Lecture 01: Curse of Dimensionality: Problem and Solutions

Nikola Zlatanov

Innopolis University

Advanced Statistics

30-th of January, 2023

Motivation - Big Data

- Under the 'umbrella' term Big Data, the following is understood to hold:

  a) Very big number of observations (or data points)

  b) Very big number of dimensions (or parameters) associated with each data point
- Example for a)
  - The monthly incomes of the residents of Tatarstan
  - Hence, each month we have 3.8 million observations (or data points), which is big.
  - In one year, we will have 45.8 million observations, which is big.
  - BUT NOTE, the dimension is one, since for each person we record only one number. $[x_1, x_2 \ldots x_{155200}]$
- Example for b)
  - The HD image of the residents of Tatarstan
  - Each image is an observation and each observation is comprised of $1440 \times 1080 = 1,555,200$ pixels, and each pixel represents one dimension. Hence, 1,555,200 dimensions per observation – big num.
  - Other examples: text, sound, video, human genome, etc

Difficulty - Big Data

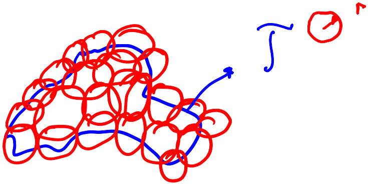
- Empirically (i.e., by experience), it is exponentially harder to deal with large number of dimensions than to deal with large number of observations
- Large number of observations is usually not a problem. This has been dealt with tools from classical statistics.
- In fact, the more observations we help, the more we can understanding the statistics (i.e., of what is going on) of the underlying random process.
- But large number of dimensions makes the problem of understanding the statistics very difficult.
- In fact, adding more dimensions to an observation makes the problem of understanding the statistics exponentially more difficult.

Volume of Big Data - Curse of Dimensional

- Example of volume of a cube in high dimensions:
  - The volume of a 2-dimensional cube, where each size is a, is $a \times a = a^2$. For $a = 2$, the volume is 2^2
  - The volume of a 3-dimensional cube, where each size is a, is $a \times a \times a = a^3$. For $a = 2$, the volume is 2^3
  - The volume of a d-dimensional cube, where each size is a, is $a \times a \times ... \times a = a^d$. For $a = 2$, the volume is 2^d.
  - Hence, the volume grows exponentially with the dimension d. This means the for higher dimensions, there is exponentially more space than in lower dimensions.
- This phenomenon is known as the "curse of high-dimensionality" (CD), and is the main problem in DS and ML.
- We have to find methods that can go around the "curse of high-dimensionality", if possible.
- This course is aimed to help you understand how to do that.

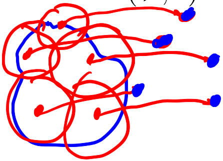
Example of ~~ACT~~ CD: Covering Number of Balls

- Assume we have an object, described as a set $\mathcal{T}$, and assume we have N balls or radius r. We would like to completely cover the object with the minimum number of balls. Let us denote this number as $N(\mathcal{T}, r)$. Now how to find $N(\mathcal{T}, r)$ for a given $\mathcal{T}$ and r?



Example of ACT: Covering Number of Balls

- Def: The covering number of a set $\mathcal{T} \subset \mathbb{R}^d$ and given radius r, is the smallest number of Euclidean balls with radius r that completely cover the set $\mathcal{T}$.
- The covering number $N(\mathcal{T}, r)$ is a descriptor of the complexity of $\mathcal{T}$. The less the number is, the easier is to describe $\mathcal{T}$ using only $N(\mathcal{T}, r)$ balls of radius r
- Other meaning: The covering is simply the quantization of $\mathcal{T}$, whereas the covering number gives us the minimum number of quantized points that describe $N(\mathcal{T}, r)$ with a quantization error r.

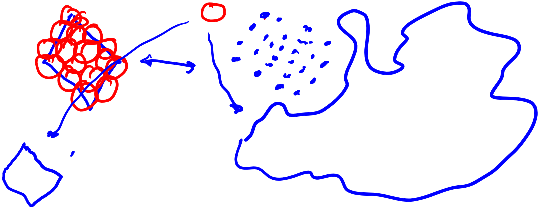


$1.353 \to 1$

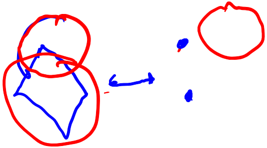
$1.56 \to 2$

Example of ACT: Covering Number of Balls

- More accurate quantization for small r, but more points.



