Master's Thesis Seminar

VC dimension of bisectors between curves

Carolin Kaffine

3rd April 2020

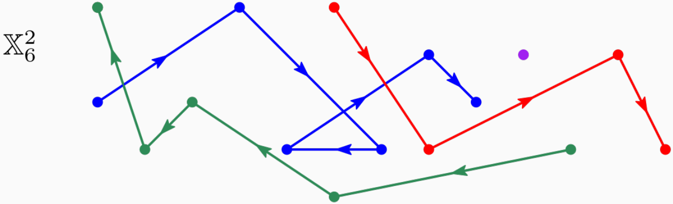
Overview

- Recap of basic definitions
- Main goals
- Upper bounds:
  - Approach 1: via composition lemma for halfspaces
  - Approach 2: via VC dimension of function spaces
- Lower bounds
- Summary of results
- Outlook

Basic definitions

Definition:

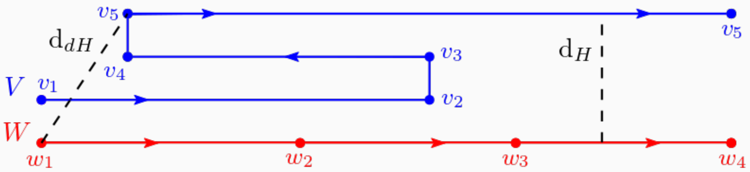
- **curve in** $\mathbb{R}^d$: continuous function $V\colon [0,1] \to \mathbb{R}^d$
- **polygonal curve**: piecewise linear
- $\mathbb{X}_k^d$: space of all piecewise linear curves in $\mathbb{R}^d$ with k vertices



Hausdorff and discrete Hausdorff distance:

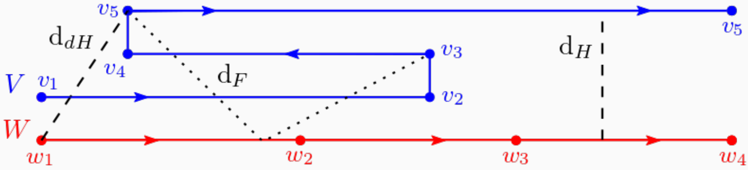
$$\mathrm{d}_H(V,W) := \max\left\{ \sup_{p\in V} \inf_{q\in W} \mathrm{d}(p,q), \sup_{q\in W} \inf_{p\in V} \mathrm{d}(p,q) \right\}$$

$$\mathrm{d}_{dH}(V,W) := \max\left\{ \max_{v\in V} \min_{w\in W} \mathrm{d}(v,w), \max_{w\in W} \min_{v\in V} \mathrm{d}(v,w) \right\}$$



Fréchet distance:

$$d_F(V, W) = \inf_{f,\,g} \max_{\alpha \in [0,1]} \left\| V(f(\alpha)) - W(g(\alpha)) \right\|,$$



Basic definitions

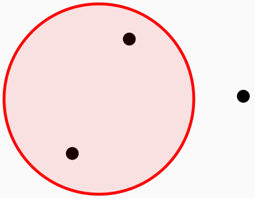
- **Range space** $(X, \mathcal{R})$: ground set X, ranges $R \in \mathcal{R} \subseteq 2^X$
- given $Y \subseteq X$, it is **shattered by** $\mathcal{R}$ if

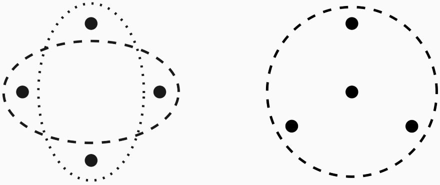
$$\{R \cap Y \mid R \in \mathcal{R}\} = 2^Y$$

- **VC dimension**: greatest cardinality of shattered subset

Examples for VC dimension

Ground set $X = \mathbb{R}^2$, ranges are disks:





$\Rightarrow$ VCdim ≥ 3 $\Rightarrow$ VCdim < 4

In general: balls and halfspaces in $\mathbb{R}^d$ have VCdim $= d + 1$.

