VC Dimension and Configurations in $\mathbb{F}_q^d$

Nathanael Grand, Mandar Juvekar, Maxwell Sun

US NSF HDR TRIPODS 1934962

August 13, 2021

Introduction

- The consideration of the VC dimension of certain hypothesis classes leads to interesting configuration problems in $\mathbb{F}_q^d$.

Introduction

- The consideration of the VC dimension of certain hypothesis classes leads to interesting configuration problems in $\mathbb{F}_q^d$.

- The VC dimension is a fundamental concept in learning theory.

VC Dimension

- Let X be a set, and $Y = \{0, 1\}$ be a set of labels

VC Dimension

- Let X be a set, and $Y = \{0, 1\}$ be a set of labels

- A hypothesis class $\mathcal{H}$ is a set of maps $h : X \to Y$.

VC Dimension

- Let X be a set, and $Y = \{0, 1\}$ be a set of labels
- A hypothesis class $\mathcal{H}$ is a set of maps $h : X \to Y$.
- For $C \subset X$, we say that $\mathcal{H}$ shatters C if for all subsets C_0 of C, there exists $h \in \mathcal{H}$ that is 1 on all points in C_0 and 0 on all points in $C \setminus C_0$.

VC Dimension (Continued)

Definition

The VC Dimension of a hypothesis class $\mathcal{H}$, denoted by VCdim($\mathcal{H}$), is the size of the largest possible subset $C \subset X$ which is shattered by $\mathcal{H}$ where no sets of size larger than VCdim($\mathcal{H}$) are shattered by $\mathcal{H}$.

- This idea was developed by Vapnik and Chervonenkis in the 1970.

Finite Fields and $\mathbb{F}_q^d$

- For a prime q we let $\mathbb{F}_q$ denote the finite field with q elements.

Finite Fields and $\mathbb{F}_q^d$

- For a prime q we let $\mathbb{F}_q$ denote the finite field with q elements.
- We let $\mathbb{F}_q^d$ denote the vector space of d-tuples of elements of $\mathbb{F}_q$.

Finite Fields and $\mathbb{F}_q^d$

- For a prime q we let $\mathbb{F}_q$ denote the finite field with q elements.
- We let $\mathbb{F}_q^d$ denote the vector space of d-tuples of elements of $\mathbb{F}_q$.
- For our purposes, q is extremely large with respect to d.

Configuration Problems in $\mathbb{F}_q^d$

- The study of the VC dimension of certain hypothesis classes in an input space contained in $\mathbb{F}_q^d$ leads to the study of interesting point configurations in that space.

Configuration Problems in $\mathbb{F}_q^d$

- The study of the VC dimension of certain hypothesis classes in an input space contained in $\mathbb{F}_q^d$ leads to the study of interesting point configurations in that space.

- We define a "norm" in $\mathbb{F}_q^d$ for $y = (y_1, y_2, \ldots, y_d) \in \mathbb{F}_q^d$ as

$$\|y\| = y_1^2 + y_2^2 + \cdots + y_d^2$$

Configuration Problems in $\mathbb{F}_q^d$

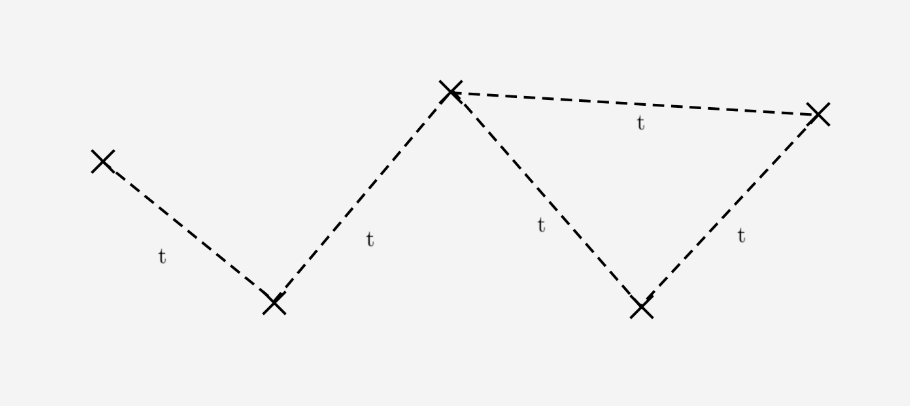
- The study of the VC dimension of certain hypothesis classes in an input space contained in $\mathbb{F}_q^d$ leads to the study of interesting point configurations in that space.

