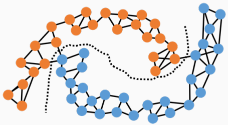






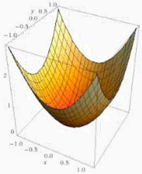


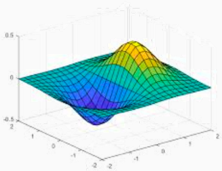


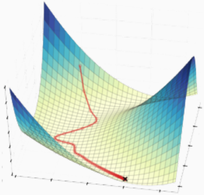


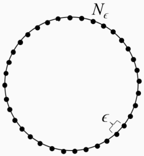
High Dimensional is Not Like Low Dimensional

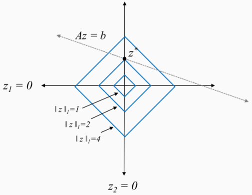
Often visualize data and algorithms in 1,2, or 3 dimensions.

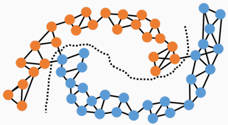












This lecture: Prove that high-dimensional space looks **very different** from low-dimensional space. These images are rarely very informative!

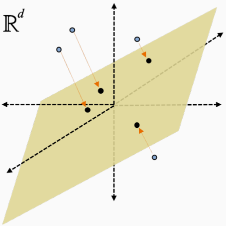
Next lecture: Ignore our own advice.

Sketching and Dimensionality Reduction

Next lecture: Ignore our own advice.

Learn about **sketching, aka dimensionality reduction** techniques that seek to approximate high-dimensional vectors with much lower dimensional vectors.

- Johnson-Lindenstrauss lemma for ℓ_2 space.
- MinHash for binary vectors (next class) .

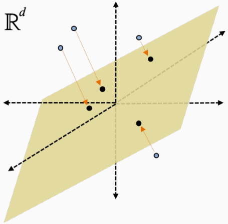


Sketching and Dimensionality Reduction

Next lecture: Ignore our own advice.

Learn about **sketching, aka dimensionality reduction** techniques that seek to approximate high-dimensional vectors with much lower dimensional vectors.

- Johnson-Lindenstrauss lemma for ℓ_2 space.
- MinHash for binary vectors (next class) .



This lecture should help you understand the potential and limitations of these methods.

Orthogonal Vectors

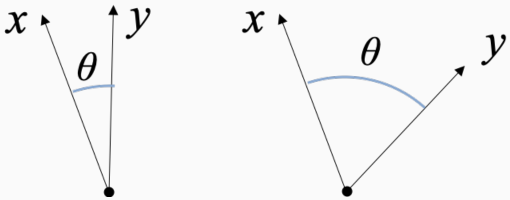
Recall the inner product between two d dimensional vectors:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = \sum_{j=1}^{d} x[j] y[j]$$

Orthogonal Vectors

Recall the inner product between two d dimensional vectors:

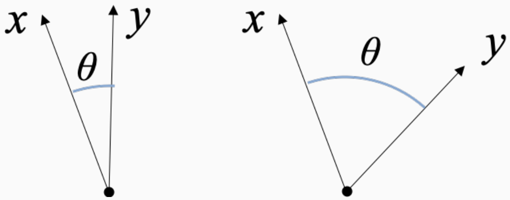
$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = \sum_{j=1}^{d} x[j] y[j]$$



Orthogonal Vectors

Recall the inner product between two d dimensional vectors:

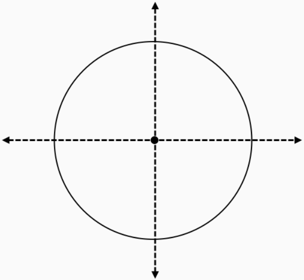
$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = \sum_{j=1}^{d} x[j] y[j]$$

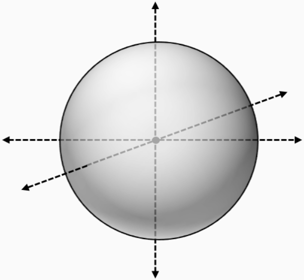


$$\langle \mathbf{x}, \mathbf{y} \rangle = \cos(\theta) \cdot \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2$$

Orthogonal Vectors

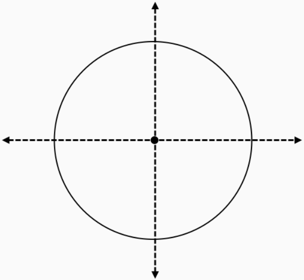
What is the largest set of **mutually orthogonal** unit vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ in d-dimensional space? I.e. with inner product $|\mathbf{x}_i^T \mathbf{x}_j| = 0$ for all i, j.

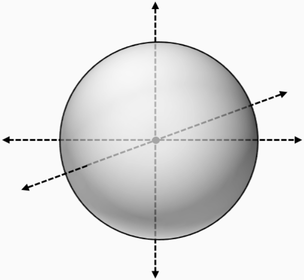




Orthogonal Vectors

What is the largest set **nearly orthogonal** unit vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ in 2 or 3 dimensions? I.e., with inner product $|\mathbf{x}_i^T \mathbf{x}_j| \leq \epsilon$ for all i, j. Consider the case when ϵ is a constant. E.g. $\epsilon = 1/10$.





Orthogonal Vectors

What is the largest set **nearly orthogonal** unit vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ in d dimensions? I.e., with inner product $|\mathbf{x}_i^T \mathbf{x}_j| \leq \epsilon$ for all i, j. Consider the case when ϵ is a constant. E.g. $\epsilon = 1/10$.

1. d
2. $\Theta(d)$
3. $\Theta(d^2)$
4. $2^{\Theta(d)}$

Orthogonal Vectors

Formal Claim: In d-dimensional space, there are $2^{\Theta(\epsilon^2 d)}$ unit vectors with all pairwise inner products $\leq \epsilon$, where $\epsilon \gg 1/\sqrt{d}$.

Orthogonal Vectors

Formal Claim: In d-dimensional space, there are $2^{\Theta(\epsilon^2 d)}$ unit vectors with all pairwise inner products $\leq \epsilon$, where $\epsilon \gg 1/\sqrt{d}$.

Proof strategy: Use the Probabilistic Method! For $t = 2^{\Theta(\epsilon^2 d)}$, define a random process which generates random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ that are unlikely to have large inner product.

Orthogonal Vectors

Formal Claim: In d-dimensional space, there are $2^{\Theta(\epsilon^2 d)}$ unit vectors with all pairwise inner products $\leq \epsilon$, where $\epsilon \gg 1/\sqrt{d}$.

Proof strategy: Use the Probabilistic Method! For $t = 2^{\Theta(\epsilon^2 d)}$, define a random process which generates random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ that are unlikely to have large inner product.

1. Claim that, with non-zero probability, $|\mathbf{x}_i^T \mathbf{x}_j| \leq \epsilon$ for all i, j.
2. Conclude that there must exists some set of t unit vectors with all pairwise inner-products bounded by ϵ.

Probabilistic Method

Claim: There is an exponential number (i.e., $2^{\Theta(d)}$) of nearly orthogonal unit vectors in d dimensional space.