Basic definitions

Shatter function (or growth function) for a range space $(X, \mathcal{R})$:

$$\pi_{(X,\mathcal{R})}(m) = \max_{Y \subseteq X, |Y| = m} \left| \{ R \cap Y \mid R \in \mathcal{R} \} \right|$$

Shatter function lemma (or Sauer's lemma)

For a range space $(X, \mathcal{R})$ with VC dimension at most δ, we have

$$\pi_{(X,\mathcal{R})}(m) \leq \Phi_\delta(m) := \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{\delta}$$

$\Rightarrow$ polynomial growth in m since $\Phi_\delta(m) \leq \left(\frac{em}{\delta}\right)^\delta \in \mathcal{O}(m^\delta)$

Bisectors between curves

Bisector range space: $(\mathbb{X}_m^d, \mathcal{B}_{\mathsf{d},k})$

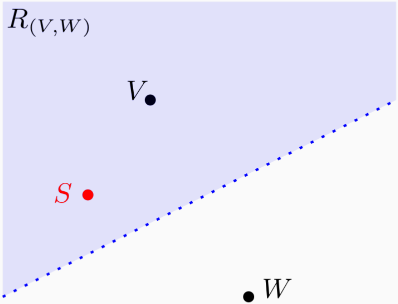
- ground set $\mathbb{X}_m^d$ = set of all curves in $\mathbb{R}^d$ with m vertices
- range set $\mathcal{B}_{\mathsf{d},k}$ with ranges

$$R_{(V,W)} = \{S \in \mathbb{X}_m^d \mid \mathsf{d}(V,S) \leq \mathsf{d}(W,S)\}$$

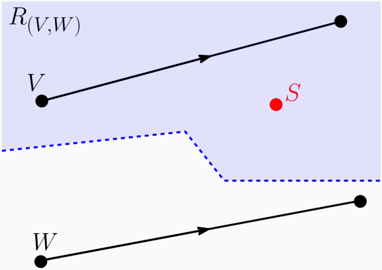
for $(V,W) \in \mathbb{X}_k^d \times \mathbb{X}_k^d$.

Bisectors in $\mathbb{R}^2$ for $m = 1$ (i.e. all distance functions are the same):

$\left(\mathbb{X}_1^2, \mathcal{B}_{\mathsf{d},1}\right)$

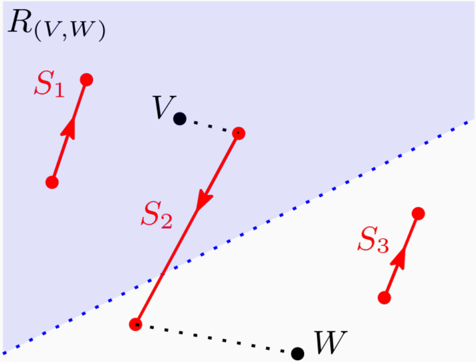


$\left(\mathbb{X}_1^2, \mathcal{B}_{\mathsf{d},2}\right)$



Bisectors between curves

For $m > 1$: no graphic representation of ranges anymore, because dimension gets higher than 3



Goal of Master's thesis

Main goal:

Find upper and lower bounds on VC dimension of the bisector range space, dependent on

- m (complexity of shattered curves)
- k (complexity of curves that define bisectors)
- d (dimension)
- d_{dH}, d_H, d_{dF}, or d_F (used distance function)

Main paper:

"The VC Dimension of Metric Balls under Fréchet and Hausdorff Distances" by A. Driemel, A. Nusser, J.M. Phillips, and I. Psarros

Upper bound 1: via composition lemma

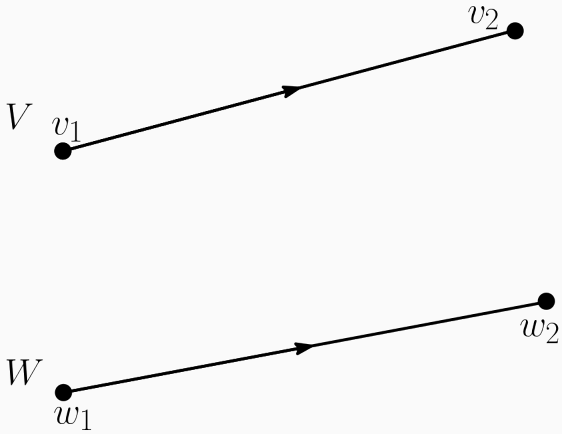
Composition lemma (simplified):

For a range space $(X, \mathcal{R})$ with VCdim $= \delta$, the range space of all unions/intersections of n ranges in $\mathcal{R}$ has VC dimension $\mathcal{O}(n\delta \log n)$.