- We define a "norm" in $\mathbb{F}_q^d$ for $y = (y_1, y_2, \ldots, y_d) \in \mathbb{F}_q^d$ as

$$\|y\| = y_1^2 + y_2^2 + \cdots + y_d^2$$

- We define a configuration in $\mathbb{F}_q^d$ to be a sequence of points $x_1, \ldots, x_k \in \mathbb{F}_q^d$ where the distances between points x_i, x_j with $i \neq j$ are specified to be t for some pairs x_i, x_j and not for others, with $t \in \mathbb{F}_q$.

Example of a Configuration in $\mathbb{F}_q^d$



Hypothesis Class of Spheres in $\mathbb{F}_q^d$

- $\mathcal{H}_t^d = \{h_y : y \in \mathbb{F}_q^d\}$, where

$$h_y(x) = \begin{cases} 1 & \textit{if} \quad \|y - x\| = t \\ 0 & \textit{if} \quad \|y - x\| \neq t \end{cases}$$

Hypothesis Class of Spheres in $\mathbb{F}_q^d$

- $\mathcal{H}_t^d = \{h_y : y \in \mathbb{F}_q^d\}$, where

$$h_y(x) = \begin{cases} 1 & \textit{if} \quad \|y - x\| = t \\ 0 & \textit{if} \quad \|y - x\| \neq t \end{cases}$$

- Define $\mathcal{H}_t^d(E)$ for a subset E of $\mathbb{F}_q^d$ to be the set of all predictors h_y, $y \in E$.

Hypothesis Class of Spheres in $\mathbb{F}_q^d$

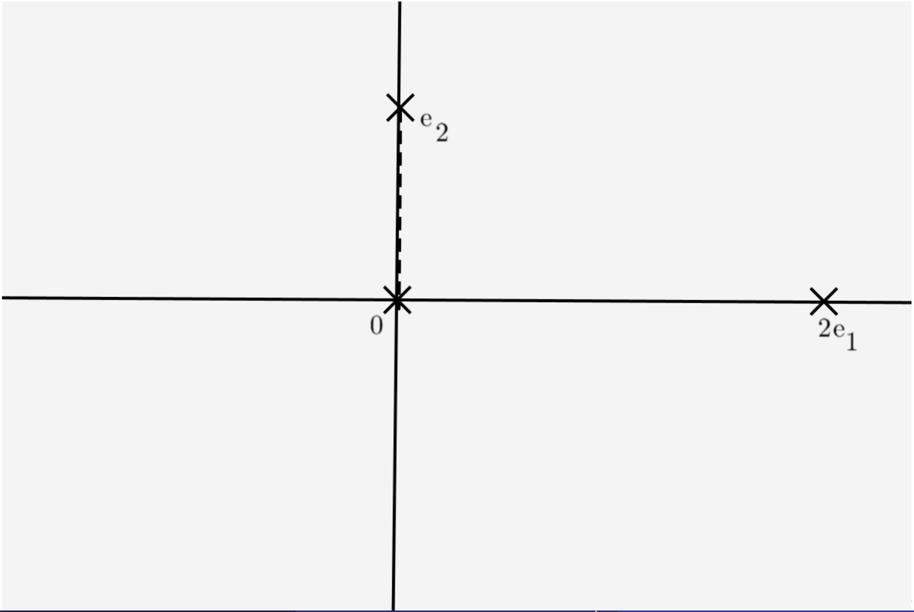
- $\mathcal{H}_t^d = \{h_y : y \in \mathbb{F}_q^d\}$, where

$$h_y(x) = \begin{cases} 1 & \text{if} \quad \|y - x\| = t \\ 0 & \text{if} \quad \|y - x\| \neq t \end{cases}$$