Proof: Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ all have independent random entries, each set to $\pm\frac{1}{\sqrt{d}}$ with equal probability. Ex.

$$x_1 = \left(\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}, -\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}, \ldots, \frac{1}{\sqrt{d}}\right)$$

Probabilistic Method

Claim: There is an exponential number (i.e., $2^{\Theta(d)}$) of nearly orthogonal unit vectors in d dimensional space.

Proof: Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ have independent random entries, each $\pm\frac{1}{\sqrt{d}}$ w.p. $1/2$.

- $\|\mathbf{x}_i\|_2^2 =$

Probabilistic Method

Claim: There is an exponential number (i.e., $2^{\Theta(d)}$) of nearly orthogonal unit vectors in d dimensional space.

Proof: Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ have independent random entries, each $\pm\frac{1}{\sqrt{d}}$ w.p. $1/2$.

- $\|\mathbf{x}_i\|_2^2 = \sum_{k=1}^{d} \frac{1}{d} = 1$

Probabilistic Method

Claim: There is an exponential number (i.e., $2^{\Theta(d)}$) of nearly orthogonal unit vectors in d dimensional space.

Proof: Let $\mathbf{x}_1, \dots, \mathbf{x}_t$ have independent random entries, each $\pm\frac{1}{\sqrt{d}}$ w.p. $1/2$.

- $\|\mathbf{x}_i\|_2^2 = \sum_{k=1}^{d} \frac{1}{d} = 1$
- $\mathbb{E}[\mathbf{x}_i^T \mathbf{x}_j] =$

Probabilistic Method

Claim: There is an exponential number (i.e., $2^{\Theta(d)}$) of nearly orthogonal unit vectors in d dimensional space.

Proof: Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ have independent random entries, each $\pm\frac{1}{\sqrt{d}}$ w.p. $1/2$.

- $\|\mathbf{x}_i\|_2^2 = \sum_{k=1}^{d} \frac{1}{d} = 1$
- $\mathbb{E}[\mathbf{x}_i^T \mathbf{x}_j] = \sum_{k=1}^{d} \mathbb{E}[\mathbf{x}_i[k]\mathbf{x}_j[k]] = \sum_{k=1}^{d} \mathbb{E}[\mathbf{x}_i[k]]\mathbb{E}[\mathbf{x}_j[k]] = 0$

Probabilistic Method

Claim: There is an exponential number (i.e., $2^{\Theta(d)}$) of nearly orthogonal unit vectors in d dimensional space.

Proof: Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ have independent random entries, each $\pm\frac{1}{\sqrt{d}}$ w.p. $1/2$.

- $\|\mathbf{x}_i\|_2^2 = \sum_{k=1}^{d} \frac{1}{d} = 1$
- $\mathbb{E}[\mathbf{x}_i^T \mathbf{x}_j] = \sum_{k=1}^{d} \mathbb{E}[\mathbf{x}_i[k]\mathbf{x}_j[k]] = \sum_{k=1}^{d} \mathbb{E}[\mathbf{x}_i[k]]\mathbb{E}[\mathbf{x}_j[k]] = 0$
- $\mathbf{x}_i[k]\mathbf{x}_j[k] = \begin{cases} +1/d & \text{w.p. } 1/2 \\ -1/d & \text{w.p. } 1/2 \end{cases}$

Probabilistic Method

Claim: There is an exponential number (i.e., $2^{\Theta(d)}$) of nearly orthogonal unit vectors in d dimensional space.

Proof: Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ have independent random entries, each $\pm\frac{1}{\sqrt{d}}$ w.p. $1/2$.

- $\|\mathbf{x}_i\|_2^2 = \sum_{k=1}^{d} \frac{1}{d} = 1$
- $\mathbb{E}[\mathbf{x}_i^T \mathbf{x}_j] = \sum_{k=1}^{d} \mathbb{E}[\mathbf{x}_i[k]\mathbf{x}_j[k]] = \sum_{k=1}^{d} \mathbb{E}[\mathbf{x}_i[k]]\mathbb{E}[\mathbf{x}_j[k]] = 0$
- $\mathbf{x}_i[k]\mathbf{x}_j[k] = \begin{cases} +1/d & \text{w.p. } 1/2 \\ -1/d & \text{w.p. } 1/2 \end{cases}$
- $\mathrm{Var}[\mathbf{x}_i^T \mathbf{x}_j] =$

Probabilistic Method

Claim: There is an exponential number (i.e., $2^{\Theta(d)}$) of nearly orthogonal unit vectors in d dimensional space.

Proof: Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ have independent random entries, each $\pm \frac{1}{\sqrt{d}}$ w.p. $1/2$.

- $\|\mathbf{x}_i\|_2^2 = \sum_{k=1}^{d} \frac{1}{d} = 1$
- $\mathbb{E}[\mathbf{x}_i^T \mathbf{x}_j] = \sum_{k=1}^{d} \mathbb{E}[\mathbf{x}_i[k]\mathbf{x}_j[k]] = \sum_{k=1}^{d} \mathbb{E}[\mathbf{x}_i[k]]\mathbb{E}[\mathbf{x}_j[k]] = 0$
- $\mathbf{x}_i[k]\mathbf{x}_j[k] = \begin{cases} +1/d & \text{w.p. } 1/2 \\ -1/d & \text{w.p. } 1/2 \end{cases}$
- $\mathrm{Var}[\mathbf{x}_i^T \mathbf{x}_j] = \sum_{k=1}^{d} \mathrm{Var}[\mathbf{x}_i[k]\mathbf{x}_j[k]] = \sum_{k=1}^{d} \frac{1}{d^2} = \frac{1}{d}$

Informal Proof

Let $Z = \mathbf{x}_i^T \mathbf{x}_j = \sum_{k=1}^{d} C_k$ where each C_k is random $\pm\frac{1}{d}$.

Informal Proof

Let $Z = \mathbf{x}_i^T \mathbf{x}_j = \sum_{k=1}^{d} C_k$ where each C_k is random $\pm\frac{1}{d}$.

Z is a sum of many i.i.d. random variables, so looks approximately Gaussian. Roughly, we expect that:

$$\Pr[|Z - \mathbb{E}Z| \geq \alpha \cdot \sigma] \leq O(e^{-\alpha^2})$$

Informal Proof

Let $Z = \mathbf{x}_i^T \mathbf{x}_j = \sum_{k=1}^{d} C_k$ where each C_k is random $\pm\frac{1}{d}$.

Z is a sum of many i.i.d. random variables, so looks approximately Gaussian. Roughly, we expect that:

$$\Pr[|Z - \mathbb{E}Z| \geq \alpha \cdot \sigma] \leq O(e^{-\alpha^2})$$

We want $\alpha \cdot \sigma = \epsilon$, where $\sigma = 1/\sqrt{d}$. Therefore, applying Chernoff with $\alpha = \epsilon\sqrt{d}$, we have:

$$\Pr[|Z| \geq \epsilon] \leq O(e^{-\epsilon^2 d}).$$

Informal Proof

Let $Z = \mathbf{x}_i^T \mathbf{x}_j = \sum_{k=1}^{d} C_k$ where each C_k is random $\pm \frac{1}{d}$.