- More accurate quantization for larger r, but less points.



Example of ACT: Covering Number of Balls

- The covering number $N(\mathcal{T}, r)$ suffers from the curse of dimensionality: The higher the dimension d of $\mathcal{T} \subset \mathbb{R}^d$ is, exponentially more balls are needed to cover $\mathcal{T}$, i.e., it is exponentially harder to describe it via balls of fixed radius.
- Fact: If $\mathcal{T} \subset \mathbb{R}^d$ is the unit (radius one) Euclidean ball in dimension d, and we need to cover it by balls of radius $r = 1/2$, the we would need $N(\mathcal{T}, 1/2) \geq 2^d$ number of balls.
- Note, the above number is exponential in d, hence, CD hits!
- Proof of the Fact: The volume of a r-radius ball in d dimension is

$$V(d,r) = \frac{\pi^{d/2}}{\Gamma(d/2+1)} r^d$$

$\Gamma(n) = (n-1)!$

Hence, $$\frac{\text{V of big ball}}{\text{V of small ball}} = \frac{V(d,1)}{V(d,1/2)} = \frac{1}{1/2^d} = 2^d$$

On the other hand $V(d,1) \leq N(\mathcal{T}, 1/2) V(d, 1/2)$, from where we obtain $N(\mathcal{T}, 1/2) \geq V(d,1/2)/V(d,1/2) = 2^d$.

Example of CD: Covering Number of Balls

- Note on the influence of the dimension on the volume of balls:
- Comparing the unit ball with any other ball of radius $r < 1$, we obtain

$r = 1 - \varepsilon$

$\varepsilon \to 0$

$$\frac{\text{V of smaller ball}}{\text{V of unit ball}} = \frac{V(d,r)}{V(d,1)} = r^d \to 0 \text{ as } d \to \infty$$

- Comparing the unit ball with any other ball of radius $r > 1$, we obtain

$r = 1 + \varepsilon$

$\varepsilon \to 0$

$$\frac{\text{V of larger ball}}{\text{V of unit ball}} = \frac{V(d,r)}{V(d,1)} = r^d \to \infty \text{ as } d \to \infty$$

- Hence, when the dimension increases, the balls bifurcate into two classes of balls compared to the unit ball
  - One class is balls with $r < 1$: These balls are almost infinitely smaller than the unit ball
  - Second class is balls with $r > 1$: These balls are almost infinitely larger than the unit ball

Example of CD: Covering Number of Balls

- The above CD happens not just for balls, but for many other objects in high-dimensions. In fact, for almost all high-dimensional objects.

  few/∞ → 0

  - This means that it is very likely we to work with high-dimensional data that suffers from the CD, since almost all high-dimensional objects suffer from the CD.
- There are only few high-dimensional objects that do not suffer from the CD. One of these is the polytope.
  - This means that it is very unlikely we to work with high-dimensional data that does not suffers from the CD, since there are few high-dimensional objects that do not suffer from the CD.
- Then how can we avoid CD? We avoid it by not seeking to obtain exactly correct results. Instead, we seek to obtain approximately correct results.

Example of Overcoming CD - Numerical Integration Dd

- Numerical integration in dimension d: Let us have a d-dimensional function $g(x_1, x_2, ..., x_d)$. We want to numerically compute the following integral

$$I_d = \int_0^1 \int_0^1 .... \int_0^1 g(x_1, x_2, ..., x_d) dx_1 dx_2 ... dx_d$$

Example of Overcoming CD - Numerical Integration D1

- When $d = 1$, we have

$$I_1 = \int_0^1 g(x)dx$$

- To approximately compute I_1, we pick n equally-distant points $\hat{x}_1, \hat{x}_2, ..., \hat{x}_n$, each in D1, with distance ϵ between neighboring points, and compute

$$\frac{1}{n}\sum_{i=1}^{n} g(\hat{x}_i) \approx I_1$$

$\varepsilon \to 0$

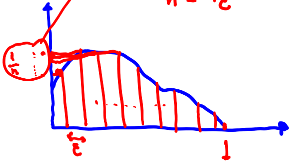
- Note that we have used $n = 1/\epsilon$ number of points

