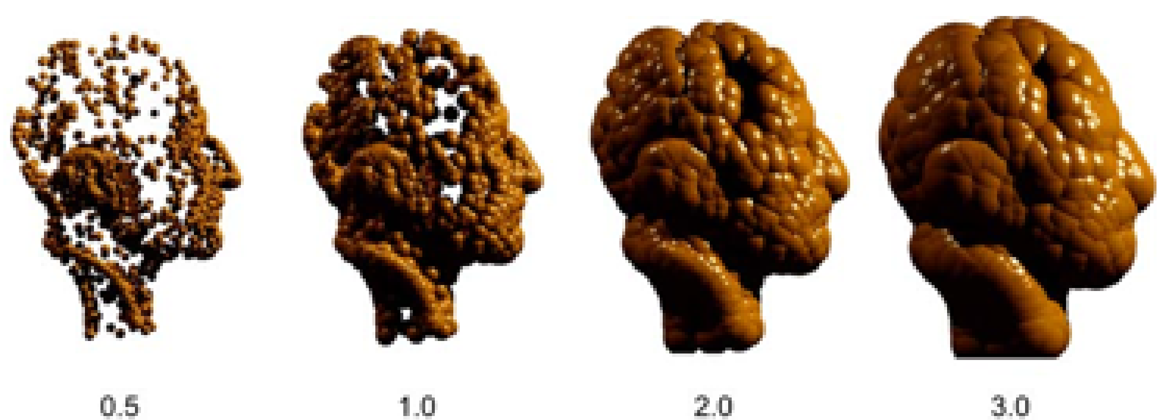
TRIPODS Summer Bootcamp: Topology and Machine Learning
Brown University, August 2018

A Gaussian Type Kernel for Persistence Diagrams

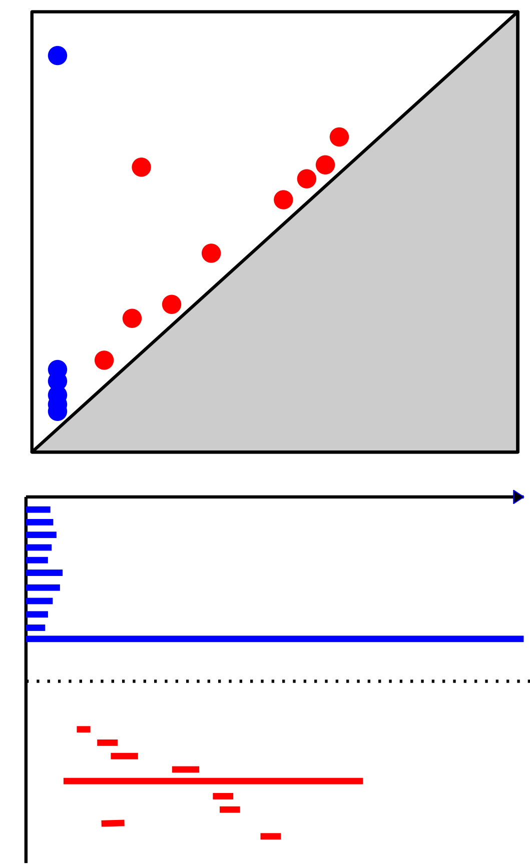
Mathieu Carrière — joint work with S. Oudot and M. Cuturi

Persistence diagrams

point cloud



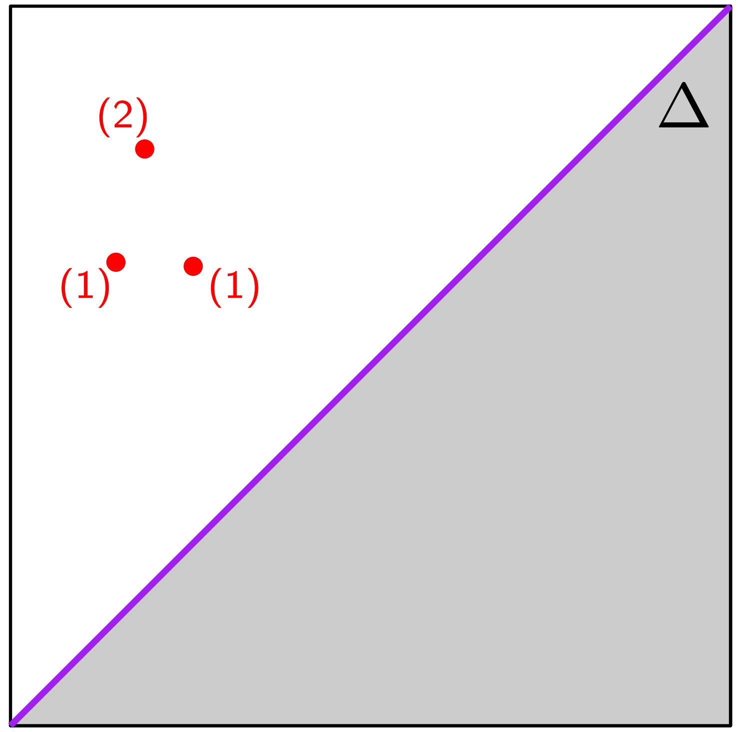
1-parameter filtration (by scale)



barcode / diagram

Persistence diagrams

Def: A finite multiset in the open half-plane $\Delta \times \mathbb{R}_+$



Persistence diagrams

Def: A finite multiset in the open half-plane $\Delta \times \mathbb{R}_+$

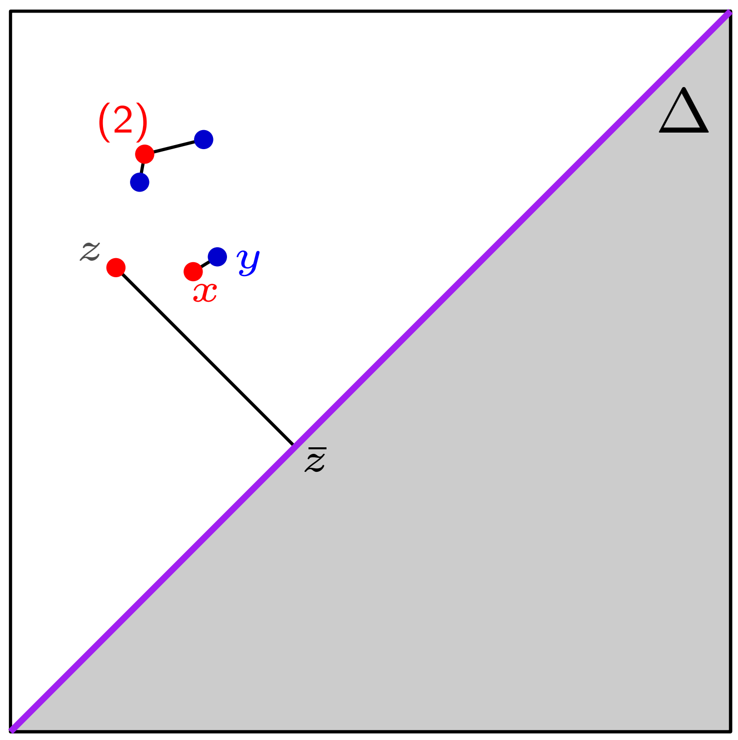
Given a partial matching $M : X \leftrightarrow Y$:

cost of a matched pair $(x, y) \in M$: $c_p(x, y) := \|x - y\|_\infty^p$

cost of an unmatched point $z \in X \sqcup Y$: $c_p(z) := \|z - \bar{z}\|_\infty^p$

cost of M:

$$c_p(M) := \left(\sum_{(x,\,y)\text{ matched}} c_p(x, y) \;+\; \sum_{z\text{ unmatched}} c_p(z) \right)^{1/p}$$



p-th diagram distance (extended metric):

$$\mathrm{d}_p(X, Y) := \inf_{M : X \leftrightarrow Y} c_p(M)$$

bottleneck distance:

$$\mathrm{d}_\infty(X, Y) := \lim_{p \to \infty} \mathrm{d}_p(X, Y)$$

Persistence diagrams

Def: A finite multiset in the open half-plane $\Delta \times \mathbb{R}_+$

Given a pa

cost of a r

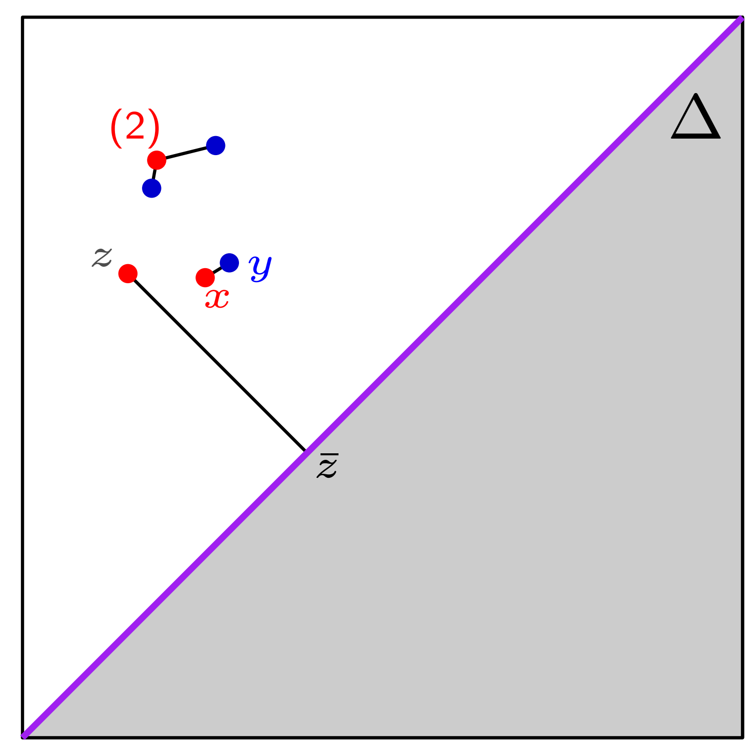
cost of an

Prop.: [Cohen-Steiner, Edelsbrunner, Harer 2007]

$$\mathrm{d}_p(D, D') \leq \left(2 \sum_{z \in X \sqcup Y} \|z - \bar{z}\|_\infty \right)^{1/p} \mathrm{d}_\infty(D, D')^{1-1/p}$$

cost of M:

$$c_p(M) := \left(\sum_{(x,\,y)\ \text{matched}} c_p(x, y) + \sum_{z\ \text{unmatched}} c_p(z) \right)^{1/p}$$



p-th diagram distance (extended metric):

$$\mathrm{d}_p(X, Y) := \inf_{M : X \leftrightarrow Y} c_p(M)$$

bottleneck distance:

$$\mathrm{d}_\infty(X, Y) := \lim_{p \to \infty} \mathrm{d}_p(X, Y)$$

Persistence diagrams

Def: A finite multiset in the open half-plane $\Delta \times \mathbb{R}_+$

Given a pa

Prop.: [Cohen-Steiner, Edelsbrunner, Harer 2007]

$$\mathrm{d}_p(D, D') \leq \left(2 \sum_{z \in X \sqcup Y} \|z - \bar{z}\|_\infty \right)^{1/p} \mathrm{d}_\infty(D, D')^{1-1/p}$$

cost of a

cost of an

cost of

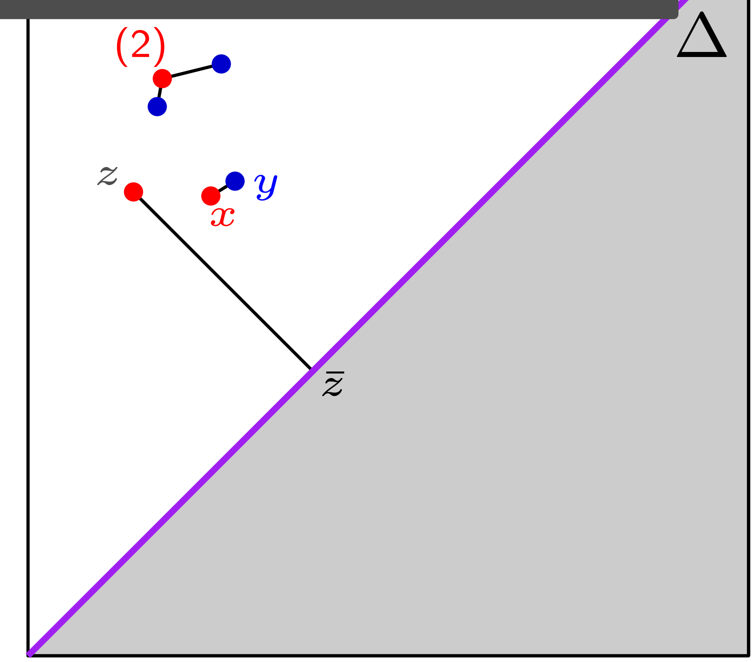
similar to Wasserstein distances between probability measures...