Z is a sum of many i.i.d. random variables, so looks approximately Gaussian. Roughly, we expect that:

$$\Pr[|Z - \mathbb{E}Z| \geq \alpha \cdot \sigma] \leq O(e^{-\alpha^2})$$

We want $\alpha \cdot \sigma = \epsilon$, where $\sigma = 1/\sqrt{d}$. Therefore, applying Chernoff with $\alpha = \epsilon\sqrt{d}$, we have:

$$\Pr[|Z| \geq \epsilon] \leq O(e^{-\epsilon^2 d}).$$

By a union bound, for any pairs x_i, x_j from $x_1, x_2, x_3, \ldots, x_t$,

$$\Pr[|\mathbf{x}_i^T \mathbf{x}_j| \geq \epsilon] \leq \binom{t}{2} \cdot O(e^{-\epsilon^2 d}) \leq 0.1$$

Informal Proof

Let $Z = \mathbf{x}_i^T \mathbf{x}_j = \sum_{k=1}^{d} C_k$ where each C_k is random $\pm\frac{1}{d}$.

Z is a sum of many i.i.d. random variables, so looks approximately Gaussian. Roughly, we expect that:

$$\Pr[|Z - \mathbb{E}Z| \geq \alpha \cdot \sigma] \leq O(e^{-\alpha^2})$$

We want $\alpha \cdot \sigma = \epsilon$, where $\sigma = 1/\sqrt{d}$. Therefore, applying Chernoff with $\alpha = \epsilon\sqrt{d}$, we have:

$$\Pr[|Z| \geq \epsilon] \leq O(e^{-\epsilon^2 d}).$$

By a union bound, for any pairs x_i, x_j from $x_1, x_2, x_3, \ldots, x_t$,

$$\Pr[|\mathbf{x}_i^T \mathbf{x}_j| \geq \epsilon] \leq \binom{t}{2} \cdot O(e^{-\epsilon^2 d}) \leq 0.1$$

Suffices to set $t = e^{O(\epsilon^2 d)}$.

Formal Proof

Use an exponential concentration inequality!

Theorem (Chernoff Bound)

Let $X_1, X_2, \ldots, X_d$ be independent $\{0,1\}$-valued random variables and let $S = \sum_{i=1}^{d} X_i$. We have for any $\epsilon < 1$:

$$\Pr[|S - \mathbb{E}[S]| \geq \epsilon \mathbb{E}[S]] \leq 2e^{\frac{-\epsilon^2 \mathbb{E}[S]}{3}}.$$

Does not immediately apply because we have random variables that are $\pm 1/d$, not $0, 1$.

Formal Proof

Use an exponential concentration inequality!

Theorem (Chernoff Bound)

Let $X_1, X_2, \ldots, X_d$ be independent $\{0,1\}$-valued random variables and let $S = \sum_{i=1}^{d} X_i$. We have for any $\epsilon < 1$:

$$\Pr[|S - \mathbb{E}[S]| \geq \epsilon \mathbb{E}[S]] \leq 2e^{\frac{-\epsilon^2 \mathbb{E}[S]}{3}}.$$

Does not immediately apply because we have random variables that are $\pm 1/d$, not $0, 1$.

Formal Proof

Use an exponential concentration inequality!

Theorem (Chernoff Bound)

Let $X_1, X_2, \ldots, X_d$ be independent $\{0,1\}$-valued random variables and let $S = \sum_{i=1}^{d} X_i$. We have for any $\epsilon < 1$:

$$\Pr[|S - \mathbb{E}[S]| \geq \epsilon \mathbb{E}[S]] \leq 2e^{\frac{-\epsilon^2 \mathbb{E}[S]}{3}}.$$

Does not immediately apply because we have random variables that are $\pm 1/d$, not $0, 1$.

Common trick: shift and scale to transform to the binary case.

Formal Proof

$$\mathbf{x}_i^T \mathbf{x}_j = Z = \sum_{k=1}^{d} C_k = \frac{2}{d} \sum_{k=1}^{d} \frac{d}{2} \cdot C_k \qquad \text{Let } \left(\frac{d}{2} C_k = B_k - \frac{1}{2} \right)$$

$$= \frac{2}{d} \cdot \left(\sum_{k=1}^{d} B_k - 1/2 \right)$$

$$= \frac{2}{d} \cdot \left(-\frac{d}{2} + \sum_{k=1}^{d} B_k \right)$$

Formal Proof

$$\mathbf{x}_i^T \mathbf{x}_j = Z = \sum_{k=1}^{d} C_k = \frac{2}{d} \sum_{k=1}^{d} \frac{d}{2} \cdot C_k \qquad \text{Let } \left(\frac{d}{2} C_k = B_k - \frac{1}{2} \right)$$

$$= \frac{2}{d} \cdot \left(\sum_{k=1}^{d} B_k - 1/2 \right)$$

$$= \frac{2}{d} \cdot \left(-\frac{d}{2} + \sum_{k=1}^{d} B_k \right)$$

Observe B_k is uniform in $\{0, 1\}$.

Formal Proof

$$\mathbf{x}_i^T \mathbf{x}_j = Z = \frac{2}{d} \cdot \left(-\frac{d}{2} + \sum_{k=1}^{d} B_k \right), \quad B_k \text{ uniform in } \{0, 1\}.$$

$$\Pr[|Z| > \epsilon] = \Pr\left[\left| \sum_{k=1}^{d} B_k - \frac{d}{2} \right| > \frac{\epsilon d}{2} \right]$$

$$= \Pr\left[\left| \sum_{k=1}^{d} B_k - \mathbb{E}\left[\sum_{k=1}^{d} B_k \right] \right| > \epsilon \cdot \mathbb{E}\left[\sum_{k=1}^{d} B_k \right] \right]$$

Chernoff Bound

Theorem (Chernoff Bound)

Let $X_1, X_2, \ldots, X_d$ be independent $\{0,1\}$-valued random variables and let $S = \sum_{i=1}^{d} X_i$. We have for any $\epsilon < 1$:

$$\Pr[|S - \mathbb{E}[S]| \geq \epsilon \mathbb{E}[S]] \leq 2e^{\frac{-\epsilon^2 \mathbb{E}[S]}{3}}.$$

Apply with $X_1, \ldots, X_d = B_1, \ldots, B_d$:

$$\Pr[|S - \mathbb{E}[S]| \geq \epsilon \mathbb{E}[S]] \leq 2e^{\frac{-\epsilon^2 \mathbb{E}[S]}{3}} = 2e^{\frac{-\epsilon^2 d}{6}}$$

Probabilistic Method

Conclusion from Chernoff bound:

$$\text{For } \underline{\text{any}}\ i, j \text{ pair, } \Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 1 - 2e^{-\epsilon^2 d/6}.$$

By a union bound:

$$\text{For } \underline{\text{all}}\ i, j \text{ pairs simultaneously, } \Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 1 - \binom{t}{2} \cdot 2e^{-\epsilon^2 d/6}.$$

Probabilistic Method

Conclusion from Chernoff bound:

$$\text{For } \underline{\text{any}}\ i, j \text{ pair, } \Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 1 - 2e^{-\epsilon^2 d/6}.$$

By a union bound:

$$\text{For } \underline{\text{all}}\ i, j \text{ pairs simultaneously, } \Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 1 - \binom{t}{2} \cdot 2e^{-\epsilon^2 d/6}.$$

Suffices to set $t = e^{O(\epsilon^2 d)}$ to get for all pairs x_i, x_j in the set $x_1, x_2, \ldots, x_t$, $\Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 0.9$.