$n \to \infty$

$n \cdot \varepsilon = 1$

$n = 1/\varepsilon$

- Figure:



Example of Overcoming CD - Numerical Integration D2

- When $d = 2$, we have

$$I_2 = \int_0^1 \int_0^1 g(x_1, x_2) dx_1 dx_2$$

$[x_{11}\ x_{12}]$ $[x_{21}\ x_{22}]$ $[x_{n1}\ x_{n2}]$

- To approximately compute I_2, we pick n equally-distant points $\boldsymbol{\hat{x}_1}, \boldsymbol{\hat{x}_1}, ..., \boldsymbol{\hat{x}_n}$ in D2, with distance ϵ between neighboring points in each dimension, and perform

$$\frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{\hat{x}_i}) = \frac{1}{n} \sum_{i=1}^{n} g(\hat{x}_{1i}, \hat{x}_{2i}) \approx I_2$$

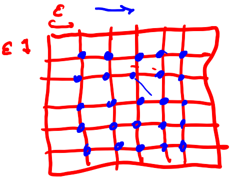
- Note that each $\boldsymbol{\hat{x}_i}$ represents a vector of two points given by $\boldsymbol{\hat{x}_i} = [\hat{x}_{1i}, \hat{x}_{2i}]$, where $\hat{x}_{1i}$ is the point from first dimension and $\hat{x}_{2i}$ is the point from second dimension.
- Note that $\epsilon = x_{1i+1} - x_{1i} = x_{2i+1} - x_{2i}$
- Note that we have used $n = 1/\epsilon^2$ number of points, which can be seen in the next figure.

Example of Overcoming CD - Numerical Integration D2

- Figure:

$\hat{n}_2 \cdot \varepsilon = 1$

$\hat{n}_2 = 1/\varepsilon$



$\hat{n}_1 \cdot \varepsilon = 1 \Rightarrow \hat{n}_1 = 1/\varepsilon$

$\hat{n}_1 \cdot \hat{n}_2 = \frac{1}{\varepsilon^2}$

Example of Overcoming CD - Numerical Integration Dd

- When $d = d$, we have

$$I_d = \int_0^1 \int_0^1 .... \int_0^1 g(x_1, x_2, ..., x_d) dx_1 dx_2 ... dx_d$$

each vec has d elem but we used $n = (\frac{1}{\epsilon})^d$ vectors

To approximate compute I_d, we pick n equally-distant points $\hat{\boldsymbol{x}}_{\boldsymbol{1}}, \hat{\boldsymbol{x}}_{\boldsymbol{1}}, ..., \hat{\boldsymbol{x}}_{\boldsymbol{n}}$ in Dd, with distance ϵ between neighboring points in each dimension, and perform

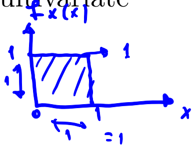
$$\frac{1}{n} \sum_{i=1}^{n} g(\hat{\boldsymbol{x}}_{\boldsymbol{i}}) = \frac{1}{n} \sum_{i=1}^{n} g(\hat{x}_{1i}, \hat{x}_{2i}, ..., \hat{x}_{di}) \approx I_d$$

- Note that each $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$ represents a vector of d points given by $\hat{\boldsymbol{x}}_{\boldsymbol{1}} = [\hat{x}_{1i}, \hat{x}_{2i}, ..., \hat{x}_{di}]$, where $\hat{x}_{ki}$ is the point in the k-th dimension of $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$, for $k = 1, 2, ..., d$ and $i = 1, 2, ..., n$.
- Note that we have used $n = 1/\epsilon^d$ number of points, and this cannot be even illustrated by a figure.
- If d is large we will be out of memory $\rightarrow$ Uncomputable for large d!