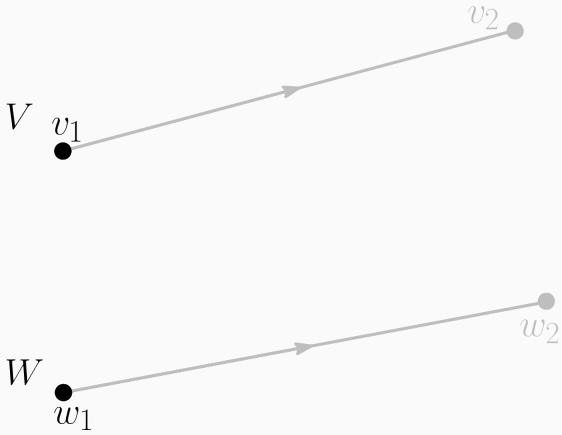
Idea (for $m = 1$):

- Write bisector ranges as unions and intersections of halfspaces
- By the composition lemma, we can bound VC dimension of bisector range space by considering VC dimension of halfspaces (which is $d + 1$)

Upper bound 1: via composition lemma

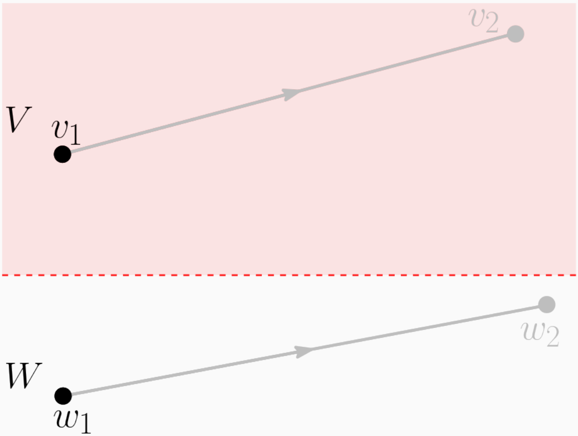


Upper bound 1: via composition lemma



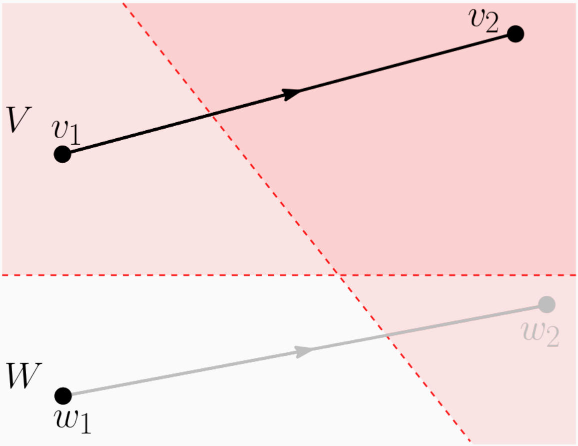
Upper bound 1: via composition lemma

Step 1: For two points: bisector range $R_{(v_1,w_1)}$ is halfspace



Upper bound 1: via composition lemma

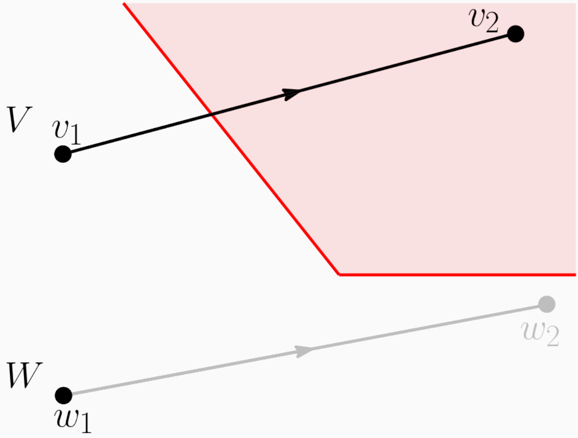
Step 1: For two points: bisector range $R_{(v_1,w_1)}$ is halfspace



Upper bound 1: via composition lemma

Step 1: For two points: bisector range $R_{(v_1,w_1)}$ is halfspace

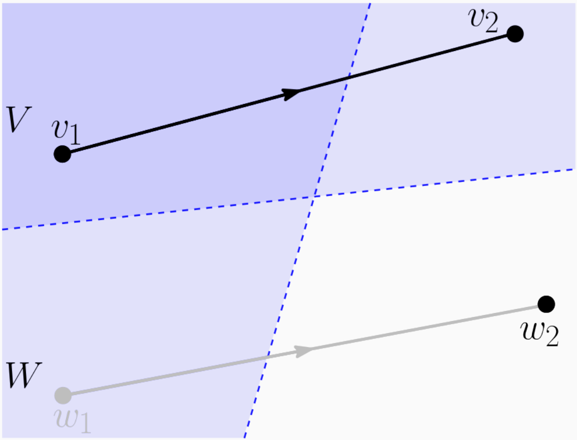
Step 2: Ranges get intersected when adding a point to one curve