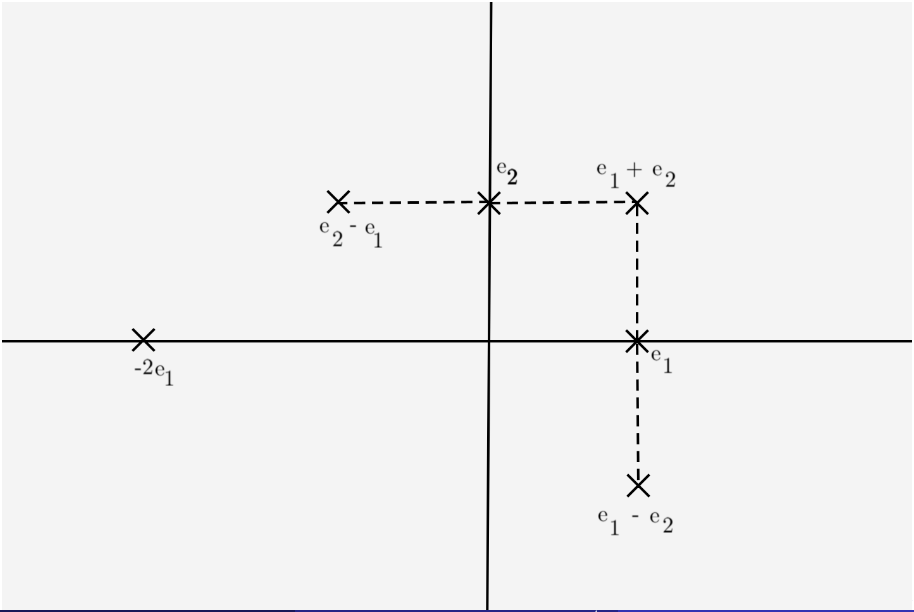
- Define $\mathcal{H}_t^d(E)$ for a subset E of $\mathbb{F}_q^d$ to be the set of all predictors h_y, $y \in E$.

- If $E = \mathbb{F}_q^d$, we will see later that $\text{VCdim}(\mathcal{H}_t^d(E)) = d + 1$.

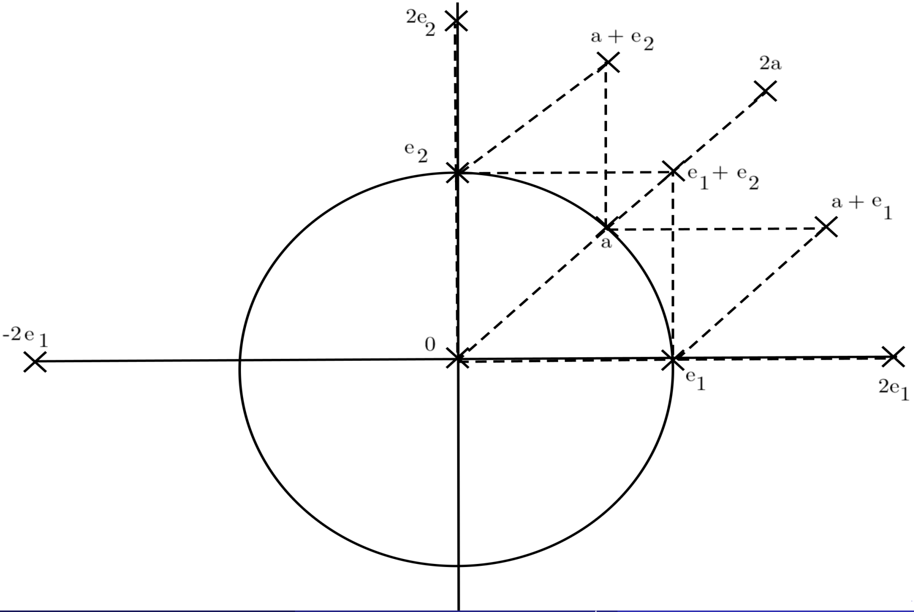
Shattering of 1 point in $\mathbb{F}_q^2$



Shattering of 2 points in $\mathbb{F}_q^2$



Shattering of 3 points in $\mathbb{F}_q^2$



VC Dimension of a Hypothesis Class of Spheres

- We conjecture that that there exists $\alpha < d$ such that if $|E| > q^{\alpha}$, then

$$\text{VCdim}(\mathcal{H}_t^d(E)) = d + 1$$

for all $t \neq 0$.

