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 \( V : [0, 1] \rightarrow \mathbb{R}^d \)
- **polygonal curve**: piecewise linear
- \( \mathbb{X}^d_k \): space of all piecewise linear curves in \( \mathbb{R}^d \) with \( k \) vertices
Basic definitions

Hausdorff and discrete Hausdorff distance:

\[ d_H(V, W) := \max \left\{ \sup_{p \in V} \inf_{q \in W} d(p, q), \sup_{q \in W} \inf_{p \in V} d(p, q) \right\} \]

\[ d_{dH}(V, W) := \max \left\{ \max_{v \in V} \min_{w \in W} d(v, w), \max_{w \in W} \min_{v \in V} d(v, w) \right\} \]
Fréchet distance:

\[ d_F(V, W) = \inf_{f, g} \max_{\alpha \in [0,1]} \| V(f(\alpha)) - W(g(\alpha)) \|, \]
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 $\geq 3$

$\Rightarrow$ 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 \(\delta\), 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: \((\mathbb{X}_m^d, \mathcal{B}_{d,k})\)

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

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

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

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

$$(X_1^2, \mathcal{B}_{d,1}) \quad \text{and} \quad (X_1^2, \mathcal{B}_{d,2})$$
For $m > 1$: no graphic representation of ranges anymore, because dimension gets higher than 3
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 \(O(n\delta \log n)\).

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

Step 1: For two points, the bisector range $R(v_1, w_1)$ is a halfspace.

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

Step 3: The final range of two curves is the 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

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
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 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

$\mathcal{O}(((k - 1)^2(d + 1) \log((k - 1)^2))) = \mathcal{O}(k^2 d \log k)$
Theorem:
Let \( h: \mathbb{R}^a \times \mathbb{R}^b \rightarrow \{0, 1\} \) and
\[
H = \{ x \mapsto h(\alpha, x) \mid \alpha \in \mathbb{R}^a \}.
\]

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 \( H \) is \( \leq 4a(t + 2) \).
Upper bound 2: via Thm on VC dimension of function spaces

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

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

H = \{ S \mapsto h((V,W), S) \mid S \in X^d_m, (V, W) \in X^d_k \times X^d_k \}
\]

• so \( h: (X^d_k \times X^d_k) \times X^d_m \to \{0, 1\} \), i.e. \( a = 2dk \)

• 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:

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

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

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

**Step 1**: Calculate \((d(v, s))^2\) 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 $(d(v, s))^2$ for all $v \in V$, $s \in S$
Example for discrete Hausdorff distance:

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

**Step 2:** Find $d_{dH}(V, S) = \max \left( \max_{s \in S} \min_{v \in V} d(v, s), \max_{v \in V} \min_{s \in S} d(v, s) \right)$
Example for discrete Hausdorff distance:

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

Step 2: Find \(d_{dH}(V, S) = \max \left( \max_{s \in S} \min_{v \in V} d(v, s), \max_{v \in V} \min_{s \in S} d(v, s) \right)\)
Example for discrete Hausdorff distance:

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

Step 2: Find $d_{dH}(V, S) = \max \left( \max_{s \in S} \min_{v \in V} d(v, s), \max_{v \in V} \min_{s \in S} d(v, s) \right)$
Upper bound 2: via Thm on VC dimension of function spaces

**Example** for discrete Hausdorff distance:

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

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

**Step 3:** Do same for \(W\) and take minimum of \(d_{dH}(V, S)\) and \(d_{dH}(W, S)\)
Example for discrete Hausdorff distance:

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

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

Step 3: Do same for \(W\) and take minimum of \(d_{dH}(V, S)\) and \(d_{dH}(W, S)\)
Example for discrete Hausdorff distance:

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

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

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 $d_{dH}(V, S)$

- All in all: $t \in \mathcal{O}(mkd)$ simple operations

$\Rightarrow$ $\text{VCdim} \leq 4 \cdot 2dk(c \cdot mkd - 1) \in \mathcal{O}(mk^2d^2)$
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 $\text{VCdim} \geq d + 1$
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 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 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 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 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 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 \((\text{for } m = 1)\):

- 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 $\text{VCdim} \in \Omega(\max(k, d))$
Summary of results so far:

Upper bounds:

<table>
<thead>
<tr>
<th>Distance function</th>
<th>$m$ arbitrary</th>
<th>$m = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>discrete Hausdorff</td>
<td>$O(mk^2d^2)$</td>
<td>$O(dk^2 \log k)$</td>
</tr>
<tr>
<td>Hausdorff</td>
<td>$-$</td>
<td>$O(dk^2 \log k)$</td>
</tr>
<tr>
<td>discrete Fréchet</td>
<td>$-$</td>
<td>$O(k^2d^2)$</td>
</tr>
<tr>
<td>Fréchet</td>
<td>$-$</td>
<td>$O(k^2d^2)$</td>
</tr>
</tbody>
</table>

Lower bound:

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