Upper bound 1: via composition lemma

Step 1: For two points: bisector range $R_{(v_1,w_1)}$ is halfspace

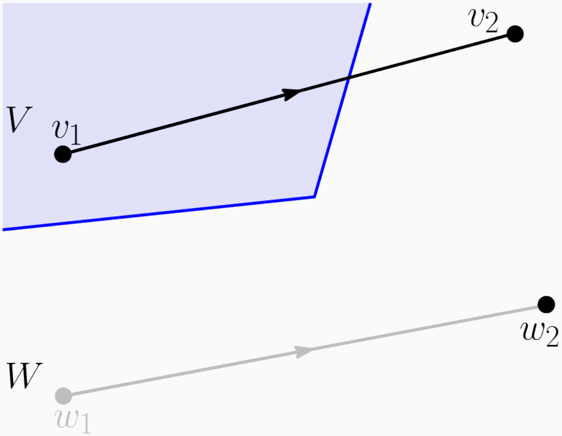
Step 2: Ranges get intersected when adding a point to one curve



Upper bound 1: via composition lemma

Step 1: For two points: bisector range $R_{(v_1,w_1)}$ is halfspace

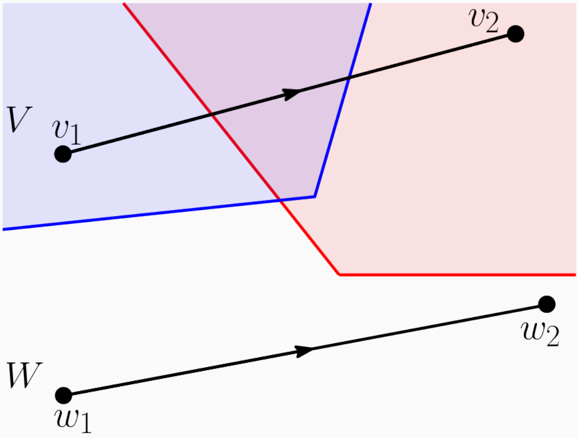
Step 2: Ranges get intersected when adding a point to one curve



Upper bound 1: via composition lemma

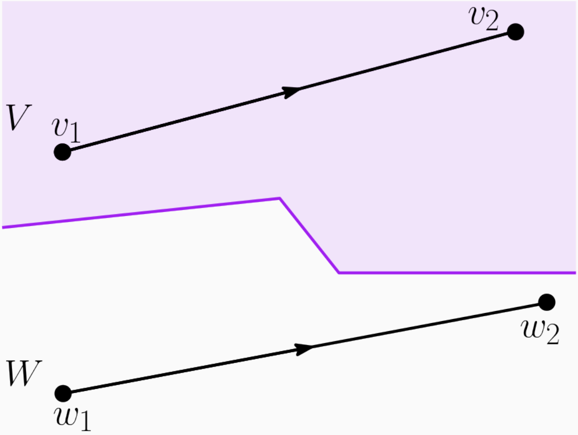
Step 1: For two points: bisector range $R_{(v_1,w_1)}$ is halfspace

Step 2: Ranges get intersected when adding a point to one curve



Upper bound 1: via composition lemma

Step 1: For two points: bisector range $R_{(v_1,w_1)}$ is halfspace
Step 2: Ranges get intersected when adding a point to one curve
Step 3: Final range of two curves is union of intersections

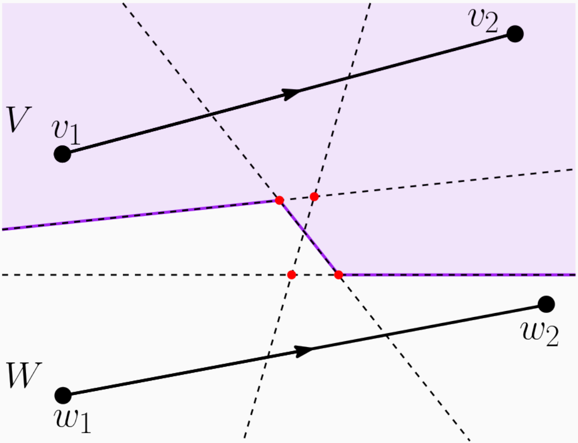


Upper bound 1: via composition lemma

Step 1: For two points: bisector range $R_{(v_1,w_1)}$ is halfspace

Step 2: Ranges get intersected when adding a point to one curve

Step 3: Final range of two curves is union of intersections



Upper bound 1: via composition lemma

For general d:

- range $R_{(V,W)} = \bigcup_{w \in W} \bigcap_{v \in V} h(v, w)$, where $h(v, w)$ is halfspace of points that are closer to v than to w

- if V, W have length k, we took $(k-1)^2 \in \mathcal{O}(k^2)$ unions and intersections

  $\Rightarrow$ VC dimension is in $\mathcal{O}\big((k-1)^2(d+1)\log((k-1)^2)\big) = \mathcal{O}(k^2 d \log k)$

Upper bound 2: via Thm on VC dimension of function spaces

Theorem:

Let $h\colon \mathbb{R}^a \times \mathbb{R}^b \to \{0,1\}$ and

$$H = \{x \mapsto h(\alpha, x) \,|\, \alpha \in \mathbb{R}^a\}.$$

Suppose h can be computed by an algorithm that takes $(\alpha, x) \in \mathbb{R}^a \times \mathbb{R}^b$ as input and returns $h(\alpha, x)$ after no more than t simple operations.
Then, the VC dimension of H ist $\leq 4a(t+2)$.

Upper bound 2: via Thm on VC dimension of function spaces

- Write bisector range $R_{(V,W)}$ as function $h\big((V,W),\cdot\big)$ that takes a curve S and outputs 1 if S is closer to V than to W, and 0 else

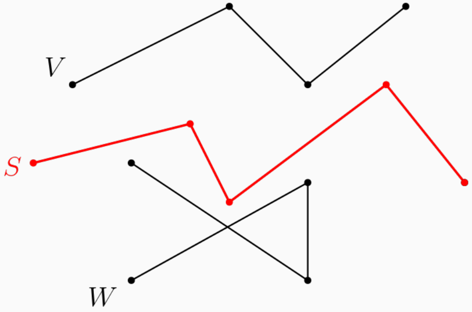
  $\Rightarrow$ Bisector range space can be written as $(\mathbb{X}^d_m, H)$ for

  $$H = \left\{S \mapsto h\big((V,W),S\big) \,\middle|\, S \in \mathbb{X}^d_m, (V,W) \in \mathbb{X}^d_k \times \mathbb{X}^d_k\right\}$$

- so $h\colon \underbrace{(\mathbb{X}^d_k \times \mathbb{X}^d_k)}_{\cong \mathbb{R}^{2dk}} \times \mathbb{X}^d_m \to \{0,1\}$, i.e. $\boxed{a = 2dk}$
- it remains to compute t, i.e. check how fast h can be computed

Upper bound 2: via Thm on VC dimension of function spaces

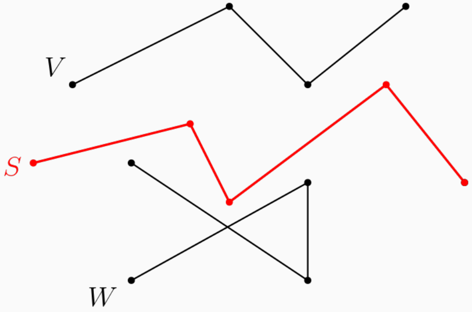
Example for discrete Hausdorff distance:



Upper bound 2: via Thm on VC dimension of function spaces

Example for discrete Hausdorff distance:

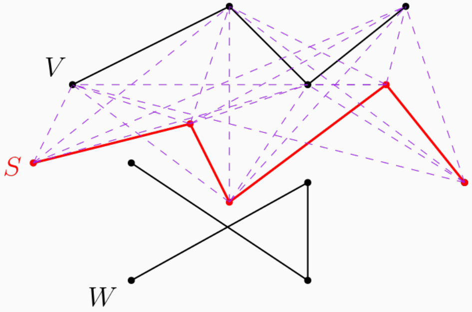
Step 1: Calculate $\big(\mathrm{d}(v,s)\big)^2$ for all $v \in V$, $s \in S$