Probabilistic Method

Conclusion from Chernoff bound:

$$\text{For } \underline{\text{any}} \ i, j \text{ pair, } \Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 1 - 2e^{-\epsilon^2 d/6}.$$

By a union bound:

$$\text{For } \underline{\text{all}} \ i, j \text{ pairs simultaneously, } \Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 1 - \binom{t}{2} \cdot 2e^{-\epsilon^2 d/6}.$$

Suffices to set $t = e^{O(\epsilon^2 d)}$ to get for all pairs x_i, x_j in the set $x_1, x_2, \ldots, x_t$, $\Pr[|\mathbf{x}_i^T \mathbf{x}_j| < \epsilon] \geq 0.9$.

Therefore, there exists some set of $t = e^{O(\epsilon^2 d)}$ unit vectors with all pairwise inner products $\leq \epsilon$.

Orthogonal Vectors

Final result: In d-dimensional space, there are $2^{\Theta(\epsilon^2 d)}$ unit vectors with all pairwise inner products $\leq \epsilon$.

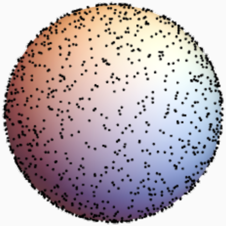
Corollary of proof: Random vectors tend to be far apart (and roughly equidistant) in high-dimensions.

Orthogonal Vectors

Final result: In d-dimensional space, there are $2^{\Theta(\epsilon^2 d)}$ unit vectors with all pairwise inner products $\leq \epsilon$.

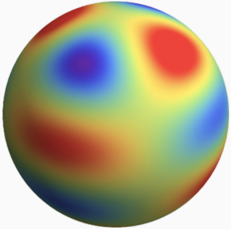
Corollary of proof: Random vectors tend to be far apart (and roughly equidistant) in high-dimensions.

$$\|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2\langle x_i, x_j \rangle \approx 2$$



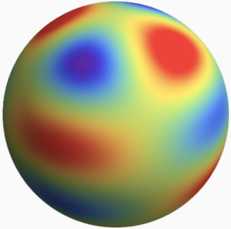
Curse of Dimensionality

Curse of dimensionality: Suppose we want to use e.g. k-nearest neighbors to learn a function or classify points in $\mathbb{R}^d$. If our data distribution is truly random, we typically need an exponential amount of data.



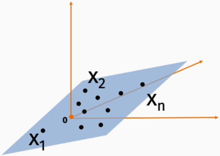
Curse of Dimensionality

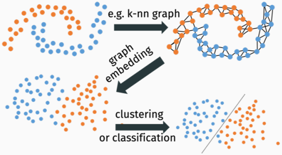
Curse of dimensionality: Suppose we want to use e.g. k-nearest neighbors to learn a function or classify points in $\mathbb{R}^d$. If our data distribution is truly random, we typically need an exponential amount of data.

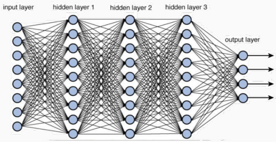


The existence of lower dimensional structure in our data is often the only reason we can hope to learn.

Low-dimensional structure.







For example, data lies on low-dimensional subspace, or does so after transformation. Or function can be represented by a restricted class of functions, like neural net with specific architecture.

Unit Ball in High Dimensions

Let $\mathcal{B}_d$ be the unit ball in d dimensions:

$$\mathcal{B}_d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}.$$

Unit Ball in High Dimensions

Let $\mathcal{B}_d$ be the unit ball in d dimensions:

$$\mathcal{B}_d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}.$$

Volume of radius R ball is $\frac{\pi^{d/2}}{(d/2)!} \cdot R^d$.

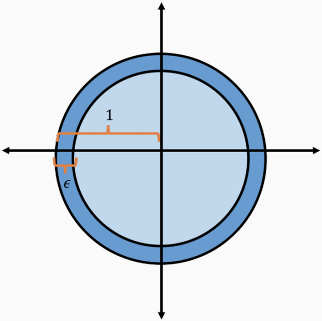
Unit Ball in High Dimensions

Let $\mathcal{B}_d$ be the unit ball in d dimensions:

$$\mathcal{B}_d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}.$$

Volume of radius R ball is $\frac{\pi^{d/2}}{(d/2)!} \cdot R^d$.

What percentage of volume of $\mathcal{B}_d$ falls within ϵ of its surface?



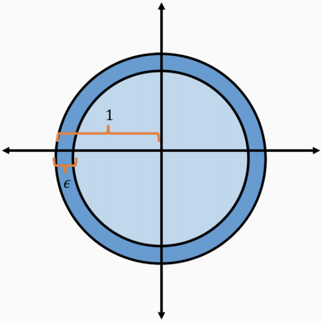
Unit Ball in High Dimensions

Let $\mathcal{B}_d$ be the unit ball in d dimensions:

$$\mathcal{B}_d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}.$$

Volume of radius R ball is $\frac{\pi^{d/2}}{(d/2)!} \cdot R^d$.

What percentage of volume of $\mathcal{B}_d$ falls within ϵ of its surface?



$$\frac{\text{Volume of Diff}}{\text{Total Volume}} = \frac{C \cdot 1^d - C \cdot (1-\epsilon)^d}{C \cdot 1^d}$$

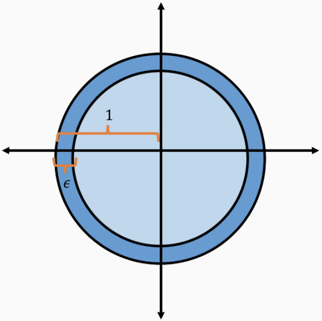
Unit Ball in High Dimensions

Let $\mathcal{B}_d$ be the unit ball in d dimensions:

$$\mathcal{B}_d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}.$$

Volume of radius R ball is $\frac{\pi^{d/2}}{(d/2)!} \cdot R^d$.

What percentage of volume of $\mathcal{B}_d$ falls within ϵ of its surface?



$$\frac{\text{Volume of Diff}}{\text{Total Volume}} = \frac{C \cdot 1^d - C \cdot (1-\epsilon)^d}{C \cdot 1^d}$$
$$= 1 - (1-\epsilon)^d$$

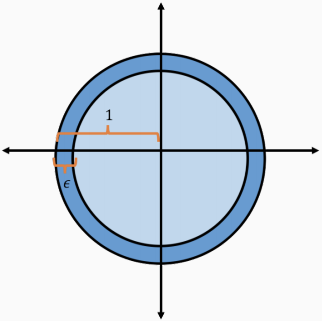
Unit Ball in High Dimensions

Let $\mathcal{B}_d$ be the unit ball in d dimensions:

$$\mathcal{B}_d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}.$$

Volume of radius R ball is $\frac{\pi^{d/2}}{(d/2)!} \cdot R^d$.

