Simplified and Improved Bounds on the VC-Dimension for Elastic Distance Measures

Frederik Brüning Anne Driemel

University of Bonn

Ioannina, March 13, 2024

VC-Dimension of a Range Space

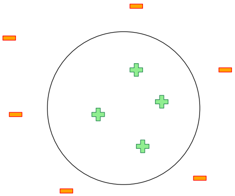
Range space $\mathcal{R}$ with ground set X:
Set $\mathcal{R}$ s.t. any $r \in \mathcal{R}$ is of the form $r \subseteq X$

Example:
Balls with ground set $X = \mathbb{R}^2$:

$$\mathcal{R} = \{b(c, \Delta) \mid \Delta \in \mathbb{R}_+, c \in \mathbb{R}^2\}$$

where

$$b(c, \Delta) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq \Delta\}$$



VC-Dimension of a Range Space

Range space $\mathcal{R}$ with ground set X:
Set $\mathcal{R}$ s.t. any $r \in \mathcal{R}$ is of the form $r \subseteq X$

Shattering:
$A \subseteq X$ is shattered by $\mathcal{R} \iff$
$\forall A' \subseteq A \; \exists r \in \mathcal{R}$ s.t. $A' = r \cap A$
(all subsets of A can be realized by ranges in $\mathcal{R}$)

VC-dim($\mathcal{R}$):
Maximal size of a shattered subset $A \subseteq X$

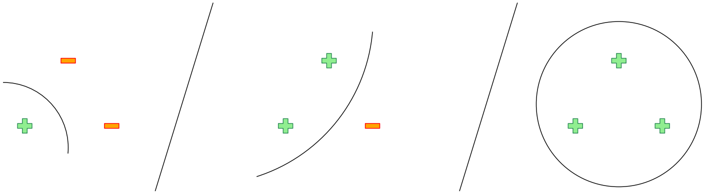
Example:
Balls with ground set $X = \mathbb{R}^2$:

$$\mathcal{R} = \{b(c, \Delta) \mid \Delta \in \mathbb{R}_+, c \in \mathbb{R}^2\}$$

where

$$b(c, \Delta) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq \Delta\}$$

VC-dim($\mathcal{R}$) $= 3$



Range space $\mathcal{R}$ with ground set X:
Set $\mathcal{R}$ s.t. any $r \in \mathcal{R}$ is of the form $r \subseteq X$

Shattering:
$A \subseteq X$ is shattered by $\mathcal{R}$ $\iff$
$\forall A' \subseteq A \, \exists r \in \mathcal{R}$ s.t. $A' = r \cap A$
(all subsets of A can be realized by ranges in $\mathcal{R}$)

VC-dim($\mathcal{R}$):
Maximal size of a shattered subset $A \subseteq X$

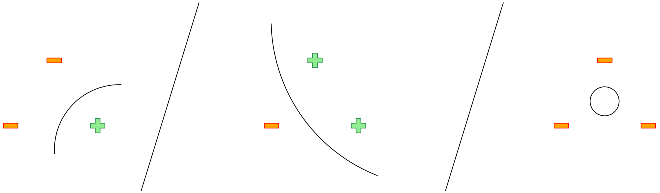
Example:
Balls with ground set $X = \mathbb{R}^2$:

$$\mathcal{R} = \{b(c, \Delta) \mid \Delta \in \mathbb{R}_+, c \in \mathbb{R}^2\}$$

where

$$b(c, \Delta) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq \Delta\}$$

VC-dim($\mathcal{R}$) $= 3$



Range space $\mathcal{R}$ with ground set X:
Set $\mathcal{R}$ s.t. any $r \in \mathcal{R}$ is of the form $r \subseteq X$

Shattering:
$A \subseteq X$ is shattered by $\mathcal{R} \iff$
$\forall A' \subseteq A \; \exists r \in \mathcal{R}$ s.t. $A' = r \cap A$
(all subsets of A can be realized by ranges in $\mathcal{R}$)

VC-dim($\mathcal{R}$):
Maximal size of a shattered subset $A \subseteq X$

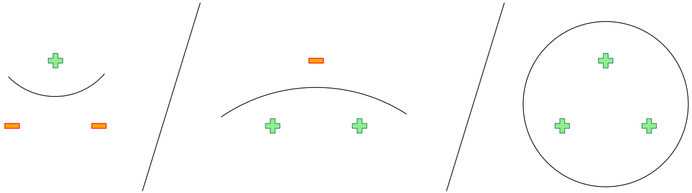
Example:
Balls with ground set $X = \mathbb{R}^2$:

$$\mathcal{R} = \{b(c, \Delta) \mid \Delta \in \mathbb{R}_+, c \in \mathbb{R}^2\}$$

where

$$b(c, \Delta) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq \Delta\}$$

VC-dim($\mathcal{R}$) $= 3$



VC-Dimension of a Range Space

Range space $\mathcal{R}$ with ground set X:
Set $\mathcal{R}$ s.t. any $r \in \mathcal{R}$ is of the form $r \subseteq X$