... but not quite the same (different masses, diagonal)

$c_p(M)$:

(x, y) matched

z unmatched



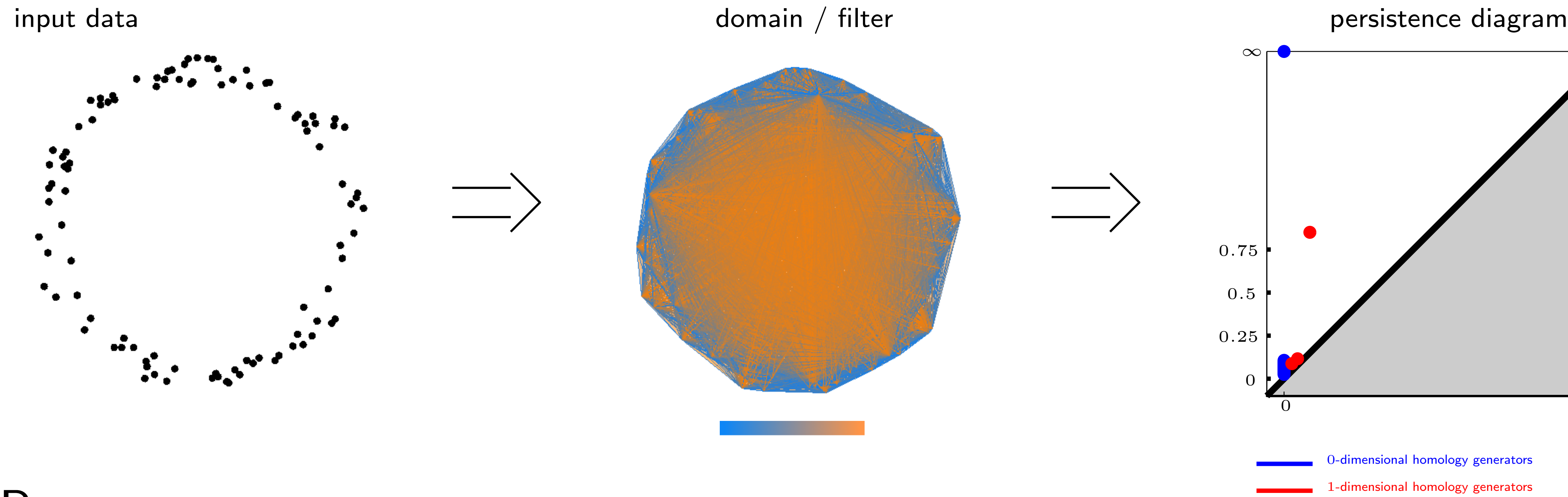
p-th diagram distance (extended metric):

$$\mathrm{d}_p(X, Y) := \inf_{M : X \leftrightarrow Y} c_p(M)$$

bottleneck distance:

$$\mathrm{d}_\infty(X, Y) := \lim_{p \to \infty} \mathrm{d}_p(X, Y)$$

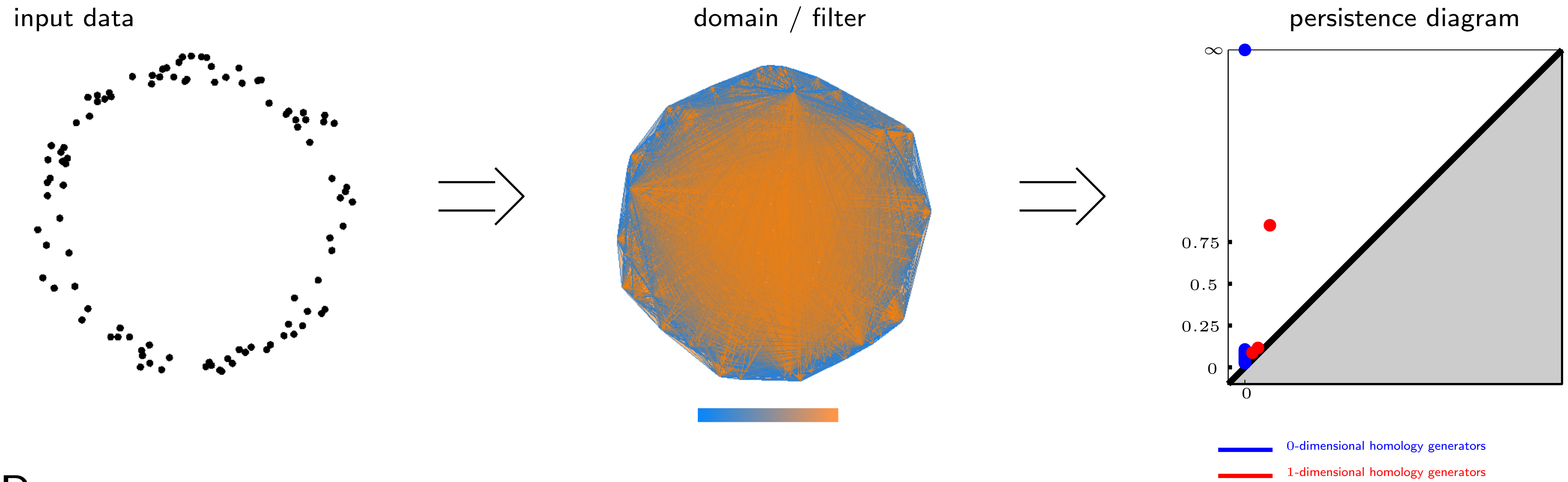
Persistence diagrams as descriptors for data



Pros:

- topological descriptors carry information of a different nature
- strong invariance and stability properties
- flexible and versatile

Persistence diagrams as descriptors for data



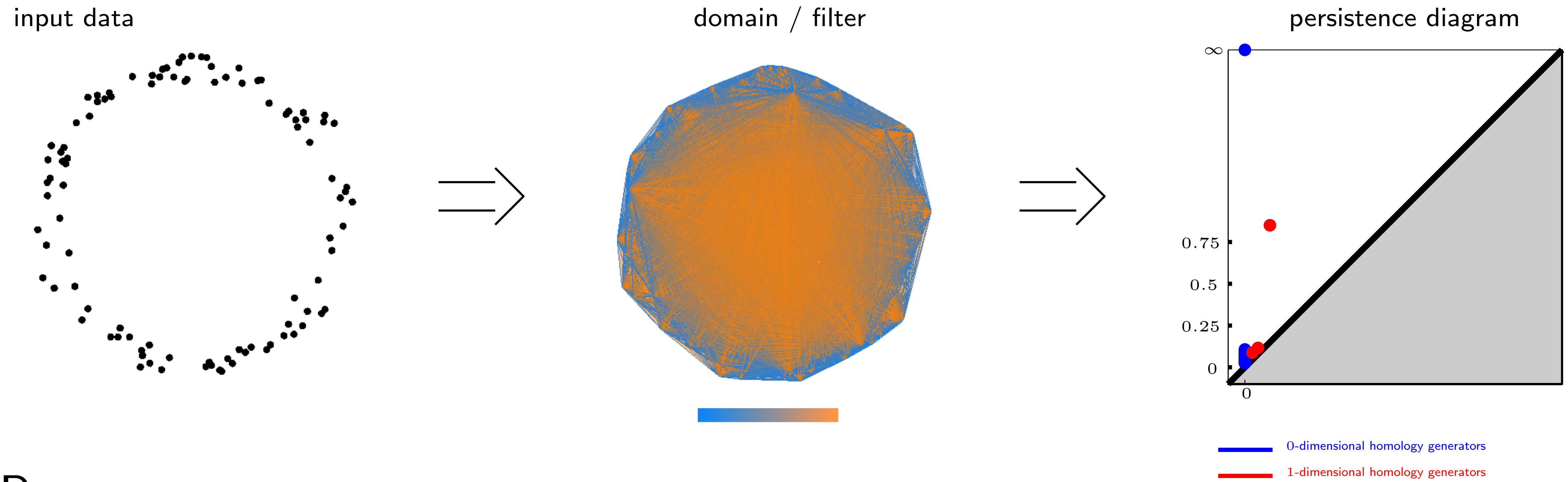
Pros:

- topological descriptors carry information of a different nature
- strong invariance and stability properties
- flexible and versatile

Cons:

- the space of persistence diagrams is not a linear space
  $\rightarrow$ bad for learning and statistics
- descriptors can be slow to compute and (more importantly) to compare
  $\rightarrow$ bad for applications

Persistence diagrams as descriptors for data



Pros:

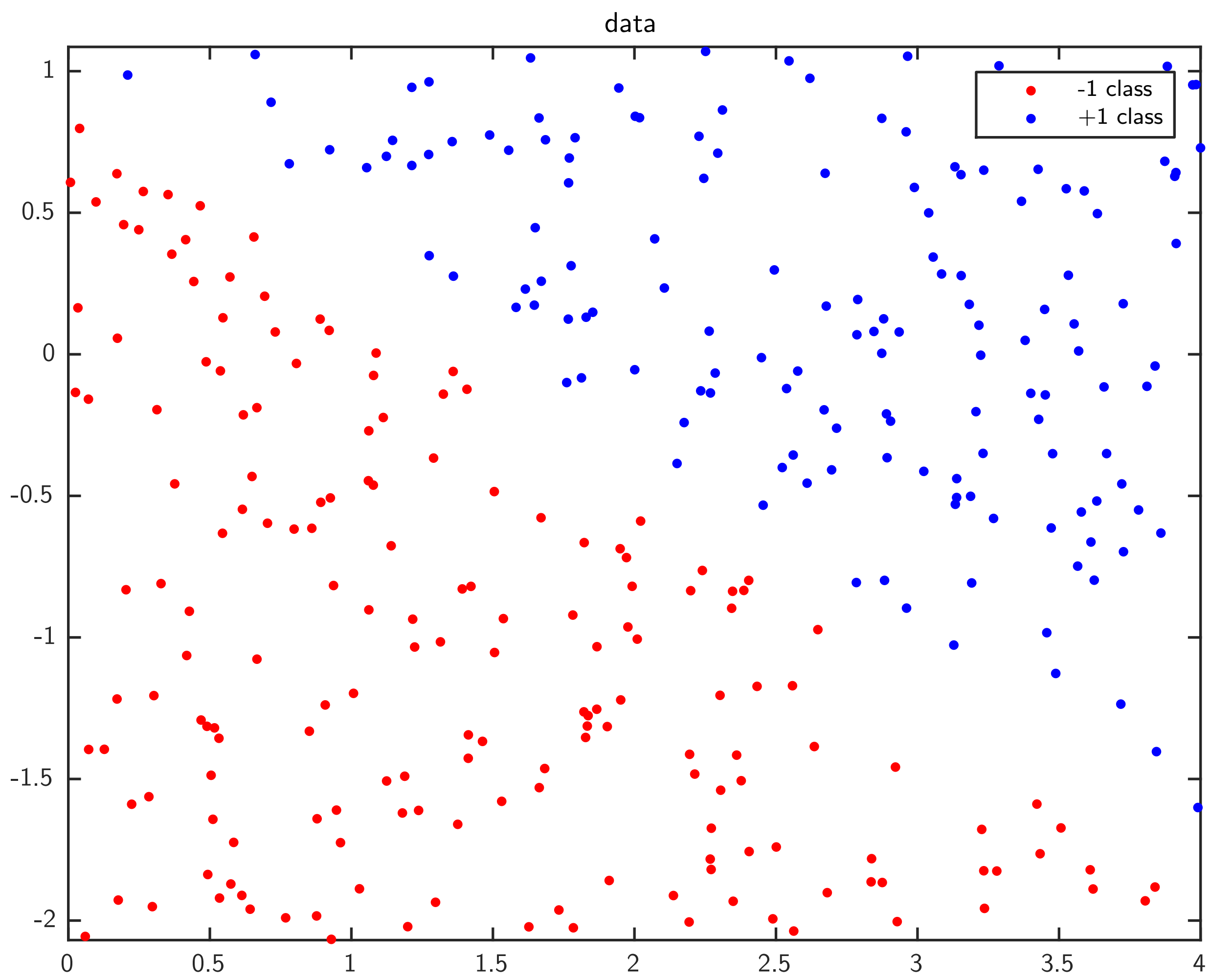
- topological descriptors carry information of a different nature
- strong invariance and stability properties
- flexible and versatile

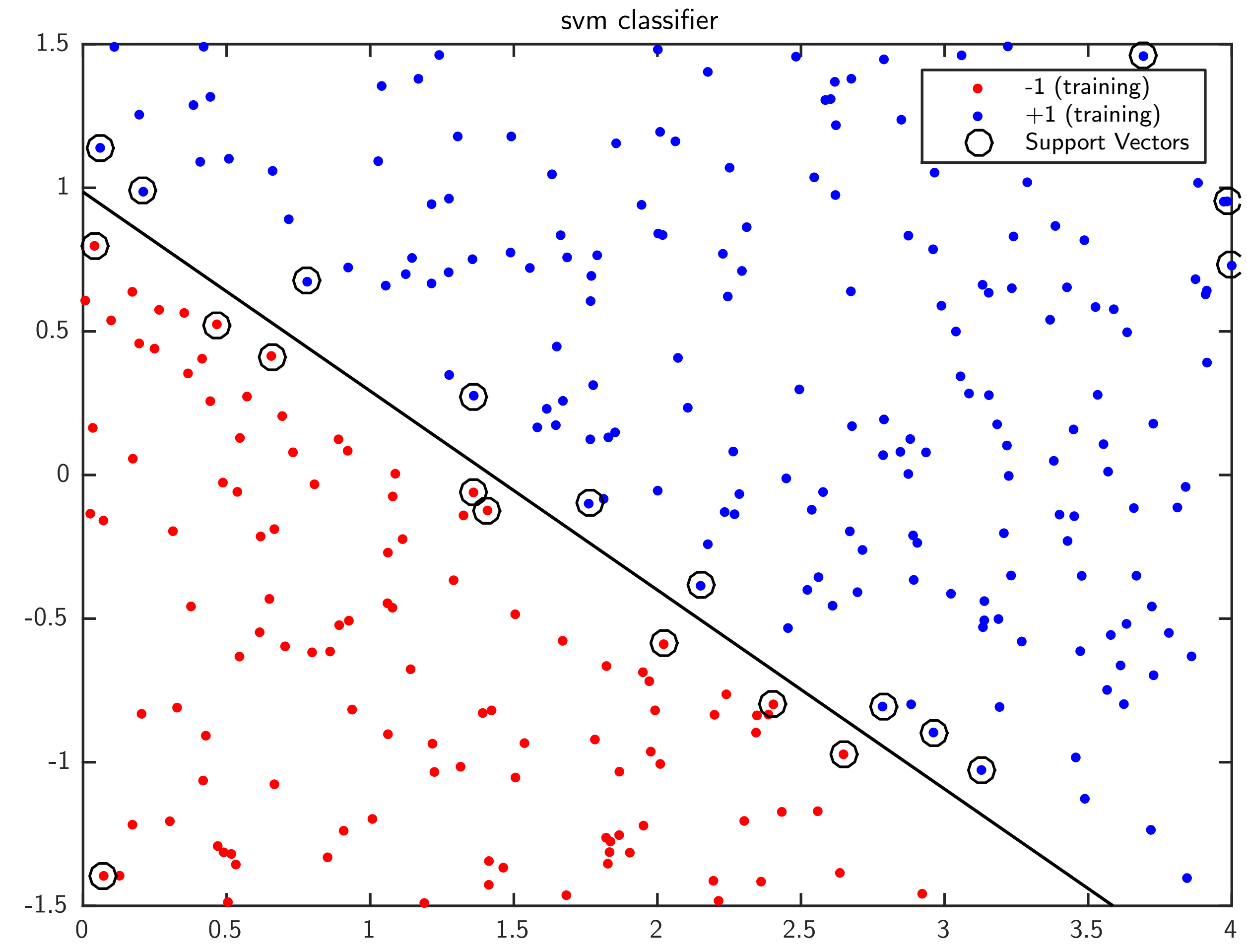
A solution:
Hilbert space embedding

Cons:

- the space of persistence diagrams is not a linear space
  $\rightarrow$ bad for learning and statistics
- descriptors can be slow to compute and (more importantly) to compare
  $\rightarrow$ bad for applications