What percentage of volume of $\mathcal{B}_d$ falls within ϵ of its surface?



$$\frac{\text{Volume of Diff}}{\text{Total Volume}} = \frac{C \cdot 1^d - C \cdot (1-\epsilon)^d}{C \cdot 1^d}$$
$$= 1 - (1-\epsilon)^d$$
$$= 1 - \left((1-\epsilon)^{1/\epsilon}\right)^{\epsilon d}$$

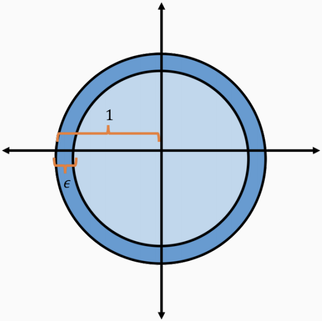
Unit Ball in High Dimensions

Let $\mathcal{B}_d$ be the unit ball in d dimensions:

$$\mathcal{B}_d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \leq 1\}.$$

Volume of radius R ball is $\frac{\pi^{d/2}}{(d/2)!} \cdot R^d$.

What percentage of volume of $\mathcal{B}_d$ falls within ϵ of its surface?



$$\begin{aligned}\frac{\text{Volume of Diff}}{\text{Total Volume}} &= \frac{C \cdot 1^d - C \cdot (1-\epsilon)^d}{C \cdot 1^d} \\ &= 1 - (1-\epsilon)^d \\ &= 1 - \left((1-\epsilon)^{1/\epsilon}\right)^{\epsilon d} \\ &\approx 1 - (1/e)^{\epsilon d} = 1 - e^{-\epsilon d}.\end{aligned}$$

Isoperimetric Inequality

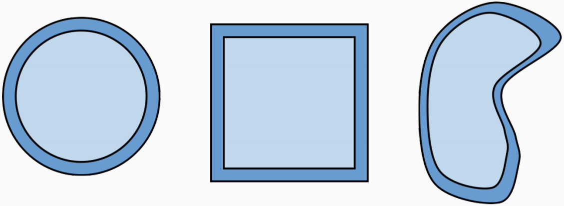
All but a vanishing small $2^{-\Theta(\epsilon d)}$ fraction of a unit ball's volume is within ϵ of its surface.

Isoperimetric Inequality: the ball has the minimum surface area/volume ratio of any shape.

Isoperimetric Inequality

All but a vanishing small $2^{-\Theta(\epsilon d)}$ fraction of a unit ball's volume is within ϵ of its surface.

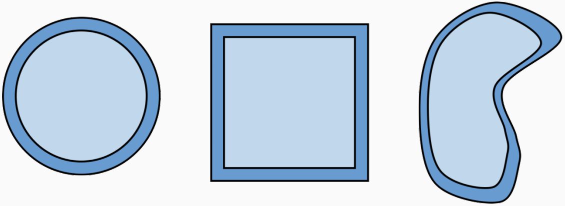
Isoperimetric Inequality: the ball has the minimum surface area/volume ratio of any shape.



Isoperimetric Inequality

All but a vanishing small $2^{-\Theta(\epsilon d)}$ fraction of a unit ball's volume is within ϵ of its surface.

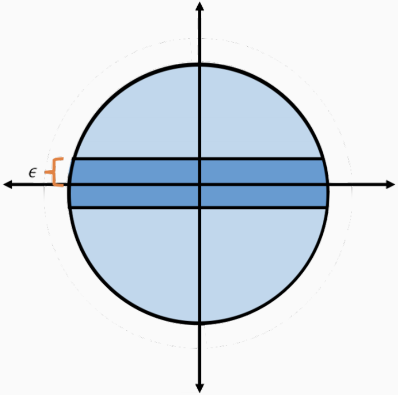
Isoperimetric Inequality: the ball has the minimum surface area/volume ratio of any shape.



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

Slices of the Unit Ball

What percentage of the volume of $\mathcal{B}_d$ falls within ϵ of its equator?

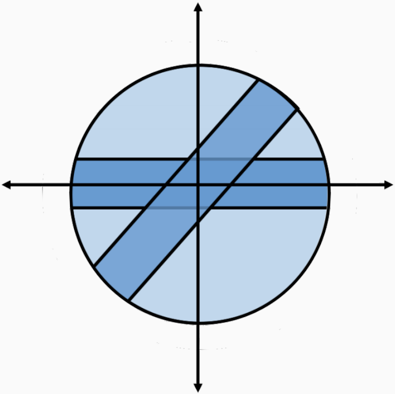


$$S = \{\mathbf{x} \in \mathcal{B}_d : |x_1| \leq \epsilon\}$$

Slices of the Unit Ball

What percentage of the volume of $\mathcal{B}_d$ falls within ϵ of its equator?

Answer: all but a $2^{-\Theta(\epsilon^2 d)}$ fraction.

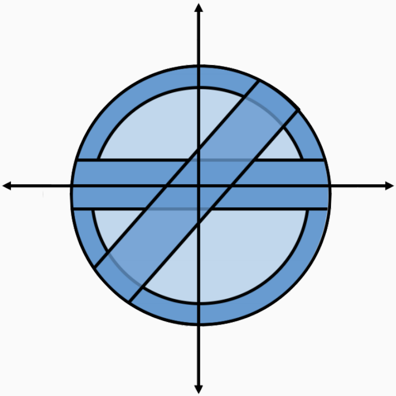


By symmetry, this is true for any equator:

$$S_{\mathbf{t}} = \{\mathbf{x} \in \mathcal{B}_d : \mathbf{x}^T \mathbf{t} \leq \epsilon\}.$$

Bizarre Shape of Unit Ball

1. $(1 - 2^{-\Theta(\epsilon d)})$ fraction of volume lies ϵ close to surface.
2. $(1 - 2^{-\Theta(\epsilon^2 d)})$ fraction of volume lies ϵ close to any equator.



High-dimensional ball looks nothing like 2D ball!

Concentration at Equator

Claim: All but a $2^{-\Theta(\epsilon^2 d)}$ fraction of the volume of the ball falls within ϵ of its equator.

Equivalent: Fix the equator to be horizontal. If we draw a point $\mathbf{x}$ randomly from the unit ball, $|x_1| \leq \epsilon$ with probability $\geq 1 - 2^{-\Theta(\epsilon^2 d)}$.

Concentration at Equator

Claim: All but a $2^{-\Theta(\epsilon^2 d)}$ fraction of the volume of the ball falls within ϵ of its equator.

Equivalent: Fix the equator to be horizontal. If we draw a point $\mathbf{x}$ randomly from the unit ball, $|x_1| \leq \epsilon$ with probability $\geq 1 - 2^{-\Theta(\epsilon^2 d)}$.

Proof Strategy:

- Instead of drawing random vectors from the unit ball, we will draw a random vector from the unit sphere (the surface of the ball).

Concentration at Equator

Claim: All but a $2^{-\Theta(\epsilon^2 d)}$ fraction of the volume of the ball falls within ϵ of its equator.

Equivalent: Fix the equator to be horizontal. If we draw a point $\mathbf{x}$ randomly from the unit ball, $|x_1| \leq \epsilon$ with probability $\geq 1 - 2^{-\Theta(\epsilon^2 d)}$.

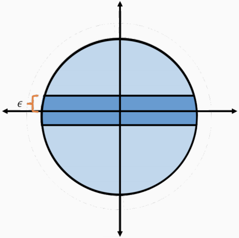
Proof Strategy:

- Instead of drawing random vectors from the unit ball, we will draw a random vector from the unit sphere (the surface of the ball).
- Let $\mathbf{w} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$, where $\mathbf{x}$ is a random vector from the unit ball.