Upper bound 2: via Thm on VC dimension of function spaces

Example for discrete Hausdorff distance:

Step 1: Calculate $\big(\mathrm{d}(v,s)\big)^2$ for all $v \in V$, $s \in S$

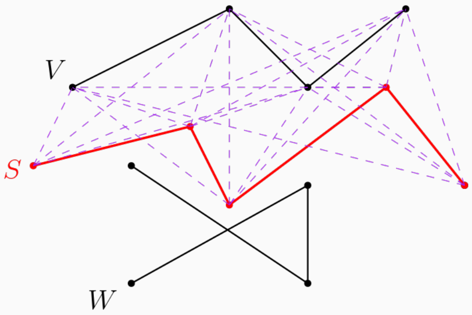


Upper bound 2: via Thm on VC dimension of function spaces

Example for discrete Hausdorff distance:

Step 1: Calculate $\big(\mathrm{d}(v,s)\big)^2$ for all $v \in V$, $s \in S$

Step 2: Find $\mathrm{d}_{dH}(V,S) = \max\big(\max_{s\in S}\min_{v\in V}\mathrm{d}(v,s), \max_{v\in V}\min_{s\in S}\mathrm{d}(v,s)\big)$

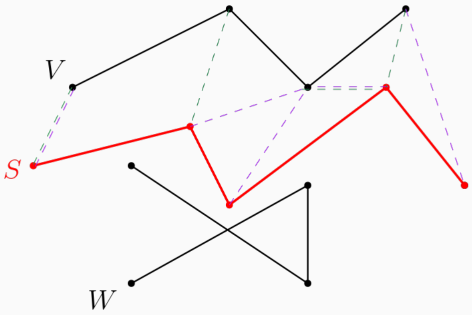


Upper bound 2: via Thm on VC dimension of function spaces

Example for discrete Hausdorff distance:

Step 1: Calculate $\big(\mathrm{d}(v,s)\big)^2$ for all $v \in V$, $s \in S$

Step 2: Find $\mathrm{d}_{dH}(V,S) = \max\big(\max_{s\in S}\min_{v\in V}\mathrm{d}(v,s), \max_{v\in V}\min_{s\in S}\mathrm{d}(v,s)\big)$

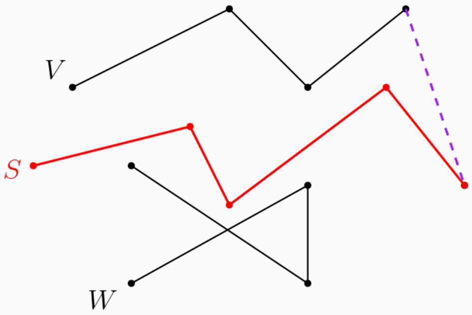


Upper bound 2: via Thm on VC dimension of function spaces

Example for discrete Hausdorff distance:

Step 1: Calculate $\big(\mathrm{d}(v, s)\big)^2$ for all $v \in V$, $s \in S$

Step 2: Find $\mathrm{d}_{dH}(V, S) = \max\big(\max_{s \in S} \min_{v \in V} \mathrm{d}(v, s), \max_{v \in V} \min_{s \in S} \mathrm{d}(v, s)\big)$



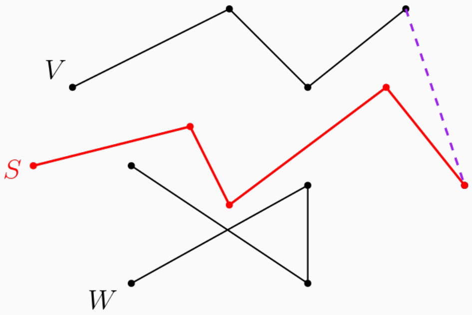
Upper bound 2: via Thm on VC dimension of function spaces

Example for discrete Hausdorff distance:

Step 1: Calculate $\big(\mathrm{d}(v,s)\big)^2$ for all $v \in V$, $s \in S$

Step 2: Find $\mathrm{d}_{dH}(V,S) = \max\big(\max_{s\in S}\min_{v\in V}\mathrm{d}(v,s), \max_{v\in V}\min_{s\in S}\mathrm{d}(v,s)\big)$

Step 3: Do same for W and take minimum of $\mathrm{d}_{dH}(V,S)$ and $\mathrm{d}_{dH}(W,S)$