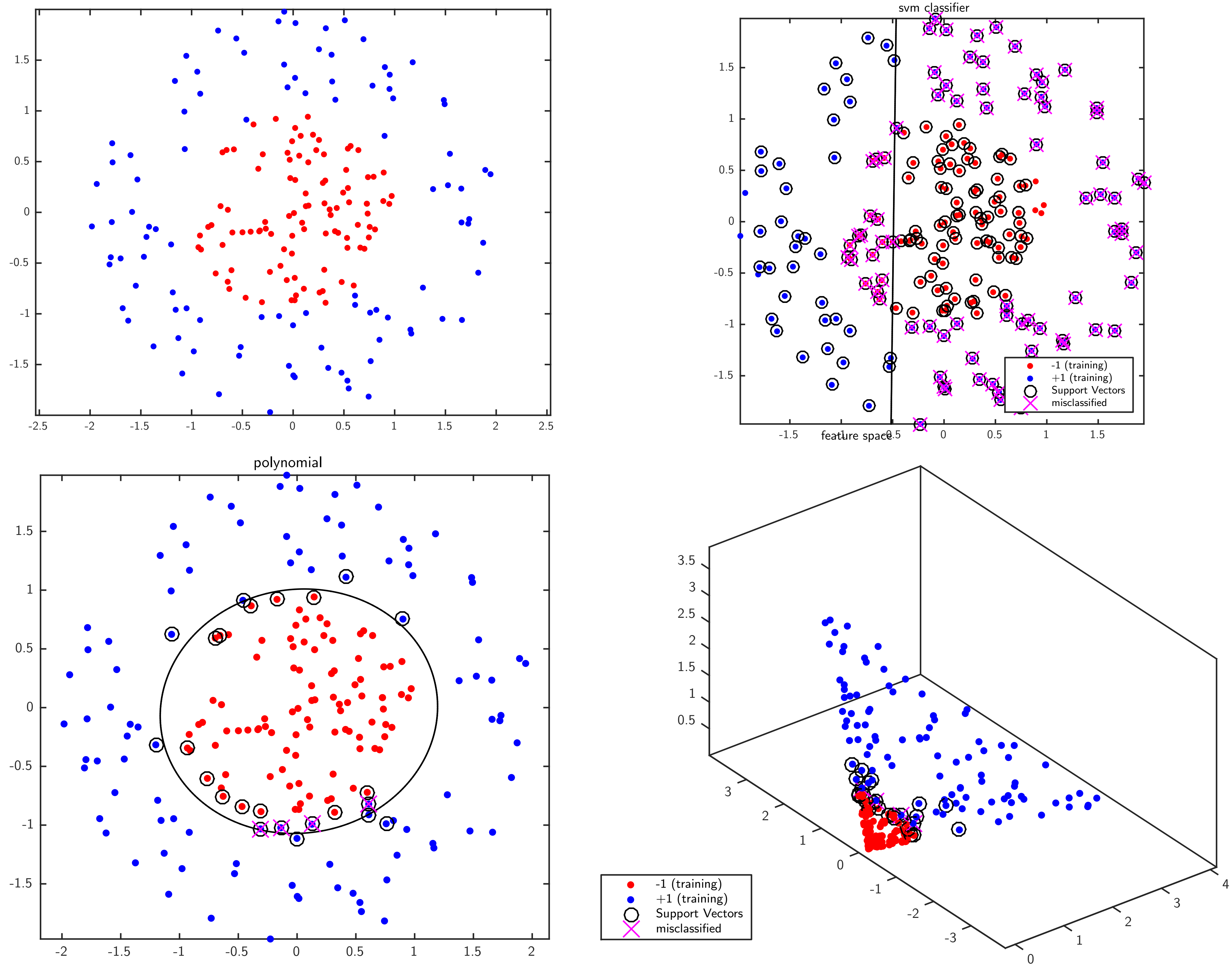
Hilbert space embedding and kernel trick





convex optimization pb. for training, immediate comput. for testing

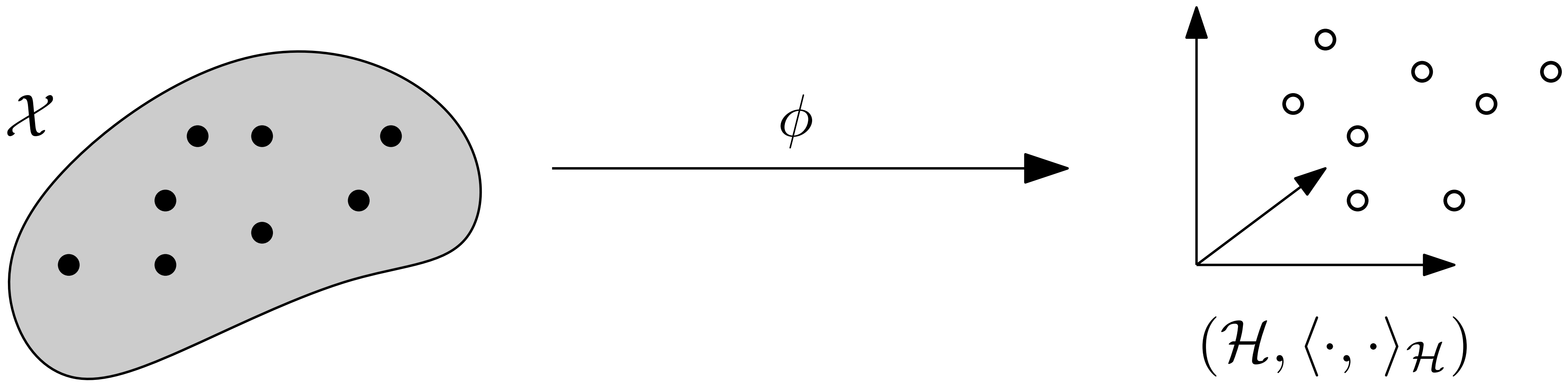
Hilbert space embedding and kernel trick



Hilbert space embedding and kernel trick

$\mathcal{X}$: a space in which we want to compare/classify elements

- feature map $\phi : \mathcal{X} \to \mathcal{H}$ equipped with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$
- lift training/testing data to $\mathcal{H}$ through ϕ then solve learning problem

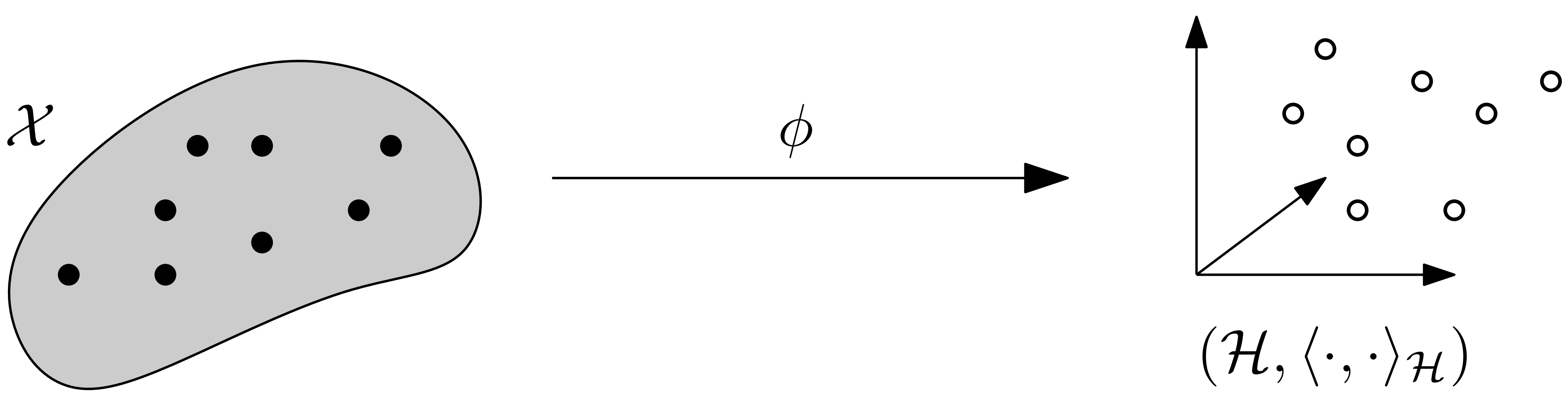


Hilbert space embedding and kernel trick

$\mathcal{X}$: a space in which we want to compare/classify elements

- feature map $\phi : \mathcal{X} \to \mathcal{H}$ equipped with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$
- lift training/testing data to $\mathcal{H}$ through ϕ then solve learning problem
- observation: many learning methods only use the inner product

$\to$ do not lift the data, instead compute the $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$



Hilbert space embedding and kernel trick

$\mathcal{X}$: a space in which we want to compare/classify elements

- feature map $\phi : \mathcal{X} \to \mathcal{H}$ equipped with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$
- lift training/testing data to $\mathcal{H}$ through ϕ then solve learning problem
- observation: many learning methods only use the inner product

$\to$ do not lift the data, instead compute the $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$

Def.: A *reproducing kernel* is a map $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $k(\cdot, \cdot) = \langle \phi(\cdot), \phi(\cdot) \rangle_{\mathcal{H}}$ for some pair $(\phi, \mathcal{H})$.

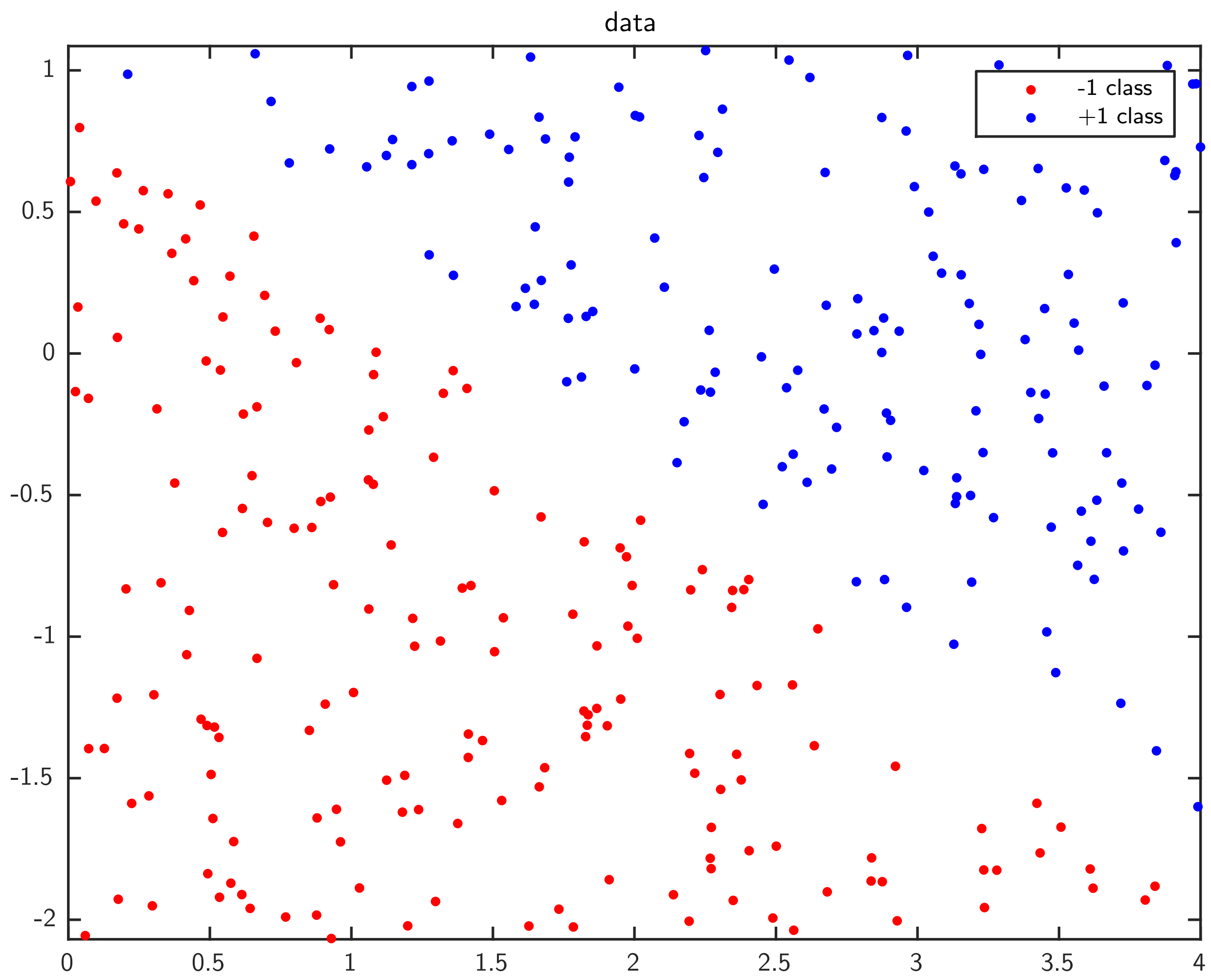
Thm.: [Moore, Aronszajn]
A pair $(\phi, \mathcal{H})$ exists (and is unique) whenever k is *positive semidefinite*, i.e. for all $n \in \mathbb{N}$ and $x_1, \cdots, x_n \in \mathcal{X}$ the Gram matrix $(k(x_i, x_j))_{i,j}$ is positive semidefinite. $\mathcal{H}$ is called the *Reproducing Kernel Hilbert Space* (RKHS) of $\mathcal{X}$.