Example of Overcoming CD - Numerical Integration Dd

- Now lets try to compute the d-dimensional integral using probabilistic tools.
- Instead of picking ϵ-distant $1/\epsilon^d$ points, lets pick n random points $\hat{\boldsymbol{x}}_{\mathbf{1}}, \hat{\boldsymbol{x}}_{\mathbf{1}}, ..., \hat{\boldsymbol{x}}_{\boldsymbol{n}}$, where each $\hat{\boldsymbol{x}}_{\boldsymbol{i}} = [\hat{x}_{1i}, \hat{x}_{2i}, ..., \hat{x}_{di}]$ is chosen independently and identically according to the multivariate uniform distribution

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{k=1}^{d} f_X(x_k),$$



where

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{if otherwise} \end{cases}$$

- Then let's compute

$$\frac{1}{n}\sum_{i=1}^{n} g(\hat{\boldsymbol{x}}_{\boldsymbol{i}}) = \frac{1}{n}\sum_{i=1}^{n} g(\hat{x}_{1i}, \hat{x}_{2i}, ..., \hat{x}_{di})$$

- How close is the above sum to I_d ?

Example of Overcoming CD - Numerical Integration Dd

- If we repeat the above process many times, we can consider that $\hat{\boldsymbol{x}}_{\boldsymbol{1}}, \hat{\boldsymbol{x}}_{\boldsymbol{1}}, ..., \hat{\boldsymbol{x}}_{\boldsymbol{n}}$ are n independent and identically distributed (i.i.d.) random vectors.
- Then we can compute the average of the sum as

$\frac{n}{n} E(g(\hat{x}_1))$

$$E\left[\frac{1}{n}\sum_{i=1}^{n} g(\hat{\boldsymbol{x}}_{\boldsymbol{i}})\right] = \frac{1}{n}\sum_{i=1}^{n} E[g(\hat{\boldsymbol{x}}_{\boldsymbol{i}})] \overset{(a)}{=} E[g(\hat{\boldsymbol{x}})],$$

where (a) follows due to the i.i.d. construction of the $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$'s. Next,

$$E[g(\hat{\boldsymbol{x}})] = \int_0^1 \int_0^1 .... \int_0^1 g(x_1, x_2, ..., x_d) \prod_{k=1}^{d} f_X(x_k) dx_1 dx_2 ... dx_d$$

$$\overset{(b)}{=} \int_0^1 \int_0^1 .... \int_0^1 g(x_1, x_2, ..., x_d) dx_1 dx_2 ... dx_d = I_d$$

where (b) follows due to the definition of $f_X(x)$.

- We have obtained that the constructed sum is an unbiased estimator of I_d.

Example of Overcoming CD - Numerical Integration Dd

- We have obtained that the average of the constructed sum is equal to I_d. But, what is the error we are making? We can compute the mean squared error (MSE) as

outside the sum

$$E\left[\left(\frac{1}{n}\sum_{i=1}^{n} g(\hat{\boldsymbol{x}}_{\boldsymbol{i}}) - I_d\right)^2\right] = \frac{1}{n^2} E\left[\left(\underbrace{\sum_{i=1}^{n} g(\hat{\boldsymbol{x}}_{\boldsymbol{i}})}_{RV} - \underbrace{nI_d}_{E(RV)}\right)^2\right]$$

$$= \frac{1}{n^2} VAR\left[\sum_{i=1}^{n} g(\hat{\boldsymbol{x}}_{\boldsymbol{i}})\right] \overset{(a)}{=} \frac{1}{n^2}\sum_{i=1}^{n} VAR\left[g(\hat{\boldsymbol{x}}_{\boldsymbol{i}})\right] = \frac{1}{n^2} n VAR[g(\hat{\boldsymbol{x}})]$$

$$= \frac{1}{n} VAR[g(\hat{\boldsymbol{x}})], \tag{1}$$

$n = 10000$

$\frac{1}{100} \cdot VAR(g(x)) \le \frac{1}{100} = 0.01 = 1\%$

where (a) is due to the i.i.d.

- Hence, the MSE decays with $1/n$. This means that the absolute error, defined as $\sqrt{\text{MSE}}$, decays with $1/\sqrt{n}$
- Note, the absolute error (and MSE) is independent from the dimension d! We have avoided the curse of dimensionality!