Shattering:
$A \subseteq X$ is shattered by $\mathcal{R}$ $\iff$
$\forall A' \subseteq A \; \exists r \in \mathcal{R}$ s.t. $A' = r \cap A$
(all subsets of A can be realized by ranges in $\mathcal{R}$)

VC-dim($\mathcal{R}$):
Maximal size of a shattered subset $A \subseteq X$

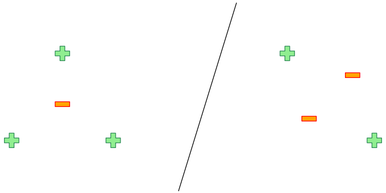
Example:
Balls with ground set $X = \mathbb{R}^2$:

$$\mathcal{R} = \{b(c, \Delta) \mid \Delta \in \mathbb{R}_+, c \in \mathbb{R}^2\}$$

where

$$b(c, \Delta) = \{x \in \mathbb{R}^2 \mid \|x - c\| \leq \Delta\}$$

VC-dim($\mathcal{R}$) $= 3$



Motivation: What are VC-Dimension bounds used for?

Sample bounds for computational tasks:

- ϵ-nets, relative-error (p, ϵ)-approximations
- test error of classification model

Applications:

- kernel density estimation
- coresets
- clustering
- object recognition
- ...

Range Spaces for Elastic Distance Measures

Range spaces:

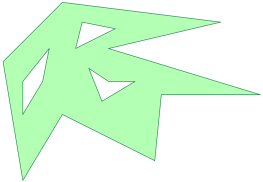
$$
\begin{aligned}
d_\rho &= \text{distance measure on } (\mathbb{R}^d)^m \\
b_\rho(c, \Delta) &= \{x \in (\mathbb{R}^d)^m \mid d_\rho(x, c) \leq \Delta\} \\
\mathcal{R}_{\rho,k} &= \{b_\rho(c, \Delta) \mid \Delta \in \mathbb{R}_+, c \in (\mathbb{R}^d)^k\}
\end{aligned}
$$

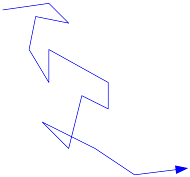
Distance measures:

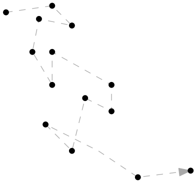
- Hausdorff
- (weak) Fréchet
- Dynamic Time Warping

Ground set $((\mathbb{R}^d)^m)$/Centers $((\mathbb{R}^d)^k)$:

- Polygonal regions
- continuous polygonal curves
- discrete polygonal curves



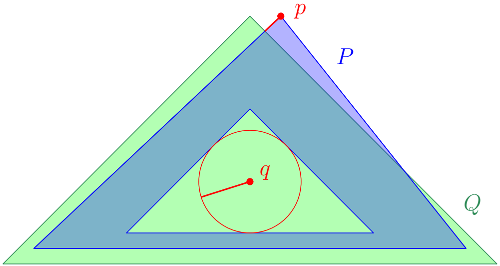




directed Hausdorff distance:
$d_{\overrightarrow{H}}(P, Q) = \sup_{p \in P} \inf_{q \in Q} \|p - q\|$

Hausdorff distance:
$d_H(P, Q) = \max\{d_{\overrightarrow{H}}(P, Q), d_{\overrightarrow{H}}(Q, P)\}$

for $P, Q \subseteq \mathbb{R}^d$



Results

		new	old
discrete polygonal curves	DTW	$O(dk^2 \log(m))$	-
		$O(dkm \log(k))$	
	Hausdorff	$O(dk \log(km))$	$O(dk \log(dkm))$
	Fréchet	$O(dk \log(km))^{(*)}$	
continuous polygonal curves	Hausdorff	$O(dk \log(km))$	$O(d^2 k^2 \log(dkm))$
	Fréchet	$O(dk \log(km))^{(*)}$	
	weak Fréchet	$O(dk \log(km))^{(*)}$	$O(d^2 k \log(dkm))$
polygons $\mathbb{R}^2$	**Hausdorff**	$O(k \log(km))$	-