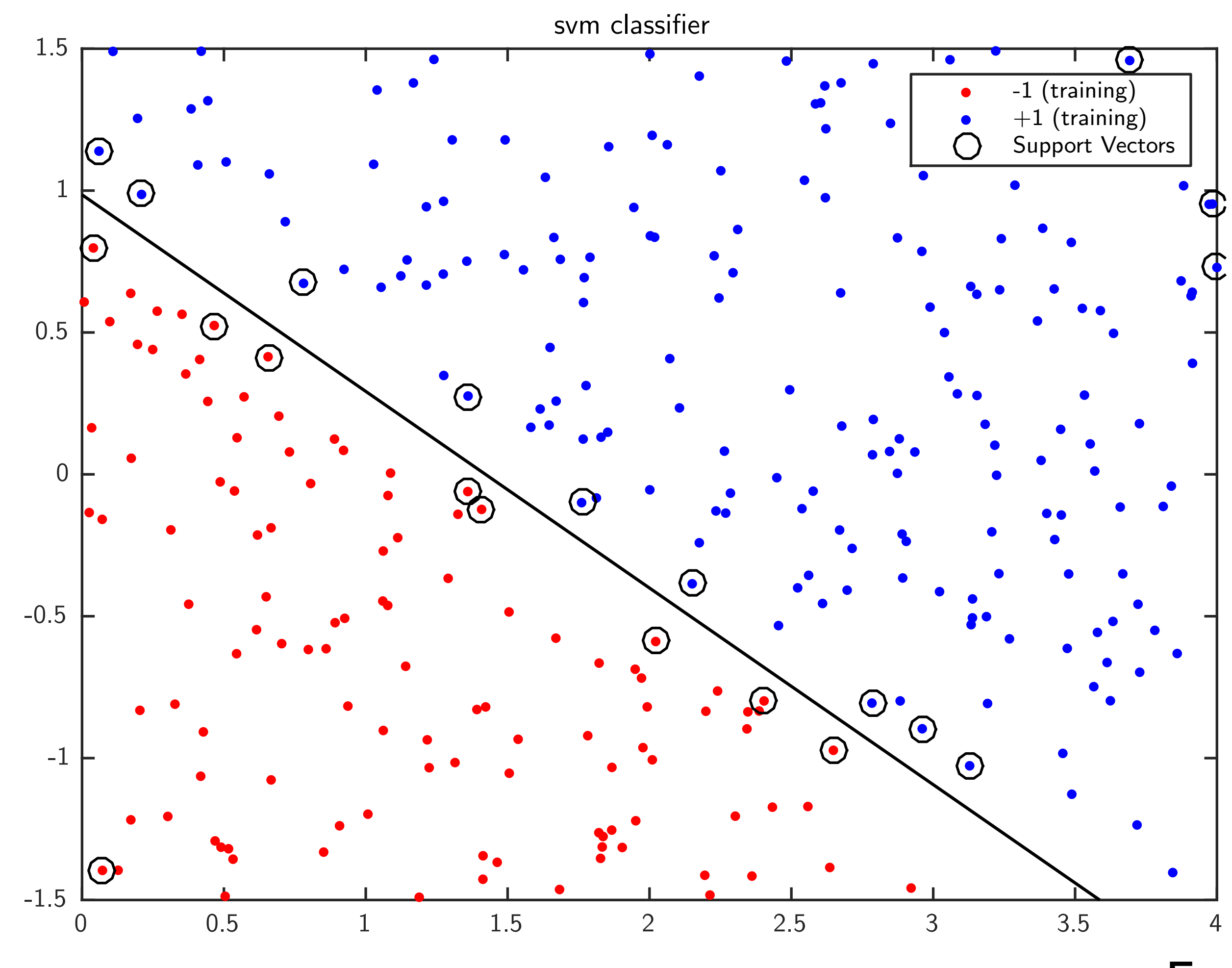
Hilbert space embedding and kernel trick

classification / regression: input: $(x_1, y_1), \cdots, (x_n, y_n) \in \mathcal{X} \times \mathcal{Y}$

$$\min_{w,\ b} \frac{1}{n} \sum_{i=1}^{n} L\left(y_i,\ \langle w, x_i \rangle\ +\ b\right) + \gamma \left\| w \right\|_p$$

(e.g. $L(y,t) = \max\{0,\ 1 - t.y\}$ for SVMs, $p = 1$ for sparsity)





Hilbert space embedding and kernel trick

classification / regression: input: $(x_1, y_1), \cdots, (x_n, y_n) \in \mathcal{X} \times \mathcal{Y}$

$$\min_{w,\ b}\ \frac{1}{n} \sum_{i=1}^{n} L\left(y_i,\ \langle w, \overset{\phi(x_i)}{\cancel{x_i}} \rangle + b\right) + \gamma \left\| w \right\|_p$$

(e.g. $L(y, t) = \max\{0,\ 1 - t.y\}$ for SVMs, $p = 1$ for sparsity)

Thm. (Representer): [Schölkopf, Herbrich, Smola]
The argmin w^* admits a representation of the form:

$$w^* = \sum_{j=1}^{n} \alpha_j \phi(x_j) = \sum_{j=1}^{n} \alpha_j k(x_j, \cdot)$$

$\rightarrow$ search argmin within $\mathrm{span}\{\phi(x_1), \cdots, \phi(x_n)\}$ and optimize for the α_i

Hilbert space embedding and kernel trick

Examples:

- linear kernel: $k(x, y) = \langle x, y \rangle$

- polynomial kernel: $k(x, y) = (\alpha \langle x, y \rangle + 1)^{\beta}$

- $\boxed{\text{Gaussian kernel: } k(x, y) = \exp\left(\frac{-\|x-y\|_2^2}{2\sigma^2}\right)}$

(the swiss knife for kernel based learning in practice)

- ...

Hilbert space embedding and kernel trick

Examples:

- linear kernel: $k(x, y) = \langle x, y \rangle$

- polynomial kernel: $k(x, y) = (\alpha \langle x, y \rangle + 1)^{\beta}$

- Gaussian kernel: $k(x, y) = \exp\left(\frac{-\|x-y\|_2^2}{2\sigma^2} \right)$

Thm.: [Berg, Christensen, Ressel 1984]
If $d : X \times X \to \mathbb{R}_+$ symmetric is *conditionally negative semidefinite*, i.e.:

$$\forall n \in \mathbb{N},\ \forall x_1, \cdots, x_n \in X,\ \sum_{i=1}^{n} \alpha_i = 0 \Longrightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \, d(x_i, x_j) \leq 0,$$

then $k(x, y) := \exp\left(-\frac{d(x,y)}{2\sigma^2} \right)$ is positive semidefinite.

Hilbert space embedding and kernel trick

Q: does this apply to persistence diagrams?

Pb: d_p is not cnsd, for any $p > 0$

Thm.: [Berg, Christensen, Ressel 1984]
If $d : X \times X \to \mathbb{R}_+$ symmetric is *conditionally negative semidefinite*, i.e.:

$$\forall n \in \mathbb{N},\ \forall x_1, \cdots, x_n \in X,\ \sum_{i=1}^{n} \alpha_i = 0 \Longrightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j\, d(x_i, x_j) \leq 0,$$

then $k(x, y) := \exp\left(-\frac{d(x,y)}{2\sigma^2}\right)$ is positive semidefinite.

Kernels for persistence diagrams

View persistence diagrams as:

- landscapes (collections of 1-d functions) [Bubenik 2012] [Bubenik, Dłotko 2015]
- discrete measures:
  - $\rightarrow$ histogram [Bendich et al. 2014]
  - $\rightarrow$ convolution with fixed kernel [Chepushtanova et al. 2015]
  - $\rightarrow$ convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]
  - $\rightarrow$ heat diffusion [Reininghaus et al. 2015] + exponential [Kwit et al. 2015]
- finite metric spaces [C., Oudot, Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

Kernels for persistence diagrams

	landscapes	discrete measures	metric spaces	polynomials
positive (semi-)definiteness	✓	✓	✓	✓
ambient Hilbert space	$L_2(\mathbb{N} \times \mathbb{R})$	$L_2(\mathbb{R}^2)$	$(\mathbb{R}^d, \Vert \cdot \Vert_2)$	$\ell_2(\mathbb{R})$
$\Vert\phi(\cdot) - \phi(\cdot)\Vert_{\mathcal{H}} \leq f(\mathrm{d}_p)$	✓	✓	✓	✓
injectivity	✓	✓	✗	✓
$\Vert\phi(\cdot) - \phi(\cdot)\Vert_{\mathcal{H}} \geq g(\mathrm{d}_p)$	?	?	✗	?
algorithmic cost	$O(n^2)$	$O(n^2)$	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(nd)$ kernel: $O(d)$
universality	✗	✓	✗	✗
additivity	✗	✓	✗	✗

Kernels for persistence diagrams

	landscapes	discrete measures	metric spaces	polynomials
positive (semi-)definiteness	✓	✓	✓	✓
ambient Hilbert space	$L_2(\mathbb{N}\times\mathbb{R})$	$L_2(\mathbb{R}^2)$	$(\mathbb{R}^d, \|\cdot\|_2)$	$\ell_2(\mathbb{R})$
$\|\phi(\cdot)-\phi(\cdot)\|_{\mathcal{H}} \leq f(\mathrm{d}_p)$	✓	✓	✓	✓
injectivity	✓	✓	✗	✓
$\|\phi(\cdot)-\phi(\cdot)\|_{\mathcal{H}} \geq g(\mathrm{d}_p)$	?	?	✗	?
algorithmic cost	$O(n^2)$	$O(n^2)$	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(nd)$ kernel: $O(d)$
universality	✗	✓	✗	✗
additivity	✗	✓	✗	✗

Kernels for persistence diagrams

	landscapes	discrete measures	metric spaces	polynomials
positive (semi-)definiteness	✓	✓	✓	✓
ambient Hilbert space	$L_2(\mathbb{N} \times \mathbb{R})$	$L_2(\mathbb{R}^2)$	$(\mathbb{R}^d, \|\cdot\|_2)$	$\ell_2(\mathbb{R})$
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq f(\mathrm{d}_p)$	✓	✓	✓	✓
injectivity	✓	✓	✗	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \geq g(\mathrm{d}_p)$	?	?	✗	?
algorithmic cost	$O(n^2)$	$O(n^2)$	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(nd)$ kernel: $O(d)$
universality	✗	✓	✗	✗
additivity	✗	✓	✗	✗

Kernels for persistence diagrams

View persistence diagrams as:

- landscapes (collections of 1-d functions) [Bubenik 2012] [Bubenik, Dłotko 2015]
- discrete measures:

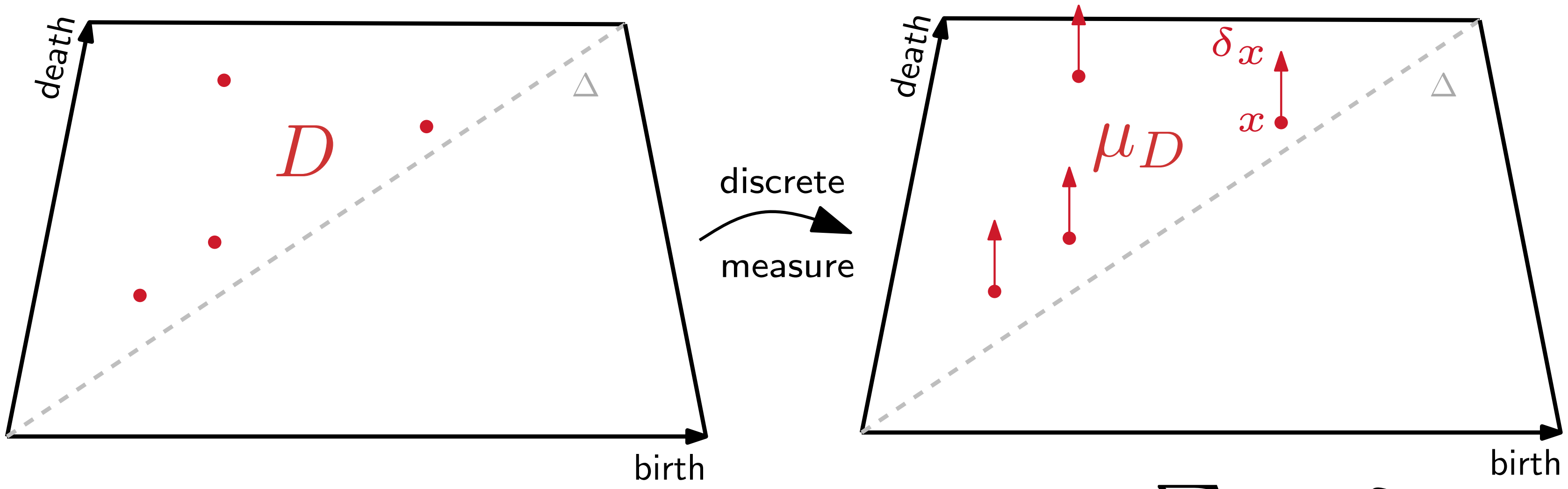
  $\rightarrow$ histogram [Bendich et al. 2014]

  $\rightarrow$ convolution with fixed kernel [Chepushtanova et al. 2015]

  $\rightarrow$ convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

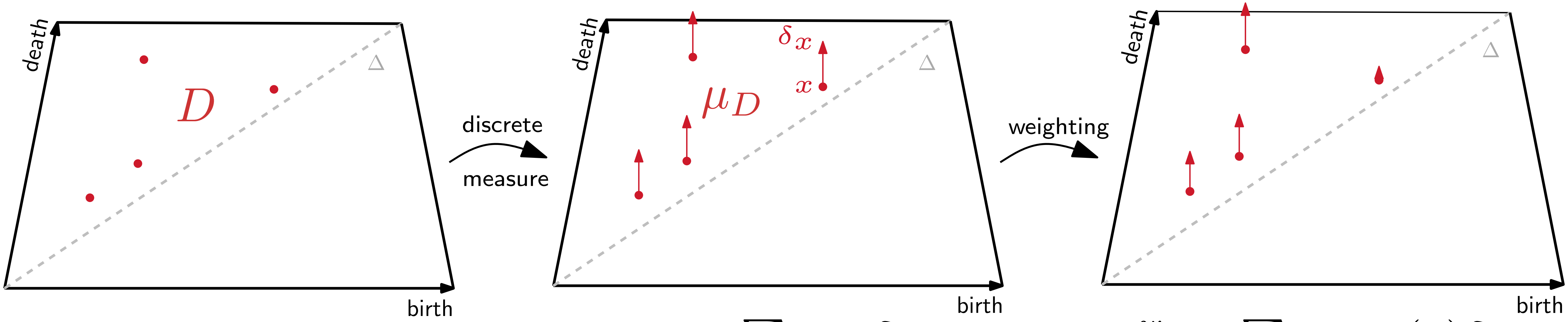
  $\rightarrow$ heat diffusion [Reininghaus et al. 2015] $+$ exponential [Kwit et al. 2015]
- finite metric spaces [C., Oudot, Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

Persistence diagrams as discrete measures (I)



$$\mu_D := \sum_{x \in D} \delta_x$$

Persistence diagrams as discrete measures (I)

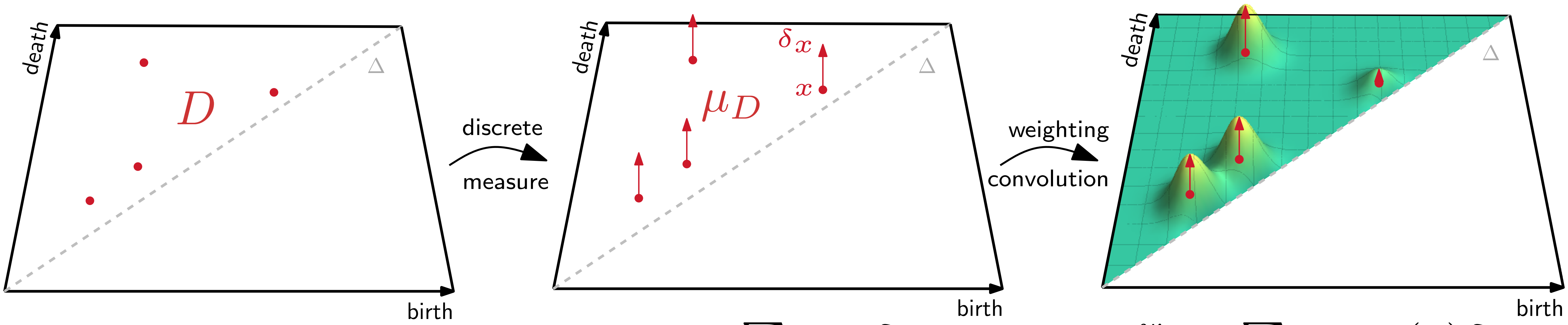


$\mu_D := \sum_{x \in D} \delta_x$ $\quad$ $\mu_D^w := \sum_{x \in D} w(x) \delta_{\bar{x}}$