VC Dimension of a Hypothesis Class of Spheres

- We conjecture that that there exists $\alpha < d$ such that if $|E| > q^{\alpha}$, then

$$\text{VCdim}(\mathcal{H}_t^d(E)) = d + 1$$

  for all $t \neq 0$.

- In other words, the sample complexity of E is the same as $\mathbb{F}_q^d$ for $|E|$ sufficiently large.

PAC Learnability, and the FTSL

- The class of functions $\mathcal{H}_t^d(E)$ corresponds to a specific learning task.

PAC Learnability, and the FTSL

- The class of functions $\mathcal{H}_t^d(E)$ corresponds to a specific learning task.
- Suppose that $S_t(p)$ denotes the sphere of radius t, centered at a point $p \in E$:
$$S_t(p) = \{x \in E : \|x - p\| = t\}$$

PAC Learnability, and the FTSL

- The class of functions $\mathcal{H}_t^d(E)$ corresponds to a specific learning task.
- Suppose that $S_t(p)$ denotes the sphere of radius t, centered at a point $p \in E$:
$$S_t(p) = \{x \in E : \|x - p\| = t\}$$
- Question: Based on a i.i.d sample of size m, sampled through a distribution $\mathcal{D}$, is there an algorithm capable of successfully determining which $S_t(p)$ corresponds to the predictor with the least error?

PAC Learnability, and the FTSL (cont.)

- If $\text{VCdim}(\mathcal{H}_t^d(E)) = d + 1$, then by the Fundamental Theorem of Statistical Learning, $\mathcal{H}_t^d(E)$ is agnostic PAC learnable.

PAC Learnability, and the FTSL (cont.)

- If $\text{VCdim}(\mathcal{H}_t^d(E)) = d + 1$, then by the Fundamental Theorem of Statistical Learning, $\mathcal{H}_t^d(E)$ is agnostic PAC learnable.

- That is, running a learning algorithm on a i.i.d sample of size $m \geq m_{\mathcal{H}_t^d(E)}$ from a distribution $\mathcal{D}$ will produce a hypothesis h such that

$$\textit{Error of } h \leq \epsilon + \textit{minimum error}$$

with probability $1 - \delta$.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$

- We will prove the following

Theorem

$$VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d)) = VCdim(\mathcal{H}_t^d) = d + 1$$

and first assume $t = 1$.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$

- We will prove the following

Theorem

$$VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d)) = VCdim(\mathcal{H}_t^d) = d + 1$$

and first assume $t = 1$.

- Take e_j to mean the j-th standard basis vector.

Sketch of Proof: $\text{VCdim}(\mathcal{H}_t^d) = d + 1$

- We will prove the following

Theorem

$$VCdim(\mathcal{H}_t^d(\mathbb{F}_q^d)) = VCdim(\mathcal{H}_t^d) = d + 1$$

and first assume $t = 1$.

- Take e_j to mean the j-th standard basis vector.

- We take the set of $d + 1$ points $T_a = C \cup \{a\}$ which all lie on the unit sphere S_1, where

$$C = \{e_1, \ldots, e_d\}$$

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- There are on the order of q^{d-1} points in S_1, so we select $a \in S_1$ such that $a \neq e_j$. Our goal is to show that we can select a which results in $C \cup \{a\}$ shattering.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- There are on the order of q^{d-1} points in S_1, so we select $a \in S_1$ such that $a \neq e_j$. Our goal is to show that we can select a which results in $C \cup \{a\}$ shattering.

- First, consider some $C_0 \subset C$. Without loss of generality, we let

$$C_0 = \{e_1, \dots, e_i\}, \quad 1 \leq i \leq d$$

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- Let y have $y_j = \frac{2}{i}$ if $1 \leq j \leq i$, and 0 otherwise. For our set C_0, $y = (\frac{2}{i}, \frac{2}{i}, \ldots, \frac{2}{i}, 0, 0, \ldots, 0)$

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- Let y have $y_j = \frac{2}{i}$ if $1 \leq j \leq i$, and 0 otherwise. For our set C_0, $y = (\frac{2}{i}, \frac{2}{i}, \ldots, \frac{2}{i}, 0, 0, \ldots, 0)$
- For $e_j \in C_0$:

$$\|e_j - y\| = 1$$

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- Let y have $y_j = \frac{2}{i}$ if $1 \leq j \leq i$, and 0 otherwise. For our set C_0, $y = (\frac{2}{i}, \frac{2}{i}, \ldots, \frac{2}{i}, 0, 0, \ldots, 0)$

- For $e_j \in C_0$:

$$\|e_j - y\| = 1$$

- Whereas, for $e_k \notin C_0$:

$$\|e_k - y\| = \frac{4}{i} + 1 \neq 1$$

Sketch of Proof: VCdim$(\mathcal{H}_t^d) = d + 1$ (cont.)