Concentration at Equator

Claim: All but a $2^{-\Theta(\epsilon^2 d)}$ fraction of the volume of the ball falls within ϵ of its equator.

Equivalent: Fix the equator to be horizontal. If we draw a point $\mathbf{x}$ randomly from the unit ball, $|x_1| \leq \epsilon$ with probability $\geq 1 - 2^{-\Theta(\epsilon^2 d)}$.



Proof Strategy:

- Instead of drawing random vectors from the unit ball, we will draw a random vector from the unit sphere (the surface of the ball).
- Let $\mathbf{w} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$, where $\mathbf{x}$ is a random vector from the unit ball.
- $\Pr\left[|x_1| \leq \epsilon\right] \geq \Pr\left[|w_1| \leq \epsilon\right]$ (why?)

Concentration at Equator

Claim: $|w_1| \leq \epsilon$ with probability $\geq 1 - 2^{-\Theta(\epsilon^2 d)}$. This then proves our statement from the previous slide.

Concentration at Equator

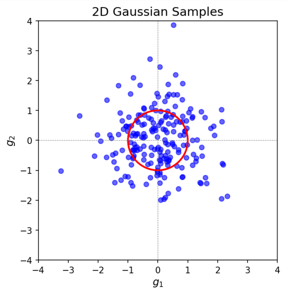
Claim: $|w_1| \leq \epsilon$ with probability $\geq 1 - 2^{-\Theta(\epsilon^2 d)}$. This then proves our statement from the previous slide.

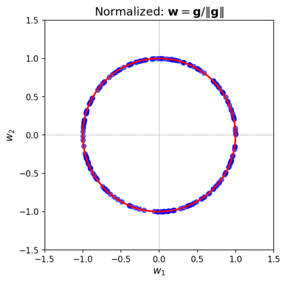
How can we generate w, which is a random vector taken from the unit sphere (the surface of the ball)?

Concentration at Equator

Claim: $|w_1| \leq \epsilon$ with probability $\geq 1 - 2^{-\Theta(\epsilon^2 d)}$. This then proves our statement from the previous slide.

How can we generate w, which is a random vector taken from the unit sphere (the surface of the ball)?





Important Fact in High Dimensional Geometry

Rotational Invariance of Gaussian distribution: Let $\mathbf{g}$ be a random Gaussian vector, with each entry drawn from $\mathcal{N}(0,1)$. Then $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$ is distributed uniformly on the unit sphere.

Why?

Important Fact in High Dimensional Geometry

Rotational Invariance of Gaussian distribution: Let $\mathbf{g}$ be a random Gaussian vector, with each entry drawn from $\mathcal{N}(0, 1)$. Then $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$ is distributed uniformly on the unit sphere.

Why? The PDF of a high-dimensional Gaussian does not depend on the direction of $\mathbf{g}$, only on its length.

Important Fact in High Dimensional Geometry

Rotational Invariance of Gaussian distribution: Let $\mathbf{g}$ be a random Gaussian vector, with each entry drawn from $\mathcal{N}(0,1)$. Then $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$ is distributed uniformly on the unit sphere.

Why? The PDF of a high-dimensional Gaussian does not depend on the direction of $\mathbf{g}$, only on its length.

Consider the probability density function of a high dimensional Gaussian:

$$p(\mathbf{g}) = p(g[1]) \cdot \ldots \cdot p(g[d]) = \prod_{i=1}^{d} ce^{-\mathbf{g}[i]^2/2}$$

Important Fact in High Dimensional Geometry

Rotational Invariance of Gaussian distribution: Let $\mathbf{g}$ be a random Gaussian vector, with each entry drawn from $\mathcal{N}(0, 1)$. Then $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$ is distributed uniformly on the unit sphere.

Why? The PDF of a high-dimensional Gaussian does not depend on the direction of $\mathbf{g}$, only on its length.

Consider the probability density function of a high dimensional Gaussian:

$$
\begin{aligned}
p(\mathbf{g}) = p(g[1]) \cdot \ldots \cdot p(g[d]) &= \prod_{i=1}^{d} c e^{-\mathbf{g}[i]^2/2} \\
&= c^d e^{\sum_{i=1}^{d} -g[i]^2/2}
\end{aligned}
$$

Important Fact in High Dimensional Geometry

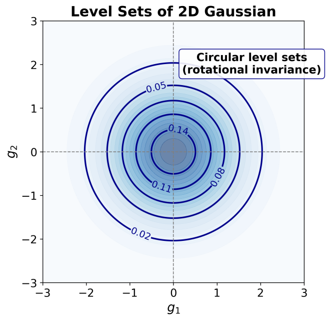
Rotational Invariance of Gaussian distribution: Let $\mathbf{g}$ be a random Gaussian vector, with each entry drawn from $\mathcal{N}(0,1)$. Then $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$ is distributed uniformly on the unit sphere.

Why? The PDF of a high-dimensional Gaussian does not depend on the direction of $\mathbf{g}$, only on its length.

Consider the probability density function of a high dimensional Gaussian:

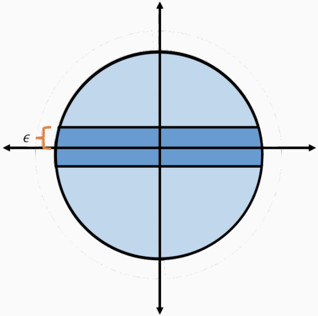
$$\begin{aligned} p(\mathbf{g}) = p(g[1]) \cdot \ldots \cdot p(g[d]) &= \prod_{i=1}^{d} c e^{-\mathbf{g}[i]^2/2} \\ &= c^d e^{\sum_{i=1}^{d} -g[i]^2/2} \\ &= c^d e^{-\|\mathbf{g}\|_2^2/2} \end{aligned}$$

Rotational Invariance of Gaussian distribution: Let $\mathbf{g}$ be a random Gaussian vector, with each entry drawn from $\mathcal{N}(0,1)$. Then $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$ is distributed uniformly on the unit sphere.



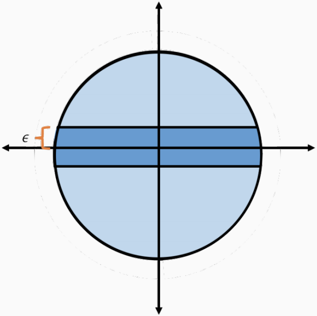
Proof Strategy

Draw $\mathbf{g} \sim \mathcal{N}(\mathbf{0}.\mathbf{I})$. Show that first entry of $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2 \leq \epsilon$ with very high probability.



Proof Strategy

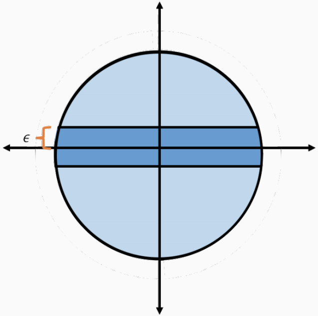
Draw $\mathbf{g} \sim \mathcal{N}(\mathbf{0}.\mathbf{I})$. Show that first entry of $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2 \leq \epsilon$ with very high probability.



1. Prove that with high probability, the first entry of $\mathbf{g}/\sqrt{d}$ is small.

Proof Strategy

Draw $\mathbf{g} \sim \mathcal{N}(\mathbf{0}.\mathbf{I})$. Show that first entry of $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2 \leq \epsilon$ with very high probability.