Pb: μ_D is unstable (points on diagonal disappear)

$$w(x) := \arctan\left(c\,\mathrm{d}(x, \Delta)^r\right),\ c, r > 0$$

Persistence diagrams as discrete measures (I)



$\mu_D := \sum_{x \in D} \delta_x$

$\mu_D^w := \sum_{x \in D} w(x)\delta_{\bar{x}}$

$\tilde{\mu}_D^w := \mu_D^w * \mathcal{N}(0, \sigma)$

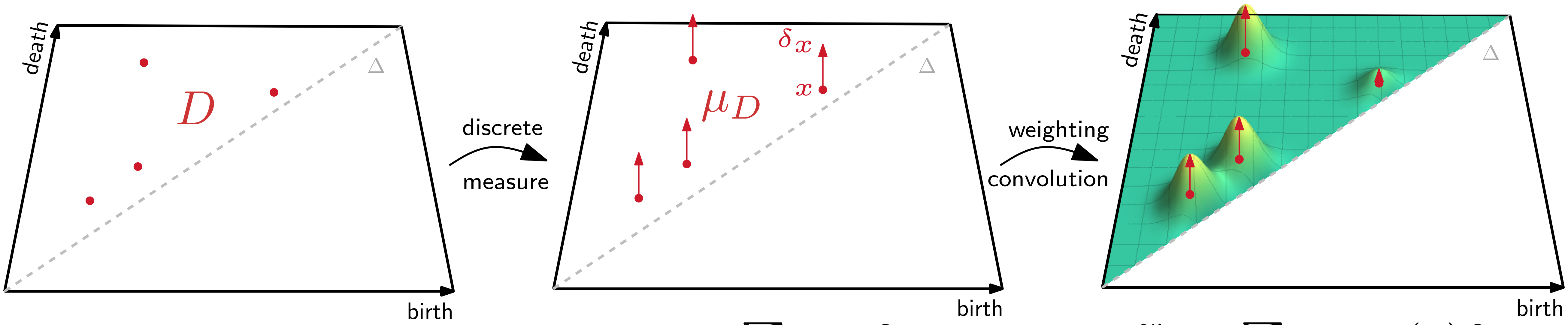
Pb: μ_D is unstable (points on diagonal disappear)

$w(x) := \arctan\left(c\, \mathrm{d}(x, \Delta)^r\right),\ c, r > 0$

Def: $\phi(D)$ is the density function of $\mu_D^w * \mathcal{N}(0, \sigma)$ w.r.t. Lebesgue measure:

$$\left(\begin{array}{l} \phi(D) := \dfrac{1}{\sqrt{2\pi}\sigma} \displaystyle\sum_{x \in D} \arctan(c\, \mathrm{d}(x, \Delta)^r) \exp\left(-\dfrac{\|\cdot - x\|^2}{2\sigma^2}\right) \\ k(D, D') := \langle \phi(D),\ \phi(D') \rangle_{L_2(\Delta \times \mathbb{R}_+)} \end{array} \right.$$

Persistence diagrams as discrete measures (I)



$\mu_D := \sum_{x\in D} \delta_x$

$\mu_D^w := \sum_{x\in D} w(x)\delta_{\bar{x}}$

$\tilde{\mu}_D^w := \mu_D^w * \mathcal{N}(0,\sigma)$

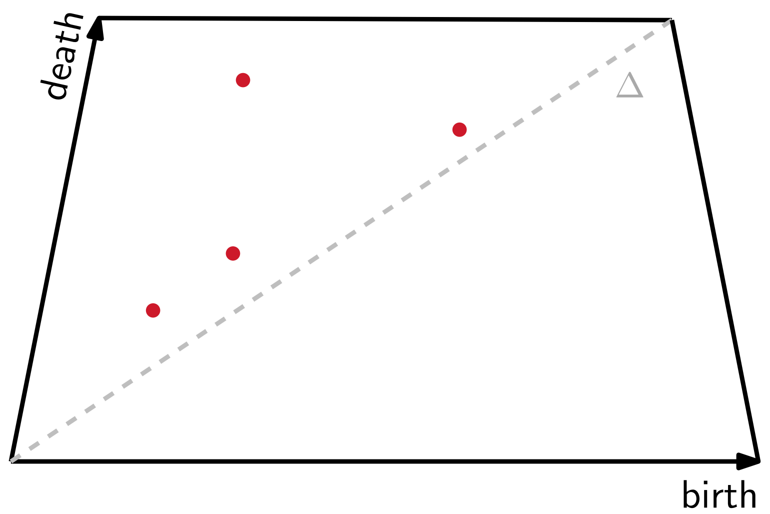
Prop.: [Kusano, Fukumisu, Hiraoka 2016-17]

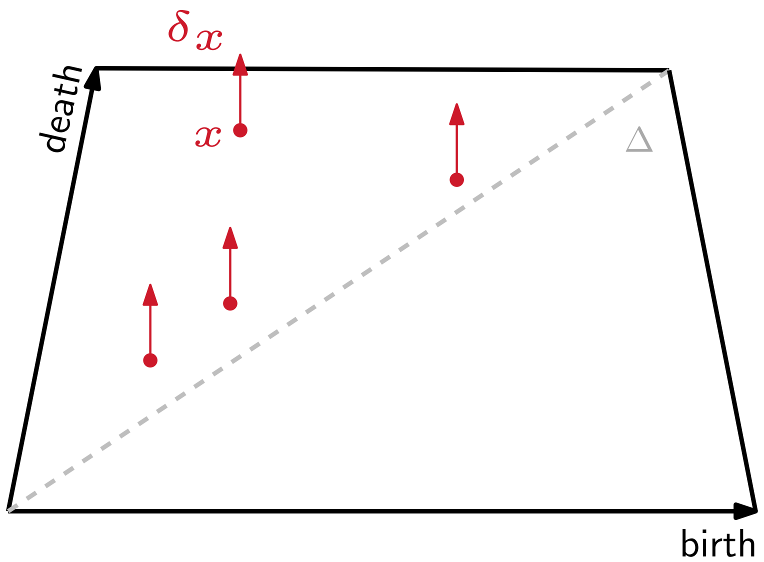
- $\|\phi(D)-\phi(D')\|_{\mathcal{H}} \leq C\, d_p(D,D')$.
- ϕ is injective and $\exp(k)$ is universal

Pb: convolution reduces discriminativity $\rightarrow$ use discrete measure instead

$$\begin{cases} \phi(D) := \dfrac{1}{\sqrt{2\pi}\sigma} \displaystyle\sum_{x\in D} \arctan(c\,\mathrm{d}(x,\Delta)^r)\, \exp\left(-\dfrac{\|\cdot - x\|^2}{2\sigma^2}\right) \\ k(D,D') := \langle \phi(D),\, \phi(D')\rangle_{L_2(\Delta\times\mathbb{R}_+)} \end{cases}$$

Persistence diagrams as discrete measures (II)

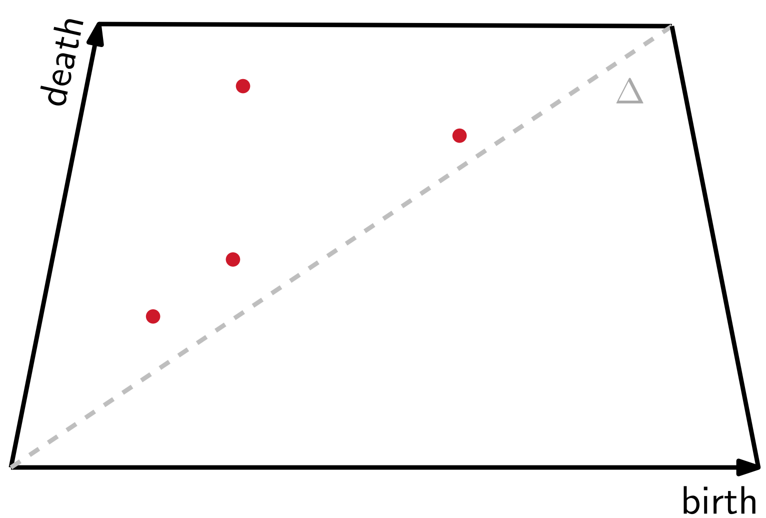


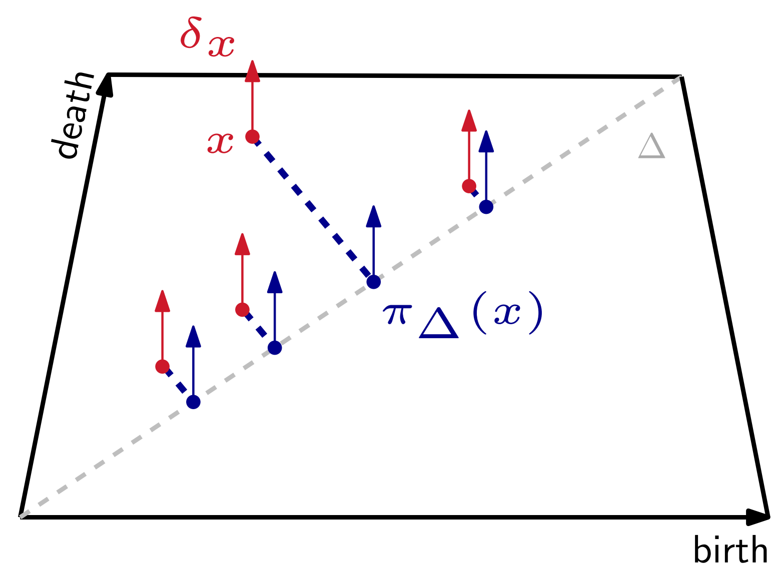


$$\mu_D := \sum_{x \in D} \delta_x$$

Pb: $\mathrm{d}_p(D, D') \not\propto W_p(\mu_D, \mu_{D'})$ (W_p does not even make sense here)

Persistence diagrams as discrete measures (II)





$\mu_D := \sum_{x \in D} \delta_x$

Pb: $\mathrm{d}_p(D, D') \not\propto W_p(\mu_D, \mu_{D'})$ (W_p does not even make sense here)

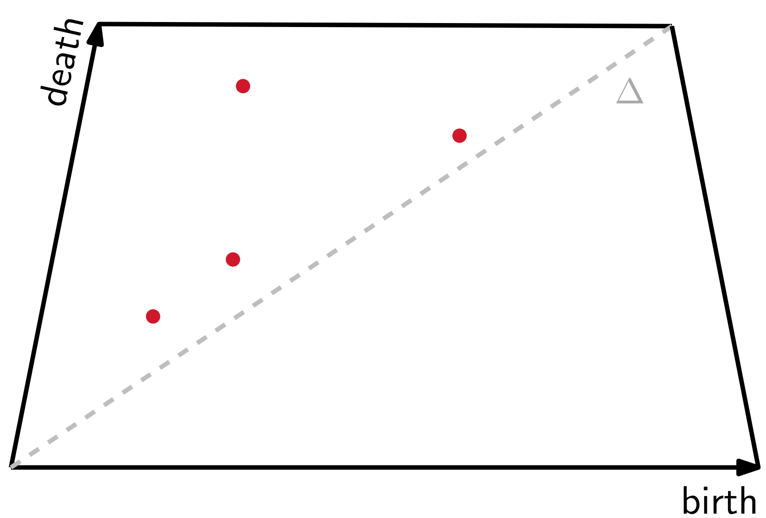
$\rightarrow$ given D, D', let

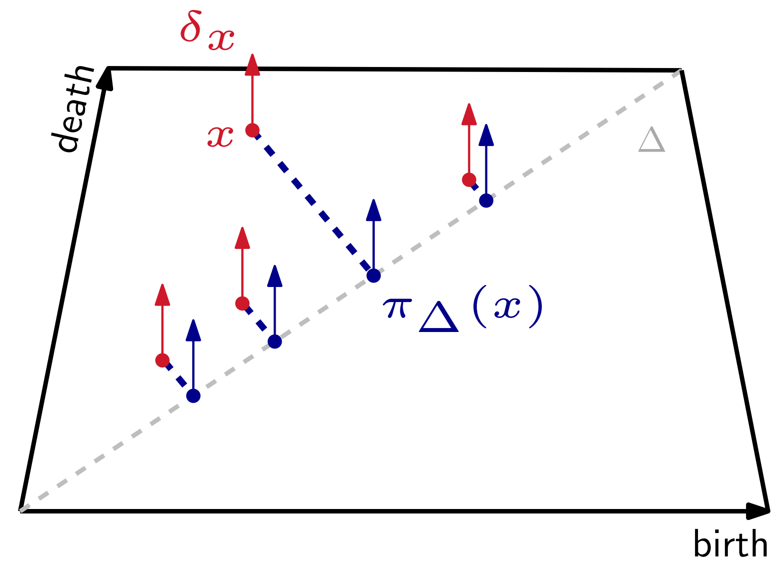
$$\bar{\mu}_D := \sum_{x \in D} \delta_x + \sum_{y \in D'} \delta_{\pi_\Delta(y)}$$

$$\bar{\mu}_{D'} := \sum_{y \in D'} \delta_y + \sum_{x \in D} \delta_{\pi_\Delta(x)}$$

Then, $\mathrm{d}_p(D, D') \leq W_p(\bar{\mu}_D, \bar{\mu}_{D'}) \leq 2\, \mathrm{d}_p(D, D')$

Persistence diagrams as discrete measures (II)





$\mu_D := \sum_{x\in D} \delta_x$

Pb: $\mathrm{d}_p(D, D') \not\propto W_p(\mu_D, \mu_{D'})$ (W_p does not even make sense here)