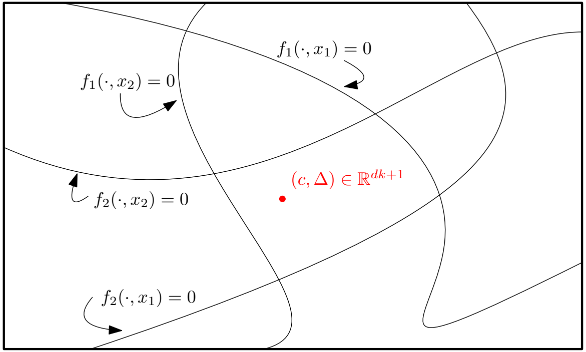
Table: Overview of VC-dimension bounds. Results marked with $^{(*)}$ were independently obtained by Cheng and Huang [2024]. The old results were obtained by Driemel, Nusser, Phillips and Psarros [2021].

Lower bound (Driemel, Nusser, Phillips and Psarros [2021]):
$\Omega(\max(dk \log(k), \log(dm)))$ for $d \geq 4$ for polygonal curves.

General Approach

Approach:

- Split query $d_H(P, Q)$ into predicates
- Express predicates as combinations of sign values of polynomials
- Bound VC-dim based on number of cells in arrangement of zero sets of polynomials.



Definition:

F: Class of polynomials of constant degree from $\mathbb{R}^{dk+1} \times \mathbb{R}^{dm}$ to $\mathbb{R}$

$\mathcal{R}$ is t-combination of $\mathrm{sgn}(F)$: $\exists$ boolean function g, $\exists f_1, \ldots, f_t \in F$ s.t. $\forall r \in \mathcal{R}$ $\exists y$ s.t.

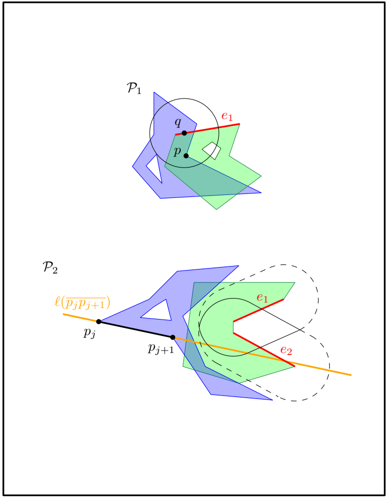
$$r = \{x \in X \mid g(\mathrm{sgn}(f_1(y, x)), \ldots, \mathrm{sgn}(f_t(y, x))) = 1\}$$

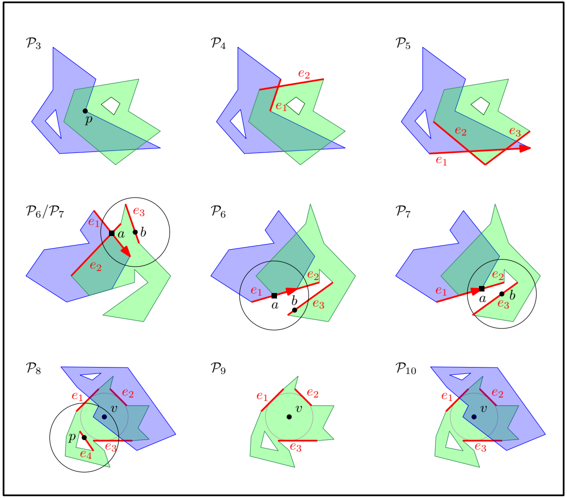
Theorem (Anthony and Bartlett 1999):

Suppose $\mathcal{R}$ is a t-combination of $\mathrm{sgn}(F)$. Then VC-dim($\mathcal{R}$) is in $O(dk \log(t))$.

Idea goes back to Goldberg and Jerrum [1993] and independently Ben-David and Lindenbaum [1993].

Predicates



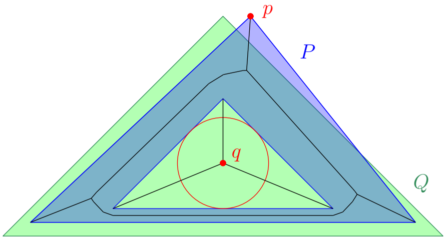


Predicates $\mathcal{P}_1$ and $\mathcal{P}_2$ are by Driemel, Nusser, Phillips and Psarros [2021]

Predicates

Cases:

$d_{\overrightarrow{H}}(P, Q)$ maximized at point p at the boundary of P
$d_{\overrightarrow{H}}(Q, P)$ maximized at point q in the interior of Q