- We now want to be able to find $a \in S_1$ such that $\|a - y\| \neq 1$.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- We now want to be able to find $a \in S_1$ such that $\|a - y\| \neq 1$.

- Consider for now the points $a \in S_1$ satisfying $\|a - y\| = 1$.

Sketch of Proof: $\text{VCdim}(\mathcal{H}_t^d) = d + 1$ (cont.)

- We now want to be able to find $a \in S_1$ such that $\|a - y\| \neq 1$.
- Consider for now the points $a \in S_1$ satisfying $\|a - y\| = 1$.
- We will show that that the set of such a is small compared to the size of S_1

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- We can simplify

$$\|a - y\| = \sum_{j=1}^{i} \left(a_j - \frac{2}{i} \right)^2 + \sum_{j=i+1}^{d} a_j^2 = 1 + \frac{4}{i} \left(1 - \sum_{j=1}^{i} a_j \right)$$

Sketch of Proof: $\text{VCdim}(\mathcal{H}_t^d) = d + 1$ (cont.)

- We can simplify

$$\|a - y\| = \sum_{j=1}^{i}\left(a_j - \frac{2}{i}\right)^2 + \sum_{j=i+1}^{d} a_j^2 = 1 + \frac{4}{i}\left(1 - \sum_{j=1}^{i} a_j\right)$$

- This distance is 1 if and only if

$$\sum_{j=1}^{i} a_j = 1$$

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d+1$ (cont.)

- This polynomial surface can be shown to have a intersection with the relation $\|a\| = 1$ that has cardinality on the order q^{d-2}.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- This polynomial surface can be shown to have a intersection with the relation $\|a\| = 1$ that has cardinality on the order q^{d-2}.

- The $i = 0$ case remains. For this, take $y = 3e_1$. A similar argument ensues where we show $\|y - e_j\| \neq 1$ and $\|y - a\| = 1$ for $O(q^{d-2})$ possible a.

Sketch of Proof: $\text{VCdim}(\mathcal{H}_t^d) = d + 1$ (cont.)

- Now, we deal with subsets including a. We wish to show there is a unit sphere containing $C_0 \cup \{a\}$ but has no other points in C as elements.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- Now, we deal with subsets including a. We wish to show there is a unit sphere containing $C_0 \cup \{a\}$ but has no other points in C as elements.

- We take h_y with y such that

$$y_j = \frac{2a_{i+1}^2}{\left(1 - \sum_{j=1}^{i} a_j\right)^2 + ia_{i+1}^2} \quad \text{and} \quad y_{i+1} = \frac{2a_{i+1}\left(1 - \sum_{j=1}^{i} a_j\right)}{\left(1 - \sum_{j=1}^{i} a_j\right)^2 + ia_{i+1}^2}$$

for $1 \leq j \leq i$ and the rest of the components are 0.

- By some algebra, it is again not hard to show that $\|y - e_j\| = 1$ for all $1 \leq j \leq i$, $\|y - a\| = 1$, and $\|y - e_j\| \neq 1$ for $j > i$ except for $O(q^{d-2})$ values of a.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- By some algebra, it is again not hard to show that $\|y - e_j\| = 1$ for all $1 \leq j \leq i$, $\|y - a\| = 1$, and $\|y - e_j\| \neq 1$ for $j > i$ except for $O(q^{d-2})$ values of a.

- Note: the above is only for $i < d$ but for $i = d$ we just take the origin. If $C_0 = \emptyset$, we take $y = 2a$, which satisfies $\|2a - e_j\| \neq 1$ if we exclude $O(q^{d-2})$ values of a.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d+1$ (cont.)