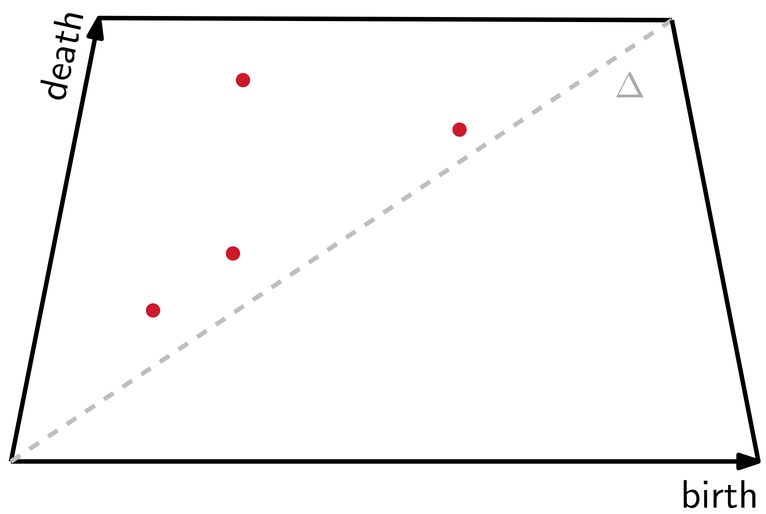
$\rightarrow$ given D, D', let

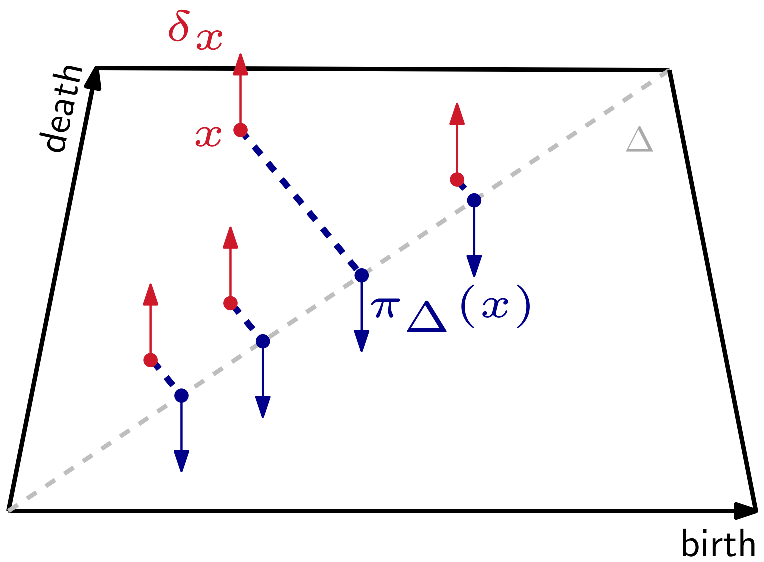
$$\bar{\mu}_D := \sum_{x\in D} \delta_x + \sum_{y\in D'} \delta_{\pi_\Delta(y)}$$

$$\bar{\mu}_{D'} := \sum_{y\in D'} \delta_y + \sum_{x\in D} \delta_{\pi_\Delta(x)}$$

Then, $\mathrm{d}_p(D, D') \leq W_p(\bar{\mu}_D, \bar{\mu}_{D'}) \leq 2\,\mathrm{d}_p(D, D')$ **Pb:** $\bar{\mu}_D$ depends on D'

Persistence diagrams as discrete measures (II)





$\mu_D := \sum_{x\in D} \delta_x$

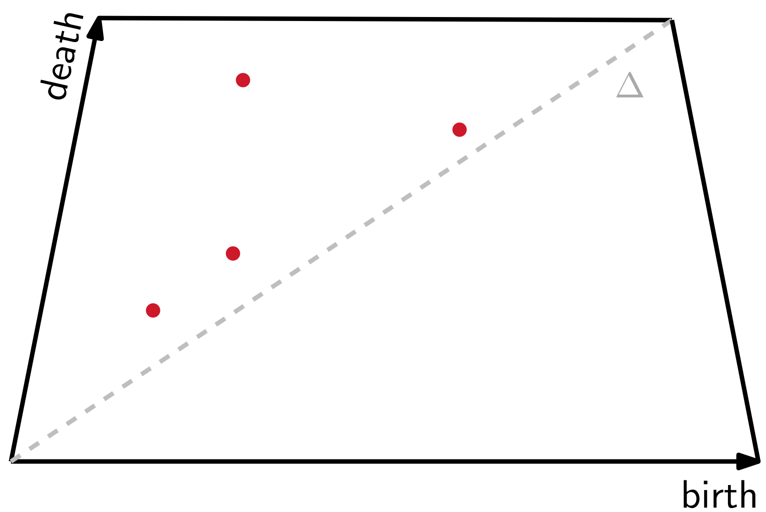
Pb: $\mathrm{d}_p(D, D') \not\propto W_p(\mu_D, \mu_{D'})$ (W_p does not even make sense here)

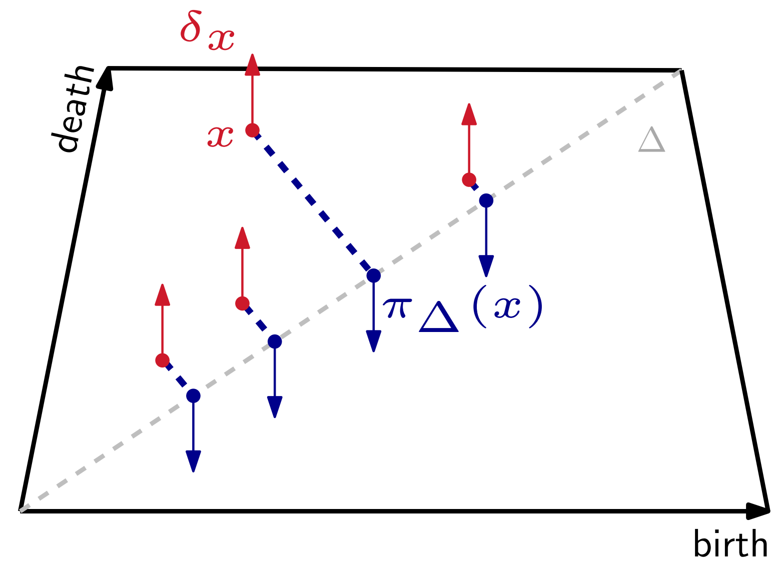
Solution: transfer mass negatively to μ_D:

$$\tilde{\mu}_D := \sum_{x\in D} \delta_x - \sum_{x\in D} \delta_{\pi_\Delta(x)} \quad \in \mathcal{M}_0(\mathbb{R}^2)$$

$\rightarrow$ signed disrete measure of total mass zero

Persistence diagrams as discrete measures (II)





$\mu_D := \sum_{x \in D} \delta_x$

Pb: $\mathrm{d}_p(D, D') \not\propto W_p(\mu_D, \mu_{D'})$ (W_p does not even make sense here)

Solution: transfer mass negatively to μ_D:

$$\tilde{\mu}_D := \sum_{x \in D} \delta_x - \sum_{x \in D} \delta_{\pi_\Delta(x)} \quad \in \mathcal{M}_0(\mathbb{R}^2)$$

$\rightarrow$ signed disrete measure of total mass zero

Q: What metric should we use?

A: Kantorovich norm $\|\cdot\|_K$

Persistence diagrams as discrete measures (II)

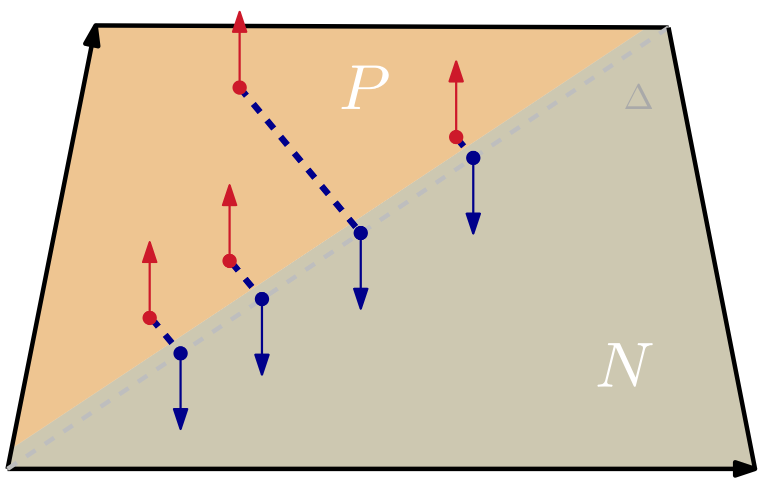
Hahn decomposition thm.: For any $\mu \in \mathcal{M}_0(X, \Sigma)$, there exist measurable sets P, N such that:

(i) $P \cup N = X$ and $P \cap N = \emptyset$

(ii) $\mu(B) \geq 0$ for every measureable set $B \subseteq P$

(iii) $\mu(B) \leq 0$ for every measureable set $B \subseteq N$

Moreover, the decomposition is essentially unique.



Solution: transfer mass negatively to μ_D:

$$\tilde{\mu}_D := \sum_{x \in D} \delta_x - \sum_{x \in D} \delta_{\pi_\Delta(x)} \quad \in \mathcal{M}_0(\mathbb{R}^2)$$

$\rightarrow$ signed disrete measure of total mass zero

Q: What metric should we use?

A: Kantorovich norm $\|\cdot\|_K$

Persistence diagrams as discrete measures (II)

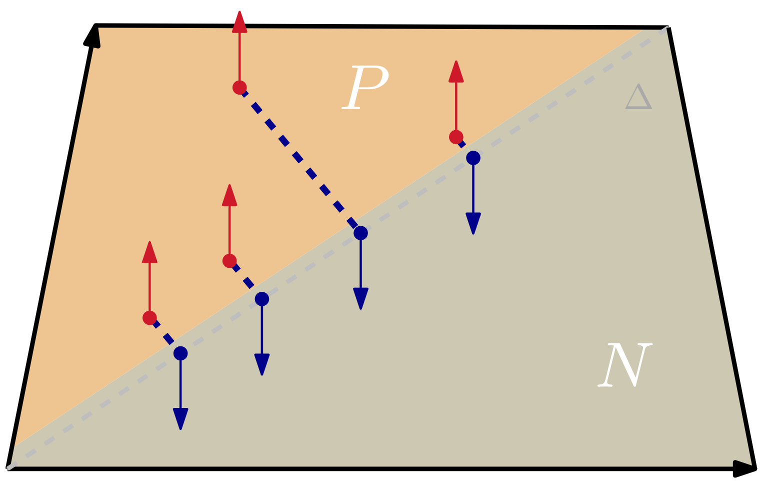
Hahn decomposition thm.: For any $\mu \in \mathcal{M}_0(X,\Sigma)$, there exist measurable sets P, N such that:

(i) $P \cup N = X$ and $P \cap N = \emptyset$

(ii) $\mu(B) \geq 0$ for every measureable set $B \subseteq P$

(iii) $\mu(B) \leq 0$ for every measureable set $B \subseteq N$

Moreover, the decomposition is essentially unique.



$\forall B \in \Sigma$, let $\mu^+(B) := \mu(B \cap P)$ and $\mu^-(B) := -\mu(B \cap N) \in \mathcal{M}_+(X)$

Def.: $\|\mu\|_K := W_1(\mu^+,\, \mu^-)$

Prop.: $\forall \mu, \nu \in \mathcal{M}_0(X)$, $\quad W_1(\mu^+ + \nu^-,\, \nu^+ + \mu^-) = \|\mu - \nu\|_K$

Persistence diagrams as discrete measures (II)

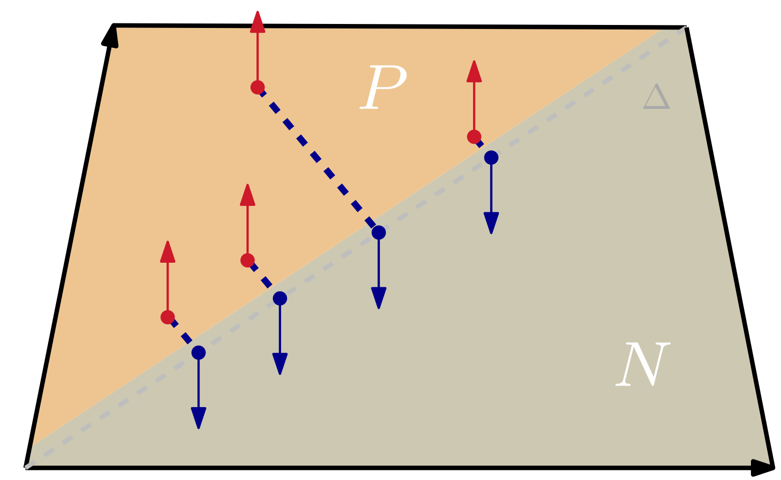
Hahn decomposition thm.: For any $\mu \in \mathcal{M}_0(X, \Sigma)$, there exist measurable sets P, N such that:

(i) $P \cup N = X$ and $P \cap N = \emptyset$

(ii) $\mu(B) \geq 0$ for every measureable set $B \subseteq P$

(iii) $\mu(B) \leq 0$ for every measureable set $B \subseteq N$

Moreover, the decomposition is essentially unique.



$\forall B \in \Sigma$, let $\mu^+(B) := \mu(B \cap P)$ and $\mu^-(B) := -\mu(B \cap N) \in \mathcal{M}_+(X)$

Def.: $\|\mu\|_K := W_1(\mu^+, \mu^-)$

Prop.: $\forall \mu, \nu \in \mathcal{M}_0(X)$, $\quad W_1(\underbrace{\mu^+ + \nu^-}_{\bar{\mu}_D}, \underbrace{\nu^+ + \mu^-}_{\bar{\mu}_{D'}}) = \|\underset{\uparrow\atop\tilde{\mu}_D}{\mu} - \underset{\uparrow\atop\tilde{\mu}_{D'}}{\nu}\|_K$

for persistence diagrams:

$$W_1(\bar{\mu}_D, \bar{\mu}_{D'}) = \|\tilde{\mu}_D - \tilde{\mu}_{D'}\|_K$$

A Wasserstein Gaussian kernel for PDs?

Thm.: [Berg, Christensen, Ressel 1984]
If $d : X \times X \to \mathbb{R}_+$ symmetric is *conditionally negative semidefinite*, i.e.:

$$\forall n \in \mathbb{N},\ \forall x_1, \cdots, x_n \in X,\ \sum_{i=1}^{n} \alpha_i = 0 \Longrightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \, d(x_i, x_j) \leq 0,$$

then $k(x, y) := \exp\left(-\frac{d(x,y)}{2\sigma^2}\right)$ is positive semidefinite.