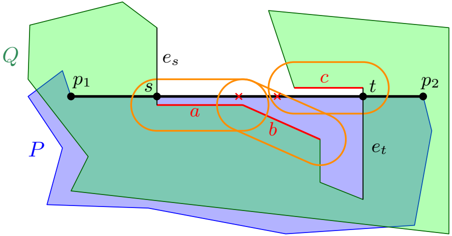


Predicates:

- $(\mathcal{B})$ (Boundary): True $\iff$ $d_{\overrightarrow{H}}(\partial P, Q) \leq \Delta$.
- $(\mathcal{I})$ (Interior): True if $d_{\overrightarrow{H}}(P, Q) \leq \Delta$. False if $d_{\overrightarrow{H}}(P, Q) > d_{\overrightarrow{H}}(\partial P, Q)$ and $d_{\overrightarrow{H}}(P, Q) > \Delta$.

$$d_{\overrightarrow{H}}(P, Q) \leq \Delta \iff (\mathcal{B}) \text{ and } (\mathcal{I}) \text{ true}$$

Predicate $(\mathcal{B})$



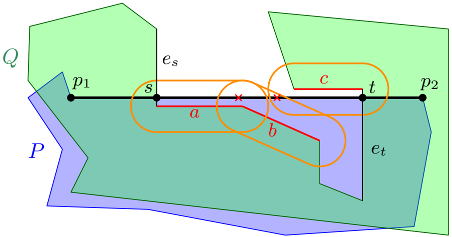
$(\mathcal{B})$ (Boundary): True $\iff d_{\vec{H}}(\partial P, Q) \leq \Delta$

$d_{\vec{H}}(\partial P, Q) \leq \Delta \iff d_{\vec{H}}(e, Q) \leq \Delta$ for every edge e of P

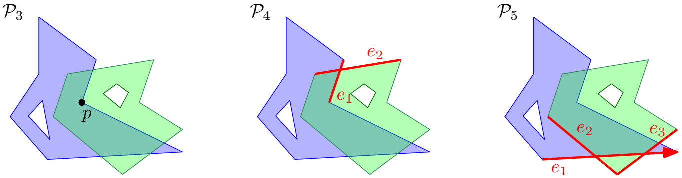
- Find for each e the part that is outside of Q (here $\overline{st}$)
- Find sequence of edges of Q such that these parts are included in their stadiums (here a, b, c)

Predicate $(\mathcal{B})$

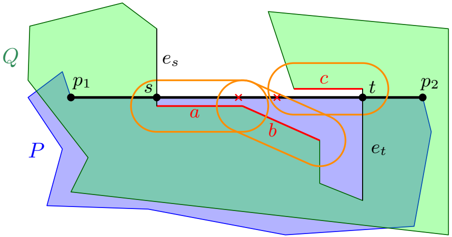
Find part of e that is outside of Q (here $\overline{st}$)



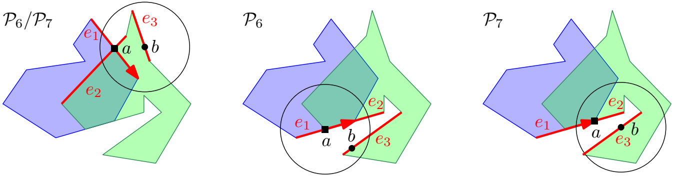
- $\mathcal{P}_3$: $p \in Q$?
- $\mathcal{P}_4$: $e_1 \cap e_2 \neq \emptyset$?
- $\mathcal{P}_5$: e_1 intersects e_2 before e_3?



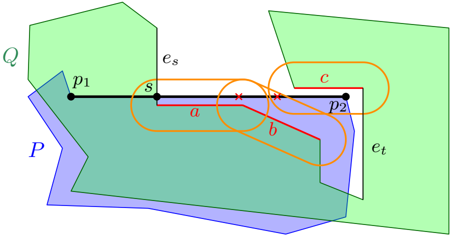
Find stadiums that include s and t (here a and c)



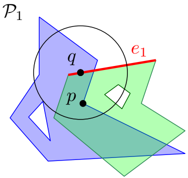
- $\mathcal{P}_6/P_7$: $\exists b$ on e_3 with $\|b-a\| \leq \Delta$?
- $\mathcal{P}_6$: a is first point on e_1 in $e_1 \cap e_2$
- $\mathcal{P}_7$: a is last point on e_1 in $e_1 \cap e_2$



Find stadiums that include s and p_2 (here a and c)

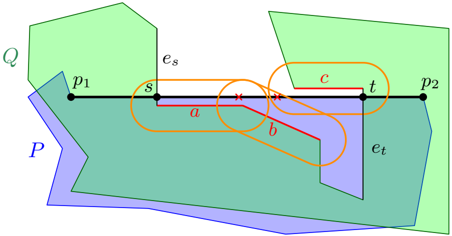


- $\mathcal{P}_1$: $\exists q$ on e_1 with $\|p - q\| \leq \Delta$?