Example of Overcoming CD - Numerical Integration Dd

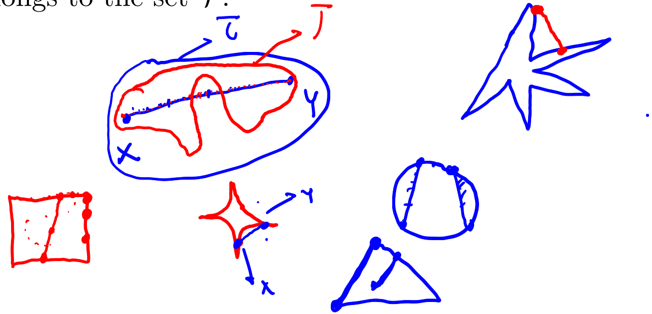
- Figure of the absolute error:

$$\sqrt{E\left\{\left(\frac{1}{n}\sum_i g(x_i) - I_d\right)^2\right\}} = O\left(\frac{1}{\sqrt{n}}\right)$$

Example of Overcoming CD - Computational Geometry

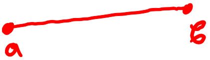
Def: A set $\mathcal{T} \subset \mathbb{R}^d$ is called a convex set if $\forall x \in \mathcal{T}$ and $\forall y \in \mathcal{T}$, the point $\lambda x + (1-\lambda)y \in \mathcal{T}$, for any $0 \leq \lambda \leq 1$.

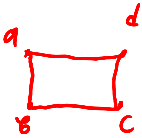
- The above simply means that any point on the line between x and y also belongs to the set $\mathcal{T}$.



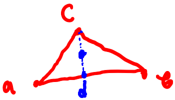
Example of Overcoming CD - Computational Geometry

- A way of transforming a non-convex set into a convex set is called taking The Convex Hull
- Def: The convex hull of $\mathcal{T} \subset \mathbb{R}^d$, denoted by $\text{conv}(\mathcal{T})$ is the smallest convex set that contains $\mathcal{T}$.
- Note $\mathcal{T} \subset \text{conv}(\mathcal{T})$.
- Examples:
  - $\mathcal{T} = \{a, b\}$, then conv $(\mathcal{T}) = [a, b]$, which denotes the line segment between a





  - $\mathcal{T} = \{a, b, c\}$, then

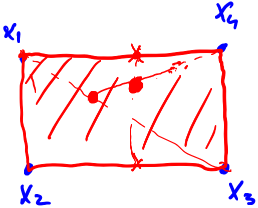


Example of Overcoming CD - Computational Geometry

- Fact: (Convex Combination) If $\boldsymbol{x} \in \text{conv}(\mathcal{T})$, then $\boldsymbol{x} = \sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i$, where $\boldsymbol{x}_i \in \mathcal{T}$ and $\lambda_i \geq 0$, $\forall i$, and $\sum_{i=1}^{m} \lambda_i = 1$.
- Caratheodori Thm: $\forall \boldsymbol{x} \in \text{conv}(\mathcal{T})$, where $\mathcal{T} \subset \mathbb{R}^d$, can be expressed as a convex combination of at most $d+1$ points (read vectors) in $\mathcal{T}$.
- The above means that any $\boldsymbol{x} \in \text{conv}(\mathcal{T})$ can be described using at most $d+1$ other vectors in $\mathcal{T}$, even though the set $\text{conv}(\mathcal{T})$ itself contains infinitely many vectors.
- Hence, we can store infinitely many vectors just by storing at most $d+1$ other vectors, which is a form of lossless compression.
- Other applications are in optimization: see paper "Fast and Accurate Least-Mean-Squares"

Example of Overcoming CD - Computational Geometry

Ex: $\mathcal{T} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}$, where $\forall \boldsymbol{x}_i \in \mathbb{R}^d$. Then $\text{conv}(\mathcal{T})$ is rectangle with ∞ points.