- By some algebra, it is again not hard to show that $\|y - e_j\| = 1$ for all $1 \leq j \leq i$, $\|y-a\| = 1$, and $\|y - e_j\| \neq 1$ for $j > i$ except for $O(q^{d-2})$ values of a.
- Note: the above is only for $i < d$ but for $i = d$ we just take the origin. If $C_0 = \emptyset$, we take $y = 2a$, which satisfies $\|2a - e_j\| \neq 1$ if we exclude $O(q^{d-2})$ values of a.
- Since every subset of C has a corresponding predictor except for a total of $O(q^{d-2})$ a, the VC dimension of $\mathcal{H}_1^d$ is at least $d+1$. We now show it is less than $d+2$.

Sketch of Proof: VCdim$(\mathcal{H}_t^d) = d + 1$ (cont.)

- Take an arbitrary set of $d + 2$ points in $\mathbb{F}_q^d$.

Sketch of Proof: $\text{VCdim}(\mathcal{H}_t^d) = d + 1$ (cont.)

- Take an arbitrary set of $d + 2$ points in $\mathbb{F}_q^d$.

- If they are in general position, a subset D of $d + 1$ of these points determine a sphere. So, the last point is either on this sphere or not. It follows that there does not exist a predictor that is 1 on D and 1 on the last point and another predictor that is 1 on D and 0 on the last point.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- Take an arbitrary set of $d + 2$ points in $\mathbb{F}_q^d$.
- If they are in general position, a subset D of $d + 1$ of these points determine a sphere. So, the last point is either on this sphere or not. It follows that there does not exist a predictor that is 1 on D and 1 on the last point and another predictor that is 1 on D and 0 on the last point.
- If there is no such D in general position, this is 'worse' in a sense. A more nuanced argument that is similar works, however.

Sketch of Proof: VCdim($\mathcal{H}_t^d$) $= d + 1$ (cont.)

- To show the result for general t that are squares, we can use a scaling argument.

Sketch of Proof: VCdim$(\mathcal{H}_t^d) = d + 1$ (cont.)

- To show the result for general t that are squares, we can use a scaling argument.

- For nonsquare t, we only need to show the result for one such t and then scale.

Sketch of Proof: $\text{VCdim}(\mathcal{H}_t^d) = d + 1$ (cont.)

- To show the result for general t that are squares, we can use a scaling argument.

- For nonsquare t, we only need to show the result for one such t and then scale.

- To do this, we take $t = s^2 + d - 1$ and then use x_j in place of e_j where x_j is s in the jth place and 1 everywhere else. The rest of the proof follows similarly.

The Next Frontier

- In the future, we wish to show that the VC dimension of $\mathcal{H}_t^d(E)$ is $d+1$ for all E that are sufficiently large.

The Next Frontier

- In the future, we wish to show that the VC dimension of $\mathcal{H}_t^d(E)$ is $d+1$ for all E that are sufficiently large.

- We want a lower bound for $|E|$ that is small compared to q^d.

The Next Frontier

- In the future, we wish to show that the VC dimension of $\mathcal{H}_t^d(E)$ is $d+1$ for all E that are sufficiently large.

- We want a lower bound for $|E|$ that is small compared to q^d.

- This leads to questions about the existence of certain configurations in such subsets E.

References

M. Bennett, J. Chapman, D. Covert, D. Hart, A. Iosevich and J. Pakianathan, *Long paths in the distance graph over large subsets of vector spaces over finite fields*, J. Korean Math. Soc. **53**, (2016).

S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, (2014).

A. Iosevich and M. Rudnev, *Erdős distance problem in vector spaces over finite fields*, Trans. Amer. Math. Soc. **359** (2007), no. 12, 6127-6142.

References

D. Covert, *The Finite Field Distance Problem*, MAA Press, an imprint of the American Mathematical Society, (2021).

A. Iosevich and H. Parshall, *Embedding distance graphs in finite field vector spaces*, (2018).

A. Iosevich, G. Jardine and B. McDonald, *Cycles of arbitrary length in distance graphs on* $\mathbb{F}_q^d$, (2021).

Thanks!

- We wish to thank our project supervisors for their help throughout the research. We look forward to continue working with them!

Thanks!

- We wish to thank our project supervisors for their help throughout the research. We look forward to continue working with them!
- Thanks to NSF for providing our groups with funding!