Pb: W_1 is not cnsd, neither is d_1

Solutions:

- relax the measures (e.g. convolution)
- relax the metric (e.g. regularization, slicing)

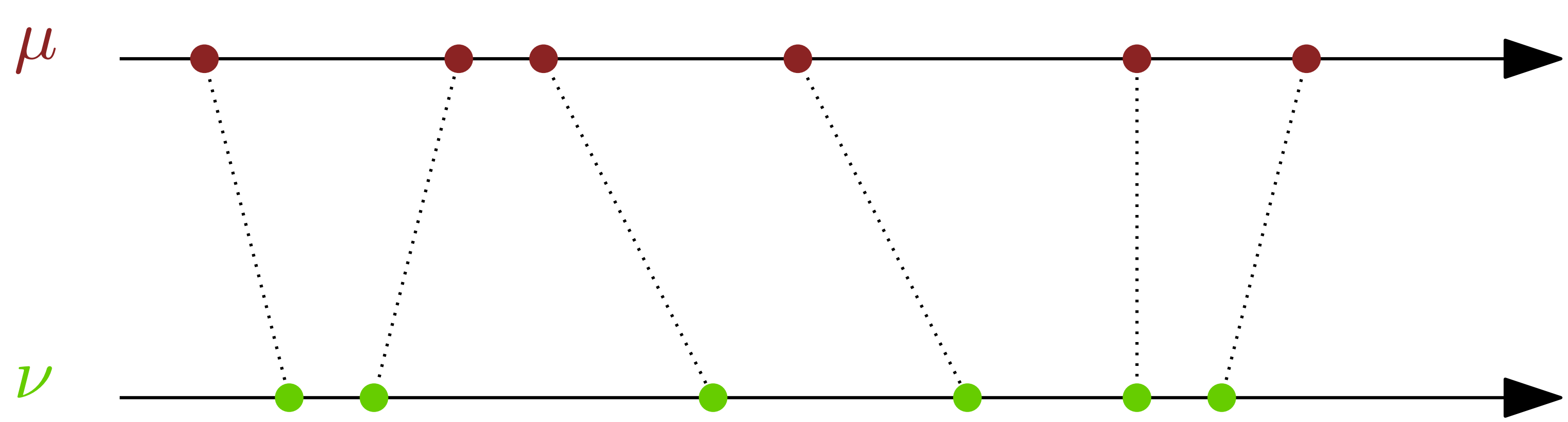
Sliced Wasserstein metric

Special case: $X = \mathbb{R}$, μ, ν discrete measures of mass n

$\mu := \sum_{i=1}^{n} \delta_{x_i}, \;\; \nu := \sum_{i=1}^{n} \delta_{y_i}$

Sort the atoms of μ, ν along the real line: $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for all i

Then: $W_1(\mu, \nu) = \sum_{i=1}^{n} |x_i - y_i| = \|(x_1, \cdots, x_n) - (y_1, \cdots, y_n)\|_1$



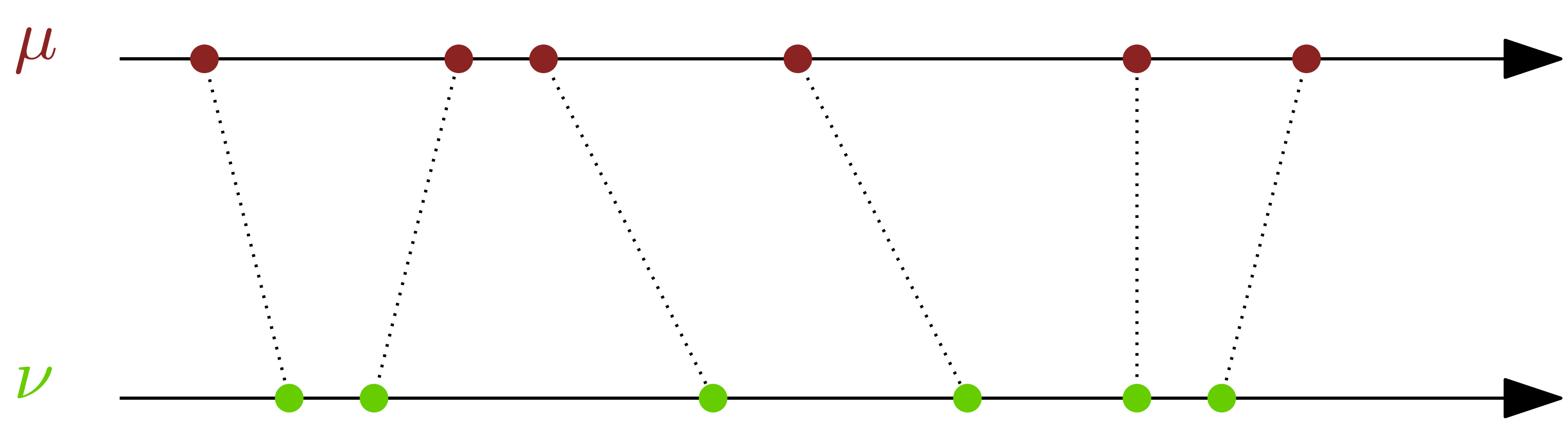
Sliced Wasserstein metric

Special case: $X = \mathbb{R}$, μ, ν discrete measures of mass n

$\mu := \sum_{i=1}^{n} \delta_{x_i}, \;\; \nu := \sum_{i=1}^{n} \delta_{y_i}$

Sort the atoms of μ, ν along the real line: $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for all i

Then: $W_1(\mu, \nu) = \sum_{i=1}^{n} |x_i - y_i| = \|(x_1, \cdots, x_n) - (y_1, \cdots, y_n)\|_1$



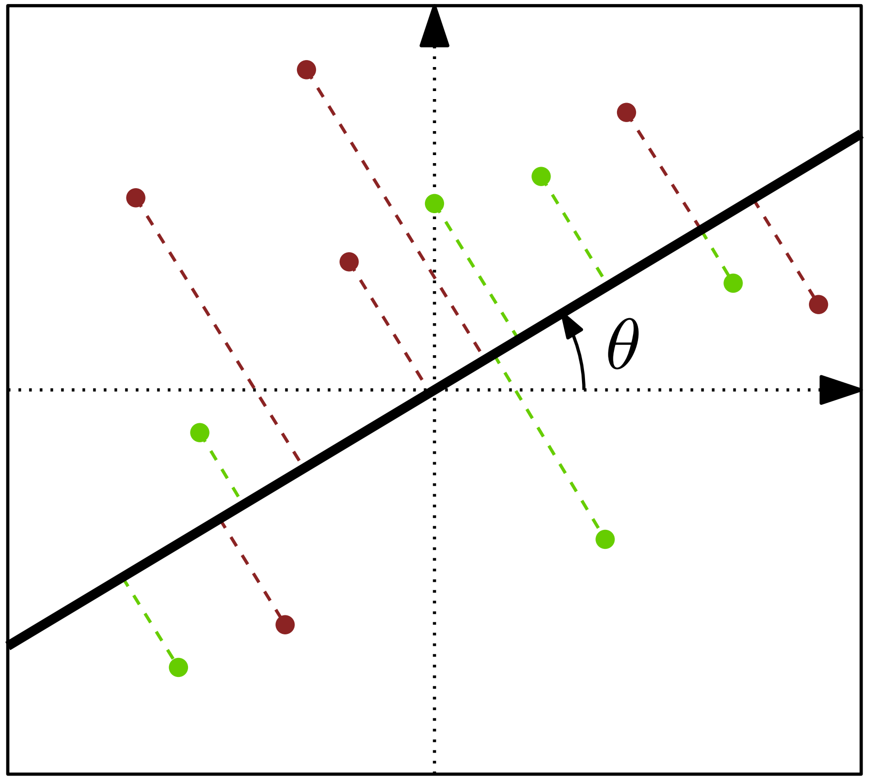
$\rightarrow W_1$ is cnsd and easy to compute (same with $\|\cdot\|_K$ for signed measures)

Sliced Wasserstein metric

Def (sliced Wasserstein distance): for $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^2)$,

$$SW_1(\mu, \nu) := \frac{1}{2\pi} \int_{\theta \in \mathbb{S}^1} W_1(\pi_\theta \# \mu,\ \pi_\theta \# \nu)\, d\theta$$

where π_θ = orthogonal projection onto line passing through origin with angle θ.



Sliced Wasserstein metric

Def (sliced Wasserstein distance): for $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^2)$,

$$SW_1(\mu, \nu) := \frac{1}{2\pi} \int_{\theta \in \mathbb{S}^1} W_1(\pi_\theta \# \mu,\, \pi_\theta \# \nu)\, d\theta$$

where $\pi_\theta =$ orthogonal projection onto line passing through origin with angle θ.

Props: (inherited from W_1 over $\mathbb{R}$) [Rabin, Peyré, Delon, Bernot 2011]

- satisfies the axioms of a metric

- well-defined barycenters, fast to compute via stochastic gradient descent, etc.

- conditionally negative semidefinite

Sliced Wasserstein kernel

Def: Given $\sigma > 0$, for any $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^2)$:

$$k_{SW}(\mu, \nu) := \exp\left(-\frac{SW_1(\mu, \nu)}{2\sigma^2}\right)$$

Corllary: [Kolouri, Zou, Rohde] (from SW cnsd and Berg's theorem) k_{SW} is positive semidefinite.

Sliced Wasserstein kernel

Def: Given $\sigma > 0$, for any $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^2)$:

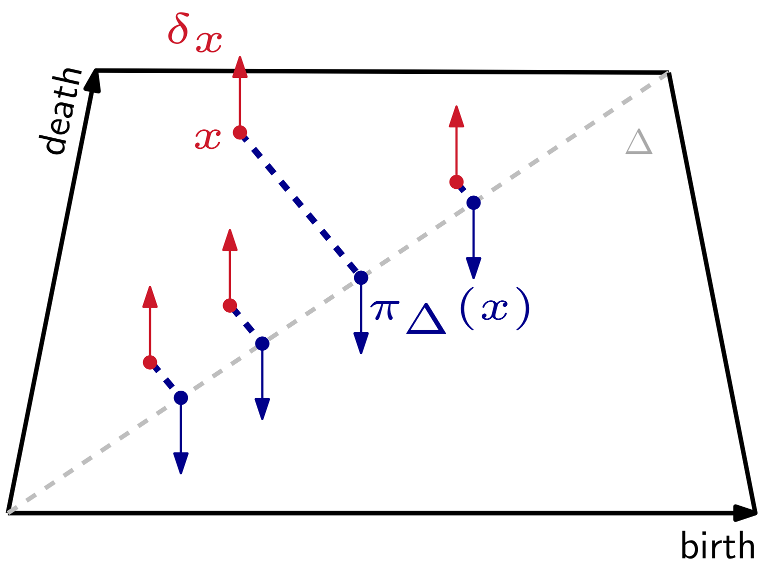
$$k_{SW}(\mu, \nu) := \exp\left(-\frac{SW_1(\mu, \nu)}{2\sigma^2}\right)$$

Corllary: [Kolouri, Zou, Rohde] (from SW cnsd and Berg's theorem)
k_{SW} is positive semidefinite.

$\rightarrow$ application to persistence diagrams:

$$D \mapsto \mu_D := \sum_{x \in D} \delta_x$$

$$\mapsto \tilde{\mu}_D := \mu_D - \pi_\Delta \# \mu_D$$



$$SW_1(D, D') := \int_{\theta \in \mathcal{S}^1} \|\pi_\theta \# \tilde{\mu}_D - \pi_\theta \# \tilde{\mu}_{D'}\|_K \, d\theta$$

$$k_{SW}(D, D') := \exp\left(-\frac{SW_1(D, D')}{2\sigma^2}\right)$$

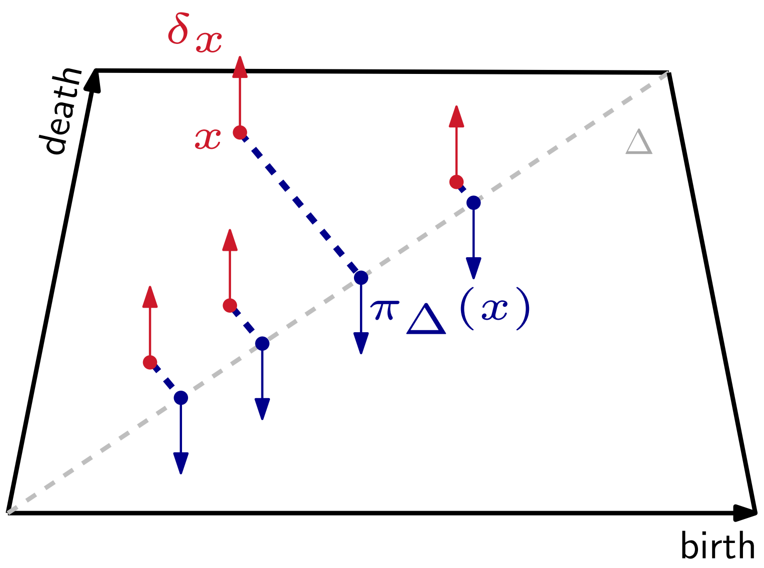
Sliced Wasserstein kernel

Def: Given $\sigma > 0$, for any $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^2)$:

$$k_{SW}(\mu,\nu) := \exp\left(-\frac{SW_1(\mu,\nu)}{2\sigma^2}\right)$$

Corllary: [Kolouri, Zou, Rohde] (from SW cnsd and Berg's theorem)
k_{SW} is positive semidefinite.

$\rightarrow$ application to persistence diagrams:



$$D \mapsto \mu_D := \sum_{x\in D} \delta_x$$

$$\mapsto \tilde{\mu}_D := \mu_D - \pi_\Delta \# \mu_D$$

$$SW_1(D,D') := \int_{\theta\in\mathcal{S}^1} \|\pi_\theta\#\tilde{\mu}_D - \pi_\theta\#\tilde{\mu}_{D'}\|_K \, d\theta$$

$$k_{SW}(D,D') := \exp\left(-\frac{SW_1(D,D')}{2\sigma^2}\right)$$

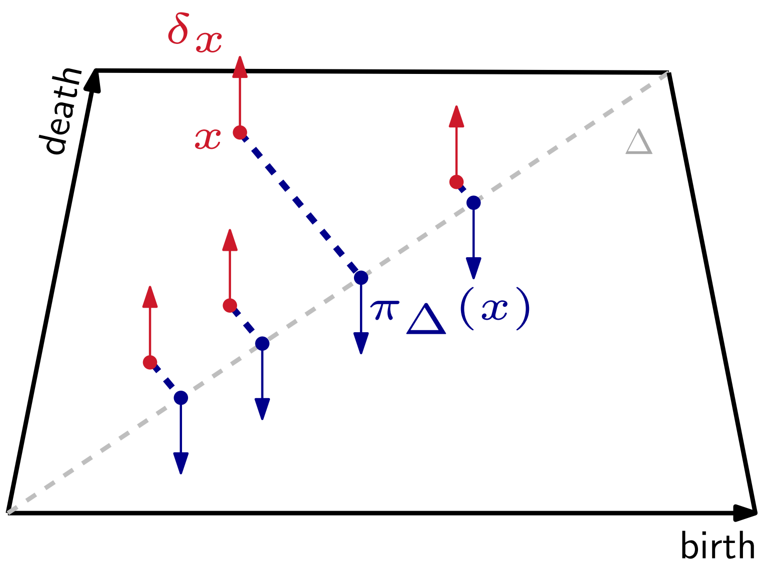
- positive semidefinite
- simple and fast to compute

Sliced Wasserstein kernel

Thm.: [C., Cuturi, Oudot]
The metrics d_1 and SW_1 on the space $\mathcal{D}_N$ of persistence diagrams of size bounded by N are strongly equivalent, namely: for $D, D' \in \mathcal{D}_N$,

$$\frac{1}{2+4N(2N-1)}\,\mathrm{d}_1(D,D') \;\leq\; SW_1(D,D') \;\leq\; 2\sqrt{2}\,\mathrm{d}_1(D,D')$$

$\rightarrow$ application to persistence diagrams:



$$D \mapsto \mu_D := \textstyle\sum_{x\in D} \delta_x$$

$$\mapsto \tilde{\mu}_D := \mu_D - \pi_\Delta \# \mu_D$$

$$SW_1(D,D') := \int_{\theta\in\mathcal{S}^1} \|\pi_\theta\#\tilde{\mu}_D - \pi_\theta\#\tilde{\mu}_{D'}\|_K \, d\theta$$

$$k_{SW}(D,D') := \exp\left(-\frac{SW_1(D,D')}{2\sigma^2}\right)$$

Sliced Wasserstein kernel

Thm.: [C., Cuturi, Oudot]
The metrics d_1 and SW_1 on the space $\mathcal{D}_N$ of persistence diagrams of size bounded by N are strongly equivalent, namely: for $D, D' \in \mathcal{D}_N$,

$$\frac{1}{2+4N(2N-1)}\,\mathrm{d}_1(D,D') \;\leq\; SW_1(D,D') \;\leq\; 2\sqrt{2}\,\mathrm{d}_1(D,D')$$

Corollary: the feature map ϕ associated with k_{SW} is weakly metric-preserving: $\exists g, h$ nonzero except at 0 such that $g \circ \mathrm{d}_1 \leq \|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq h \circ \mathrm{d}_1$.