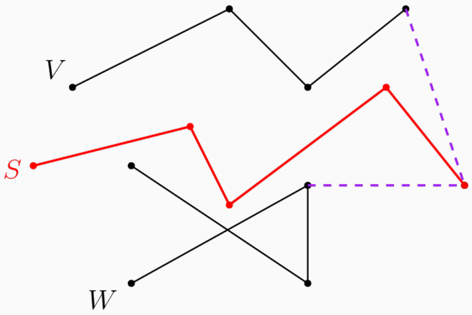
Upper bound 2: via Thm on VC dimension of function spaces

Example for discrete Hausdorff distance:

Step 1: Calculate $\big(\mathrm{d}(v,s)\big)^2$ for all $v \in V$, $s \in S$

Step 2: Find $\mathrm{d}_{dH}(V,S) = \max\big(\max_{s\in S}\min_{v\in V}\mathrm{d}(v,s), \max_{v\in V}\min_{s\in S}\mathrm{d}(v,s)\big)$

Step 3: Do same for W and take minimum of $\mathrm{d}_{dH}(V,S)$ and $\mathrm{d}_{dH}(W,S)$



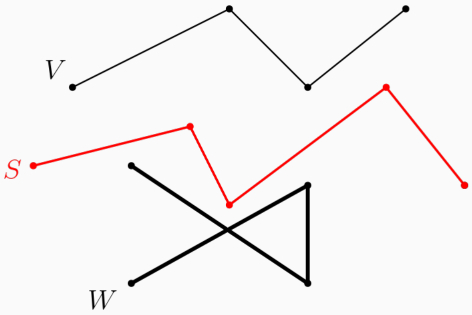
Upper bound 2: via Thm on VC dimension of function spaces

Example for discrete Hausdorff distance:

Step 1: Calculate $\big(d(v,s)\big)^2$ for all $v \in V$, $s \in S$

Step 2: Find $d_{dH}(V,S) = \max\big(\max_{s\in S}\min_{v\in V} d(v,s), \max_{v\in V}\min_{s\in S} d(v,s)\big)$

Step 3: Do same for W and take minimum of $d_{dH}(V,S)$ and $d_{dH}(W,S)$



Upper bound 2: via Thm on VC dimension of function spaces

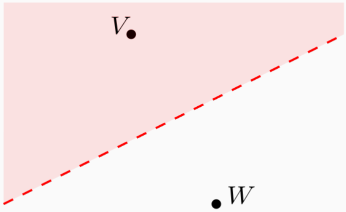
- In total: calculation of $2mk$ squared euclidean distances between vertices, each in $\mathcal{O}(d)$
- $\mathcal{O}(mk)$ comparisons to find $\mathrm{d}_{dH}(V, S)$
- All in all: $\boxed{t \in \mathcal{O}(mkd)}$ simple operations

  $\Rightarrow$ VCdim $\leq 4 \cdot 2dk(c \cdot mkd - 1) \in \mathcal{O}(mk^2d^2)$

Lower bounds

Idea: Find lower bounds for $k = 1$ and/or $m = 1$
$\Rightarrow$ valid lower bound for all distance functions and all k and m

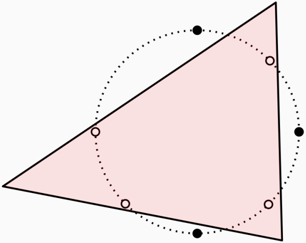
Easy lower bound (for $m = k = 1$:)



Bisector ranges look like halfspaces, so VCdim $\geq d + 1$

Lower bound (for $m = 1$):

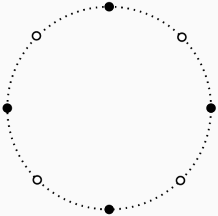
- VC dimension of (open) k-gons is $2k + 1$
- Bisector ranges can look like open k-gons, so their VC dimension is $\geq 2k + 1$



Lower bounds

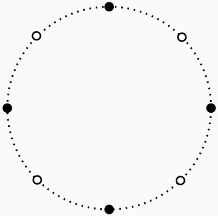
Lower bound (for $m = 1$):

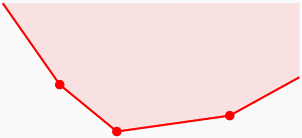
- VC dimension of (open) k-gons is $2k + 1$
- Bisector ranges can look like open k-gons, so their VC dimension is $\geq 2k + 1$



Lower bound (for $m = 1$):