1. Prove that with high probability, the first entry of $\mathbf{g}/\sqrt{d}$ is small.
2. Prove that $\mathbf{g}/\sqrt{d}$ is very very close to $\mathbf{g}/\|\mathbf{g}\|_2^2$, so this vector also has small first entry.

Concentration at Equator

Let $\mathbf{g}$ be a random Gaussian vector and $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$.

$$\mathbb{E}[\|\mathbf{g}\|_2^2] = \sum_{i=1}^{d} \mathbb{E}[g_i^2] = \sum_{i=1}^{d} \mathrm{Var}[g_i] = d.$$

Concentration at Equator

Let $\mathbf{g}$ be a random Gaussian vector and $\mathbf{w} = \mathbf{g}/\|\mathbf{g}\|_2$.

$$\mathbb{E}[\|\mathbf{g}\|_2^2] = \sum_{i=1}^{d} \mathbb{E}[g_i^2] = \sum_{i=1}^{d} \mathrm{Var}[g_i] = d.$$

Excercise for home: Prove that

$$\Pr\left[\|\mathbf{g}\|_2^2 \leq \frac{1}{2}\mathbb{E}[\|\mathbf{g}\|_2^2]\right] \leq 2^{-\Theta(d)}$$

This should intuitively make sense. Can you tell me why?

Concentration at Equator

For $1 - 2^{-\Theta(d)}$ fraction of vectors $\mathbf{g}$, $\|\mathbf{g}\|_2 \geq \sqrt{d/2}$. Condition on the event that we get a random vector in this set.

Concentration at Equator

For $1 - 2^{-\Theta(d)}$ fraction of vectors $\mathbf{g}$, $\|\mathbf{g}\|_2 \geq \sqrt{d/2}$. Condition on the event that we get a random vector in this set.

Recall that $\mathbf{w} = \frac{\mathbf{g}}{\|\mathbf{g}\|_2}$. Given this event:

$$\Pr\left[|w_1| \leq \epsilon\right] = \Pr\left[|g_1/\|\mathbf{g}\|_2| \leq \epsilon\right]$$

Concentration at Equator

For $1 - 2^{-\Theta(d)}$ fraction of vectors $\mathbf{g}$, $\|\mathbf{g}\|_2 \geq \sqrt{d/2}$. Condition on the event that we get a random vector in this set.

Recall that $\mathbf{w} = \frac{\mathbf{g}}{\|\mathbf{g}\|_2}$. Given this event:

$$\begin{aligned} \Pr\left[|w_1| \leq \epsilon\right] &= \Pr\left[|g_1/\|\mathbf{g}\|_2| \leq \epsilon\right] \\ &\geq \Pr\left[|g_1|/\sqrt{d/2} \leq \epsilon\right] \end{aligned}$$

Concentration at Equator

For $1 - 2^{-\Theta(d)}$ fraction of vectors $\mathbf{g}$, $\|\mathbf{g}\|_2 \geq \sqrt{d/2}$. Condition on the event that we get a random vector in this set.

Recall that $\mathbf{w} = \frac{\mathbf{g}}{\|\mathbf{g}\|_2}$. Given this event:

$$\begin{aligned}\Pr\left[|w_1| \leq \epsilon\right] &= \Pr\left[|g_1/\|\mathbf{g}\|_2| \leq \epsilon\right] \\ &\geq \Pr\left[|g_1|/\sqrt{d/2} \leq \epsilon\right] \\ &= \Pr\left[|g_1| \leq \epsilon \cdot \sqrt{d/2}\right]\end{aligned}$$

Concentration at Equator

For $1 - 2^{-\Theta(d)}$ fraction of vectors $\mathbf{g}$, $\|\mathbf{g}\|_2 \geq \sqrt{d/2}$. Condition on the event that we get a random vector in this set.

Recall that $\mathbf{w} = \frac{\mathbf{g}}{\|\mathbf{g}\|_2}$. Given this event:

$$\begin{aligned}
\Pr\left[|w_1| \leq \epsilon\right] &= \Pr\left[|g_1/\|\mathbf{g}\|_2| \leq \epsilon\right] \\
&\geq \Pr\left[|g_1|/\sqrt{d/2} \leq \epsilon\right] \\
&= \Pr\left[|g_1| \leq \epsilon \cdot \sqrt{d/2}\right] \\
&\geq 1 - 2^{-\theta\left((\epsilon \cdot \sqrt{d/2})^2\right)} \qquad \text{(Integrate the PDF)}
\end{aligned}$$

Concentration at Equator

For $1 - 2^{-\Theta(d)}$ fraction of vectors $\mathbf{g}$, $\|\mathbf{g}\|_2 \geq \sqrt{d/2}$. Condition on the event that we get a random vector in this set.

Recall that $\mathbf{w} = \frac{\mathbf{g}}{\|\mathbf{g}\|_2}$. Given this event:

$$\begin{aligned}
\Pr\left[|w_1| \leq \epsilon\right] &= \Pr\left[|g_1/\|\mathbf{g}\|_2| \leq \epsilon\right] \\
&\geq \Pr\left[|g_1|/\sqrt{d/2} \leq \epsilon\right] \\
&= \Pr\left[|g_1| \leq \epsilon \cdot \sqrt{d/2}\right] \\
&\geq 1 - 2^{-\theta\left((\epsilon\cdot\sqrt{d/2})^2\right)} \qquad \text{(Integrate the PDF)}
\end{aligned}$$

Concentration at Equator

For $1 - 2^{-\Theta(d)}$ fraction of vectors $\mathbf{g}$, $\|\mathbf{g}\|_2 \geq \sqrt{d/2}$. Condition on the event that we get a random vector in this set.

Recall that $\mathbf{w} = \frac{\mathbf{g}}{\|\mathbf{g}\|_2}$. Given this event:

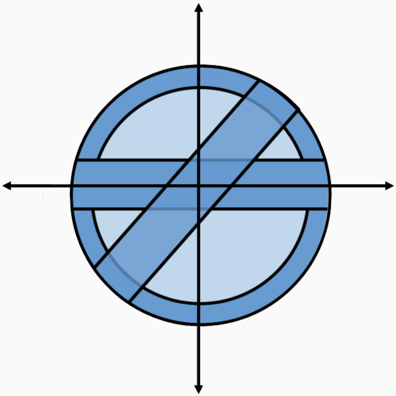
$$\begin{aligned}
\Pr\left[|w_1| \leq \epsilon\right] &= \Pr\left[|g_1/\|\mathbf{g}\|_2| \leq \epsilon\right] \\
&\geq \Pr\left[|g_1|/\sqrt{d/2} \leq \epsilon\right] \\
&= \Pr\left[|g_1| \leq \epsilon \cdot \sqrt{d/2}\right] \\
&\geq 1 - 2^{-\theta\left((\epsilon\cdot\sqrt{d/2})^2\right)} \qquad \text{(Integrate the PDF)}
\end{aligned}$$

By union bound, overall we have:

$$\Pr\left[|w_1| \leq \epsilon\right] \geq 1 - 2^{-\Theta\left(\epsilon^2 d\right)} - 2^{-\theta(d)}$$

Bizarre Shape of Unit Ball

1. $(1 - 2^{-\Theta(\epsilon d)})$ fraction of volume lies ϵ close to surface.
2. $(1 - 2^{-\Theta(\epsilon^2 d)})$ fraction of volume lies ϵ close to any equator.