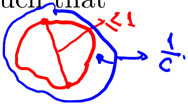


Example of Overcoming CD - Computational Geometry

- What if we want to describe/store all points in $\mathrm{conv}(\mathcal{T})$ with less then $d+1$ points/vectors, in order to save storage space, i.e., we want to do lossy compression?
- What kind of an error we would make in that compression process?
- In fact, it turns out that we can describe $\boldsymbol{x}$ with an error by choosing $m < d$, and m is independent of d. Avoided the CD!
- This is described in the following

Example of Overcoming CD - Computational Geometry

- Approximate Caratheodori Th (ACT): Let $\mathcal{T} \subset \mathbb{R}^d$, and let $\text{diam}(\mathcal{T}) \leq 1$. Then, $\forall \boldsymbol{x} \in \text{conv}(\mathcal{T})$ and $\forall k \in \mathbb{N}$, $\exists \boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k$ $\in \mathcal{T}$ such that



$$\left|\left|\boldsymbol{x} - \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{x}_k\right|\right|_2 \leq \frac{1}{\sqrt{2k}}$$

  - $\text{diam}(\mathcal{T}) \leq 1$ means that the length between any two points in $\mathcal{T}$ is at most one. This is just scaling, since we can always divide the points in $\mathcal{T}$ by a scalar such that $\text{diam}(\mathcal{T}) \leq 1$ holds.
  - Note the equal weights above, i.e., $\lambda_i = 1/k$, $\forall i$.
  - Note $||\boldsymbol{x}||_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$, where x_i is the i-th element of $\boldsymbol{x}$.

- **Compression:** Hence, instead of storing the infinite number of vectors in $\text{conv}(\mathcal{T})$ using at most $d+1$ vectors in $\mathcal{T}$, I can store $\text{conv}(\mathcal{T})$ using $k \ll d$ vectors in $\mathcal{T}$, and the squired error that I would make is at most $1/\sqrt{2k}$.

Example of Overcoming CD - Computational Geometry

$E[(X-\hat{X})^2] = E[X^2 - 2X\hat{X} + \hat{X}^2] = E(X^2) + E(\hat{X}^2) - 2E(X)\cdot E(\hat{X}) = 2E(X^2) - 2E(X)^2 = 2VAR$

$E(\hat{X}^2) = E(X^2)$

- How can we achieve the (ACT)? Using probability, and this is how:
- First, another definition of $VAR[X]$:

$$VAR[X] = \frac{1}{2}E[(X-\hat{X})^2],$$

where $\hat{X}$ is i.i.d. as X, and X is one dimensional. Prove this at home!

- Similarly, for random vectors: If $\boldsymbol{X} \in \mathbb{R}^d$, then

$$VAR[\boldsymbol{X}] = E\left[||\boldsymbol{X} - E[\boldsymbol{X}]||_2^2\right] = \frac{1}{2}E\left[||\boldsymbol{X} - \cancel{E[}\hat{\boldsymbol{X}}\cancel{]}||_2^2\right], \qquad (2)$$

where $\hat{\boldsymbol{X}}$ is i.i.d. as $\boldsymbol{X}$. Again, prove at home!

Example of Overcoming CD - Computational Geometry

- Proof of ACT: Fix $\boldsymbol{x} \in \operatorname{conv}(\mathcal{T})$. Then, we know that

$$\boldsymbol{x} = \sum_{i=1}^{d+1} \lambda_i \boldsymbol{z}_i,$$

where $\boldsymbol{z}_i \in \mathcal{T}$, $\lambda_i \geq 0$ and $\sum_{i=1}^{d+1} \lambda_i = 1$.
We now interpret λ_i's as probabilities and the above sum as expectation.
Let $\boldsymbol{Z}$ be a discrete random vector with PMF $\Pr\{\boldsymbol{Z} = \boldsymbol{z}_i\} = \lambda_i$.
Then,

$$\boldsymbol{x} = E[\boldsymbol{Z}] = \sum_{i=1}^{d+1} \lambda_i \boldsymbol{z}_i \tag{3}$$

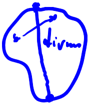
Example of Overcoming CD - Computational Geometry