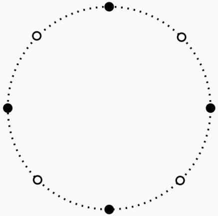
- VC dimension of (open) k-gons is $2k + 1$
- Bisector ranges can look like open k-gons, so their VC dimension is $\geq 2k + 1$

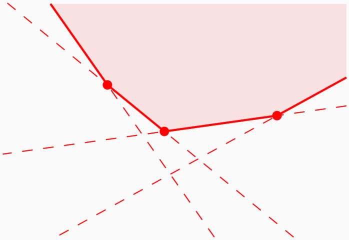




Lower bound (for $m = 1$):

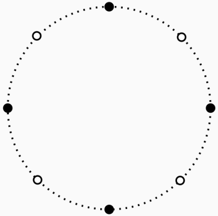
- VC dimension of (open) k-gons is $2k + 1$
- Bisector ranges can look like open k-gons, so their VC dimension is $\geq 2k + 1$

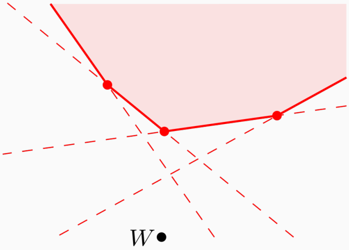




Lower bound (for $m = 1$):

- VC dimension of (open) k-gons is $2k + 1$
- Bisector ranges can look like open k-gons, so their VC dimension is $\geq 2k + 1$

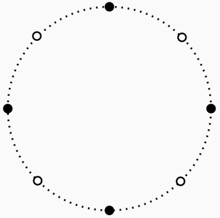


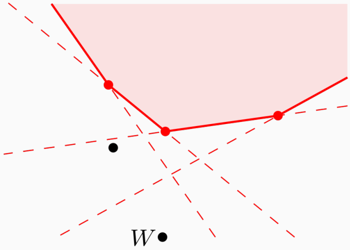


Lower bounds

Lower bound (for $m = 1$):

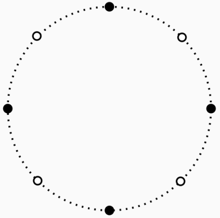
- VC dimension of (open) k-gons is $2k + 1$
- Bisector ranges can look like open k-gons, so their VC dimension is $\geq 2k + 1$

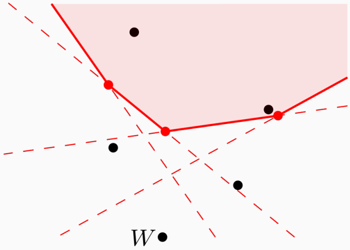




Lower bound (for $m = 1$):

- VC dimension of (open) k-gons is $2k + 1$
- Bisector ranges can look like open k-gons, so their VC dimension is $\geq 2k + 1$

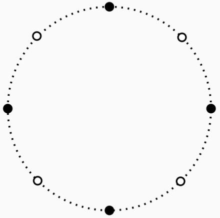


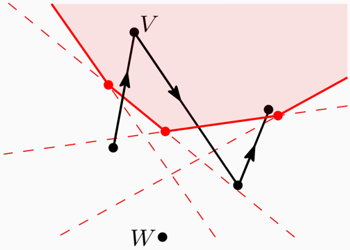


Lower bounds

Lower bound (for $m = 1$):

- VC dimension of (open) k-gons is $2k + 1$
- Bisector ranges can look like open k-gons, so their VC dimension is $\geq 2k + 1$





$\Rightarrow$ Combining the two lower bounds we get VCdim $\in \Omega(\max(k, d))$

Summary of results so far:

Upper bounds:

<table>
<tr><th>Distance function</th><th>m arbitrary</th><th>$m = 1$</th></tr>
<tr><td>discrete Hausdorff</td><td>$\mathcal{O}(mk^2d^2)$</td><td rowspan="2">$\mathcal{O}(dk^2 \log k)$</td></tr>
<tr><td>Hausdorff</td><td rowspan="3">–</td></tr>
<tr><td>discrete Fréchet</td><td rowspan="2">$\mathcal{O}(k^2d^2)$</td></tr>
<tr><td>Fréchet</td></tr>
</table>

Lower bound:

$\Omega(\max(k, d))$

Outlook

Further goals:

- establish upper bounds for other distance functions than the discrete Hausdorff distance that depend on m
- establish better lower bounds by using geometric properties of bisector range spaces
- reduce gap between upper and lower bounds