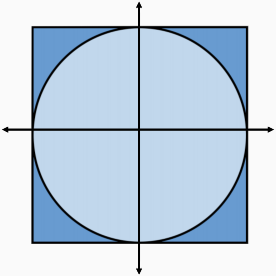


High-dimensional ball looks nothing like 2D ball!

High Dimensional Cube

Let $\mathcal{C}_d$ be the d-dimensional cube:

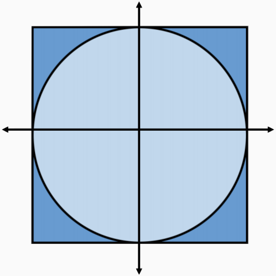
$$\mathcal{C}_d = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}(i)| \leq 1 \ \forall i\}.$$



High Dimensional Cube

Let $\mathcal{C}_d$ be the d-dimensional cube:

$$\mathcal{C}_d = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}(i)| \leq 1 \ \forall i\}.$$

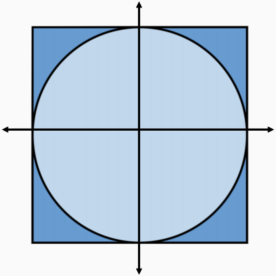


In two dimensions, the cube is pretty similar to the ball.

High Dimensional Cube

Let $\mathcal{C}_d$ be the d-dimensional cube:

$$\mathcal{C}_d = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}(i)| \leq 1 \ \forall i\}.$$



In two dimensions, the cube is pretty similar to the ball.

But volume of $\mathcal{C}_d$ is 2^d while volume of unit ball is $\frac{\sqrt{\pi}^d}{(d/2)!}$.

This is a huge gap! Cube has $O(d)^{O(d)}$ more volume.

High Dimensional Cube

Some other ways to see these shapes are very different:

- $\max_{\mathbf{x} \in \mathcal{B}_d} \|\mathbf{x}\|_2^2 =$

High Dimensional Cube

Some other ways to see these shapes are very different:

- $\max_{\mathbf{x}\in\mathcal{B}_d} \|\mathbf{x}\|_2^2 = 1$
- $\max_{\mathbf{x}\in\mathcal{C}_d} \|\mathbf{x}\|_2^2 =$

High Dimensional Cube

Some other ways to see these shapes are very different:

- $\max_{\mathbf{x} \in \mathcal{B}_d} \|\mathbf{x}\|_2^2 = 1$
- $\max_{\mathbf{x} \in \mathcal{C}_d} \|\mathbf{x}\|_2^2 = d$
- $\mathbb{E}_{\mathbf{x} \sim \mathcal{B}_d} \|\mathbf{x}\|_2^2 =$

High Dimensional Cube

Some other ways to see these shapes are very different:

- $\max_{\mathbf{x} \in \mathcal{B}_d} \|\mathbf{x}\|_2^2 = 1$
- $\max_{\mathbf{x} \in \mathcal{C}_d} \|\mathbf{x}\|_2^2 = d$
- $\mathbb{E}_{\mathbf{x} \sim \mathcal{B}_d} \|\mathbf{x}\|_2^2 = \frac{d}{d+2} \approx 1$
- $\mathbb{E}_{\mathbf{x} \sim \mathcal{C}_d} \|\mathbf{x}\|_2^2 =$

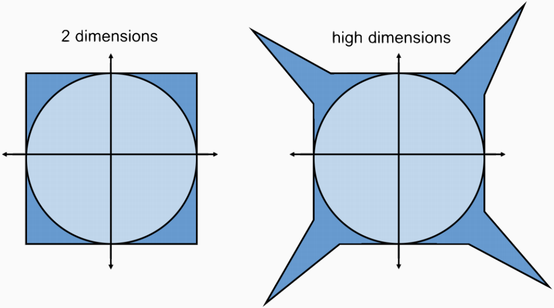
High Dimensional Cube

Some other ways to see these shapes are very different:

- $\max_{\mathbf{x} \in \mathcal{B}_d} \|\mathbf{x}\|_2^2 = 1$
- $\max_{\mathbf{x} \in \mathcal{C}_d} \|\mathbf{x}\|_2^2 = d$
- $\mathbb{E}_{\mathbf{x} \sim \mathcal{B}_d} \|\mathbf{x}\|_2^2 = \frac{d}{d+2} \approx 1$
- $\mathbb{E}_{\mathbf{x} \sim \mathcal{C}_d} \|\mathbf{x}\|_2^2 = \sum_{i=1}^{d} \mathbb{E}[x_i^2] = d \cdot \frac{1}{3}$

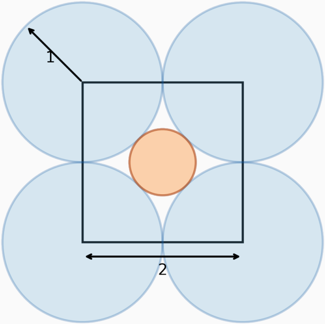
High Dimensional Cube

Almost all of the volume of the unit cube falls in its corners, and these corners lie far outside the unit ball.



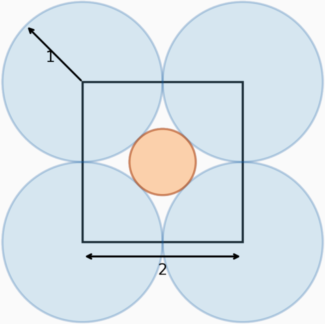
On Breaking Intuition

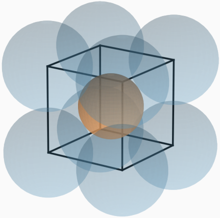
Consider the following setup:



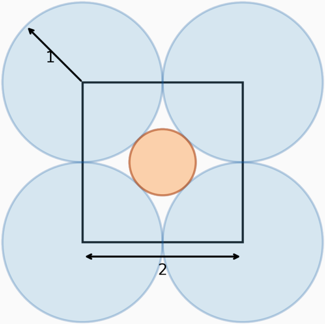
On Breaking Intuition

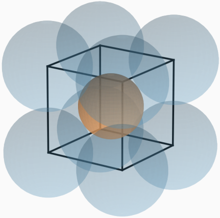
Consider the following setup:





Consider the following setup:

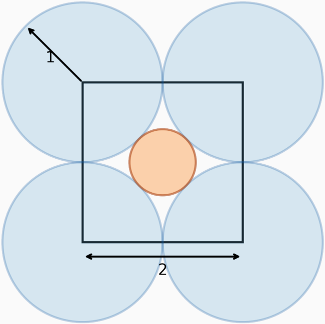


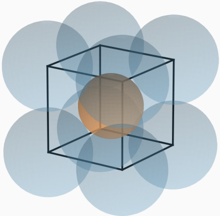


Place 2^d unit balls in box with side length 2. Look at sphere they enclose. It has radius $\sqrt{d} - 1$.

On Breaking Intuition

Consider the following setup:





Place 2^d unit balls in box with side length 2. Look at sphere they enclose. It has radius $\sqrt{d} - 1$.

So for $d \geq 10$, the central ball sticks out of the box...