Now let us take k RVs that are i.i.d. as $\boldsymbol{Z}$, given by $\boldsymbol{Z}_1, \boldsymbol{Z}_2, ..., \boldsymbol{Z}_k$. Then, we know that

$$E\left[\left|\left|\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{Z}_i - E[\boldsymbol{Z}]\right|\right|_2^2\right] \stackrel{(a)}{=} E\left[\left|\left|\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{Z}_i - \boldsymbol{x}\right|\right|_2^2\right]$$

$$= E\left[\left|\left|\frac{1}{k}\sum_{i=1}^{k}(\boldsymbol{Z}_i - \boldsymbol{x})\right|\right|_2^2\right] \stackrel{(a)}{=} \frac{1}{k^2}E\left[\left|\left|\sum_{i=1}^{k}(\boldsymbol{Z}_i - E[\boldsymbol{Z}_i])\right|\right|_2^2\right]$$

$$= \frac{1}{k^2}VAR\left[\sum_{i=1}^{k}\boldsymbol{Z}_i\right] = \frac{1}{k^2}\sum_{i=1}^{k}VAR[\boldsymbol{Z}_i] = \frac{1}{k}VAR[\boldsymbol{Z}]$$

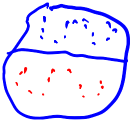


$$\stackrel{(b)}{=} \frac{1}{2k}E\left[||\boldsymbol{Z} - \boldsymbol{Z}_1]||_2^2\right] \stackrel{(c)}{\leq} \frac{1}{2k}\mathrm{diam}(\mathcal{T})^2 \stackrel{(d)}{\leq} \frac{1}{2k},$$

where (a) is due to (3), (b) is due to (2), and c is due to the fact that the diffrence between any two points cannot be larger than $\mathrm{diam}(\mathcal{T})$, and (d) is due to $\mathrm{diam}(\mathcal{T}) \leq 1$

Example of Overcoming CD - Computational Geometry

Now , since



$$E\left[\left\|\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{Z}_i - \boldsymbol{x}\right\|_2^2\right] \leq \frac{1}{2k} \tag{4}$$

then $\exists \boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_k \in \mathcal{T}$ (read there exist k realizations of $\boldsymbol{Z}$) such that

$$\left\|\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{z}_i - \boldsymbol{x}\right\|_2^2 \leq \frac{1}{2k} \tag{5}$$

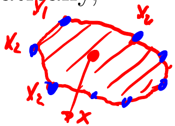
from which

$$\left\|\boldsymbol{x} - \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{z}_i\right\|_2 \leq \frac{1}{\sqrt{2k}} \tag{6}$$

Example of ACT

- We age given n images represented as vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$, where $\boldsymbol{x}_i \in \mathbb{R}^d$, $\forall i$. Let x_{ji} be the j-th element of $\boldsymbol{x}_i$, where x_{ji} represents the j-th pixel of the image $\boldsymbol{x}_i$.
- We would like to create a new image $\boldsymbol{x}$ (of some desired shapes), by mixing some or all of the given images $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$, i.e., by convex combination of the given images. Mathematically,

$$\boldsymbol{x} = \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i,$$



where $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$.

- We must assume that the desired image $\boldsymbol{x}$ can indeed the formed as a convex combination of $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$. More precisely, $\boldsymbol{x}$ belongs to the convex hull created from $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$, i.e., $\boldsymbol{x} \in \text{conv}(\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\})$. Otherwise, the desired image cannot be created.

Example of ACT

- What the Caratheodori Theorem tells us, is that we need to mix at most $d+1$ images to create the desire image $\boldsymbol{x}$. If the image is HD, thereby contains, $d = 1440 \times 1080 = 1,555,200$ pixels , we need to mix $1,555,200+1$ images to obtain the desire image.
- What the Approximate Caratheodori Theorem tells us, is that we need to mix at most k images to create the desire image $\boldsymbol{x}$. If we mix $k = 100$ images, then the absolute mean error between the original and the approximated image will be

$\frac{1}{\sqrt{2\cdot 100}}$ $1/(2*10) = 0.05 = 5\%$, which might be a very close approximation that the eye cannot notice.

- Note: The Approximate Caratheodori Theorem shows only existence of the points that make the approximations, but do not shows us how to obtain these points, other by brute-force search. Hence, the search for such an algorithm is ongoing.