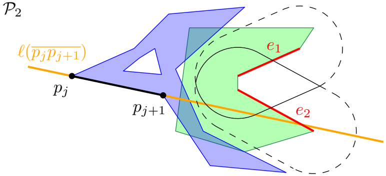


Predicate $(\mathcal{B})$

Find sequence of stadiums that include $\overline{st}$ (here a, b, c)

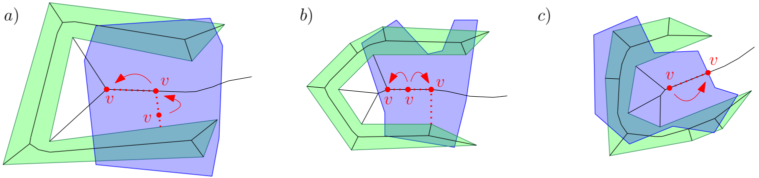


- $D_{\Delta,2}(e_1, e_2)$: intersection of stadiums around e_1, e_2
- $\mathcal{P}_2$: $\ell(\overline{p_j p_{j+1}}) \cap D_{\Delta,2}(e_1, e_2) \neq \emptyset$?



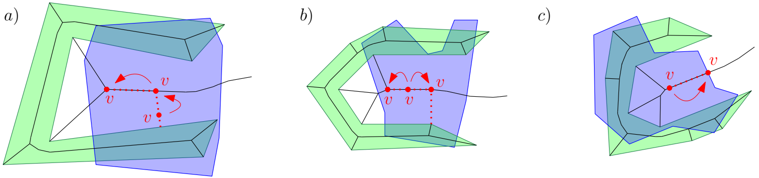
Predicate $(\mathcal{I})$

$(\mathcal{I})$ (Distance realized in interior): We only check vertices of Voronoi diagram of edges of Q.

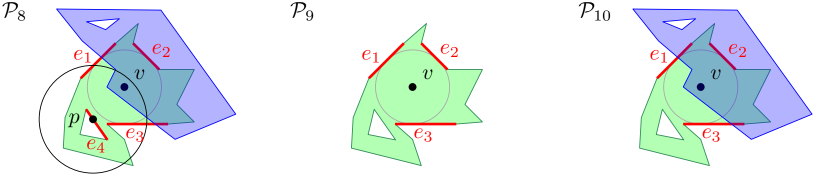


Predicate ($\mathcal{I}$)

($\mathcal{I}$) (Distance realized in interior): We only check vertices of Voronoi diagram of edges of Q.



We check distances of all Voronoi vertices to all edges of Q.
A Voronoi vertex is relevant if it is inside of P and outside of Q.



- $\mathcal{P}_8$: $\exists p$ on e_4 with $\|v - p\| \leq \Delta$?
- $\mathcal{P}_9$: $v \in Q$?
- $\mathcal{P}_{10}$: $v \in P$?

Remaining step

Lemma

For any two polygonal regions P and Q (that may contain holes), given the truth values of all predicates of the type $\mathcal{P}_1, \ldots, \mathcal{P}_{10}$ one can determine whether $d_{\overrightarrow{H}}(P, Q) \leq \Delta$.

Remaining step:

Express predicates $\mathcal{P}_1, \ldots, \mathcal{P}_{10}$ as combinations of sign values of polynomials.

Results

		new	old
discrete polygonal curves	DTW	$O(dk^2 \log(m))$	-
		$O(dkm \log(k))$	
	Hausdorff	$O(dk \log(km))$	$O(dk \log(dkm))$
	Fréchet	$O(dk \log(km))^{(*)}$	
continuous polygonal curves	Hausdorff	$O(dk \log(km))$	$O(d^2 k^2 \log(dkm))$
	Fréchet	$O(dk \log(km))^{(*)}$	
	weak Fréchet	$O(dk \log(km))^{(*)}$	$O(d^2 k \log(dkm))$
polygons $\mathbb{R}^2$	**Hausdorff**	$O(k \log(km))$	-

Table: Overview of VC-dimension bounds. Results marked with $^{(*)}$ were independently obtained by Cheng and Huang [2024]. The old results were obtained by Driemel, Nusser, Phillips and Psarros [2021]

Lower bound (Driemel, Nusser, Phillips and Psarros [2021]):
$\Omega(\max(dk \log(k), \log(dm)))$ for $d \geq 4$ for polygonal curves.