Sliced Wasserstein kernel

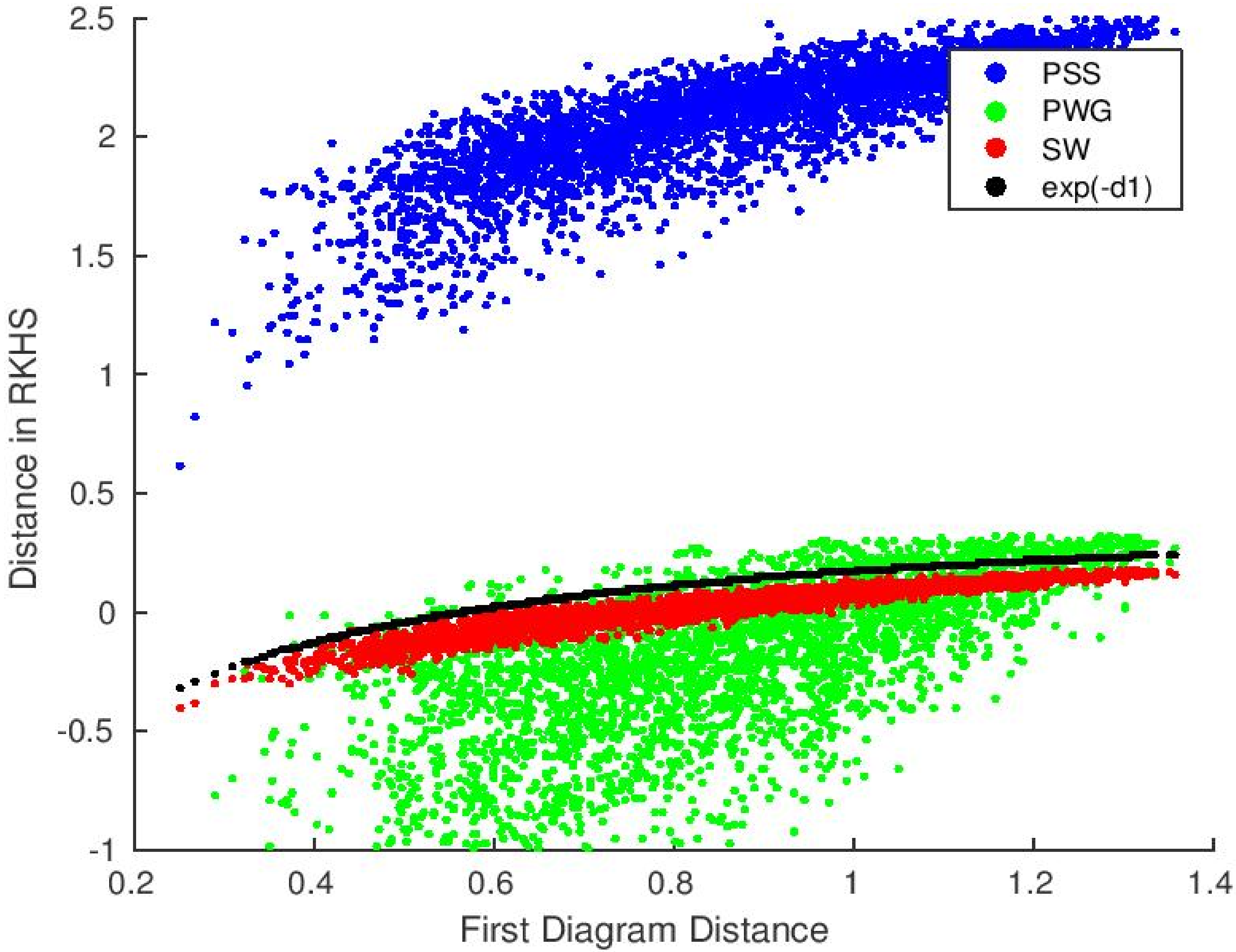
Thm.: [C., Cuturi, Oudot]
The metrics d_1 and SW_1 on the space $\mathcal{D}_N$ of persistence diagrams of size bounded by N are strongly equivalent, namely: for $D, D' \in \mathcal{D}_N$,

$$\frac{1}{2+4N(2N-1)}\,\mathrm{d}_1(D,D') \;\leq\; SW_1(D,D') \;\leq\; 2\sqrt{2}\,\mathrm{d}_1(D,D')$$

Corollary: the feature map ϕ associated with k_{SW} is weakly metric-preserving: $\exists g, h$ nonzero except at 0 such that $g \circ \mathrm{d}_1 \leq \|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq h \circ \mathrm{d}_1$.

Thm.: [C., Bauer]
Any feature map $\phi : \mathcal{D}_N \to \mathbb{R}^d$ *cannot* preserve metrics, whatever d and N are, i.e. either lower bound is 0 or upper bound is $+\infty$

Metric distortion in practice

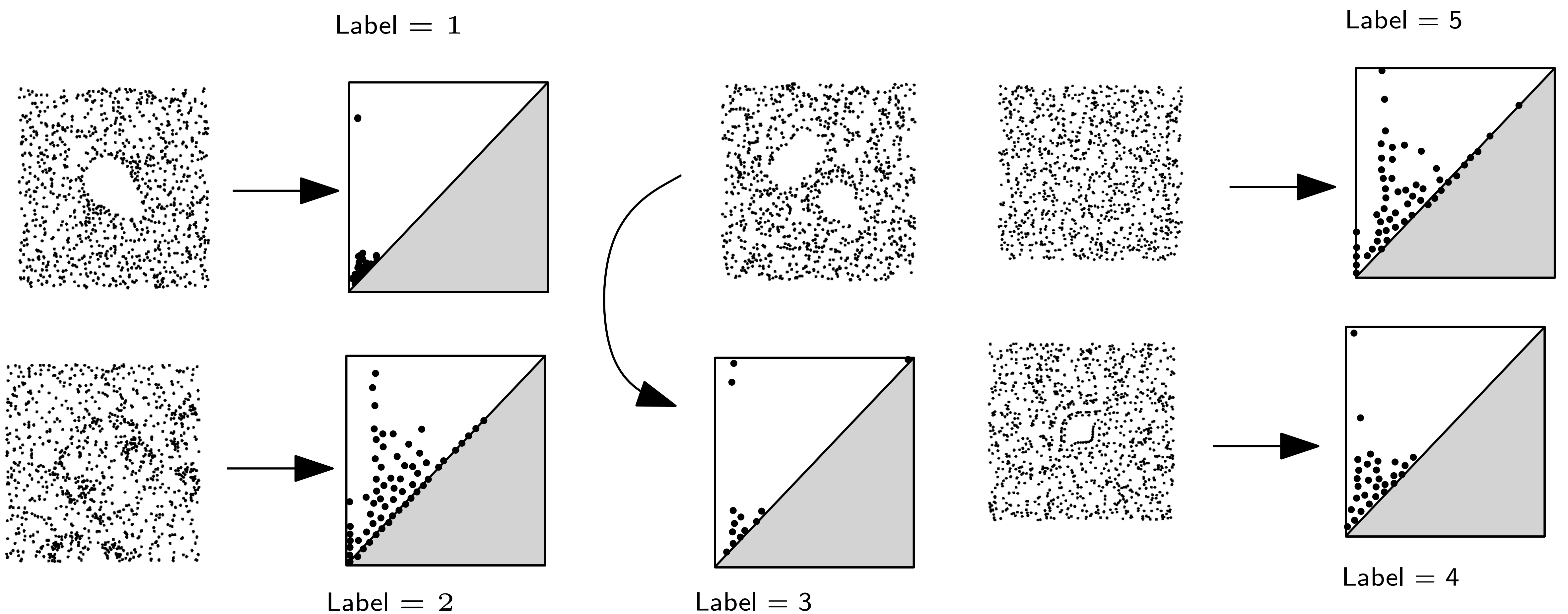


Application to supervised orbits classification

Goal: classify orbits of *linked twisted map*, modelling fluid flow dynamics

Orbits described by (depending on parameter r):

$$\begin{cases} x_{n+1} &= x_n + r\, y_n(1-y_n) \mod 1 \\ y_{n+1} &= y_n + r\, x_{n+1}(1-x_{n+1}) \mod 1 \end{cases}$$



Application to supervised orbits classification

Goal: classify orbits of *linked twisted map*, modelling fluid flow dynamics

Orbits described by (depending on parameter r):

$$\begin{cases} x_{n+1} = x_n + r\, y_n(1 - y_n) \mod 1 \\ y_{n+1} = y_n + r\, x_{n+1}(1 - x_{n+1}) \mod 1 \end{cases}$$

Accuracies (%) using only TDA descriptors (kernels on barcodes):

	k_{PSS}	k_{PWG}	k_{SW}
Orbit	64.0 ± 0.0	78.7 ± 0.0	$\mathbf{83.7} \pm 1.1$

(PDs as discrete measures)

Application to supervised orbits classification

Goal: classify orbits of *linked twisted map*, modelling fluid flow dynamics

Orbits described by (depending on parameter r):

$$\begin{cases} x_{n+1} = x_n + r\, y_n(1-y_n) \mod 1 \\ y_{n+1} = y_n + r\, x_{n+1}(1-x_{n+1}) \mod 1 \end{cases}$$

Accuracies (%) using only TDA descriptors (kernels on barcodes):

	k_{PSS}	k_{PWG}	k_{SW}
Orbit	64.0 ± 0.0	78.7 ± 0.0	$\mathbf{83.7} \pm 1.1$

(PDs as discrete measures)

Running times (in seconds on N-sized parameter space from 100 orbits):

	k_{PSS}	k_{PWG}	k_{SW}
Orbit	$N \times 9183.4 \pm 65.6$	$N \times 69.2 \pm 0.9$	$385.8 \pm 0.2 + NC$

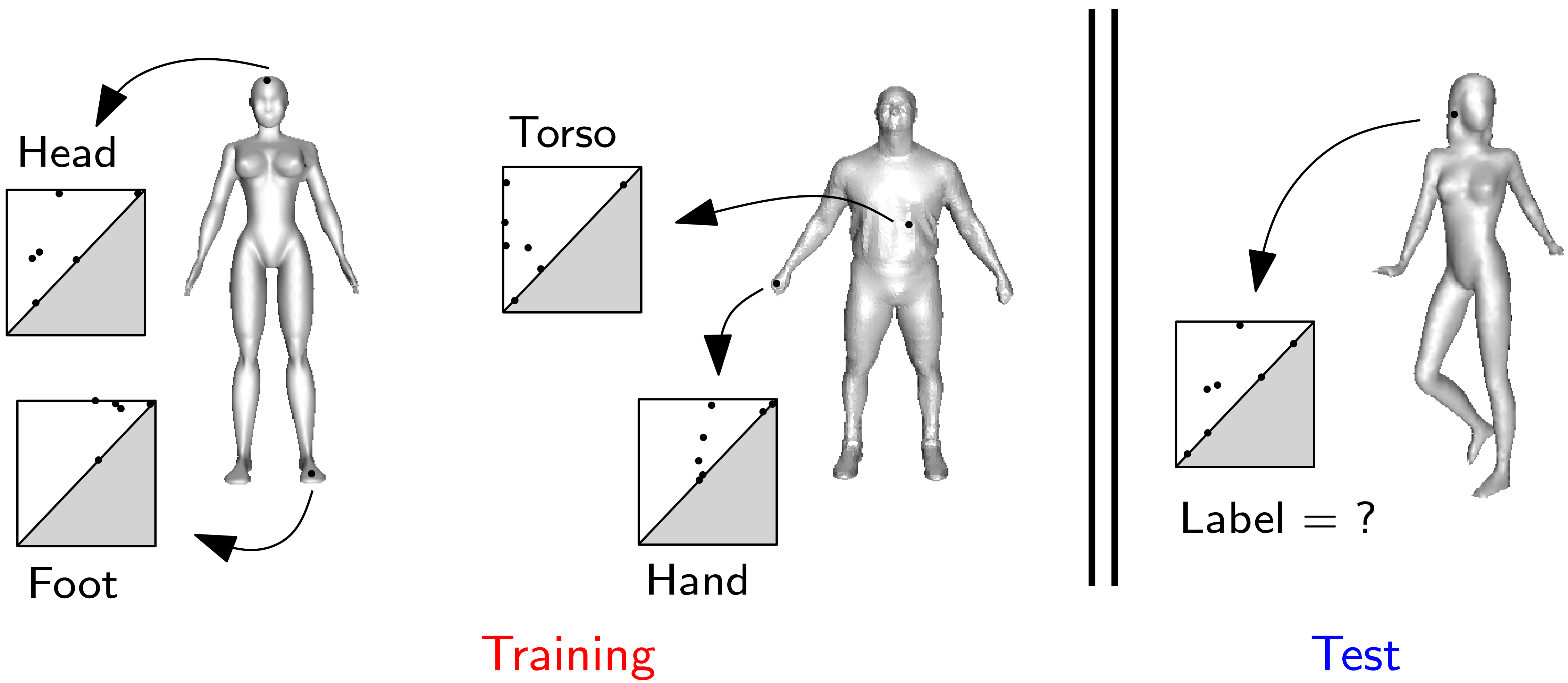
($\phi(\cdot)$ recomputed for each σ) (SW_1 computed only once)

Application to supervised shape segmentation

Goal: segment 3d shapes based on examples

Approach:

- train a (multiclass) classifier on PDs extracted from the training shapes
- apply classifier to PDs extracted from query shape



Application to supervised shape segmentation

Goal: segment 3d shapes based on examples

Approach:

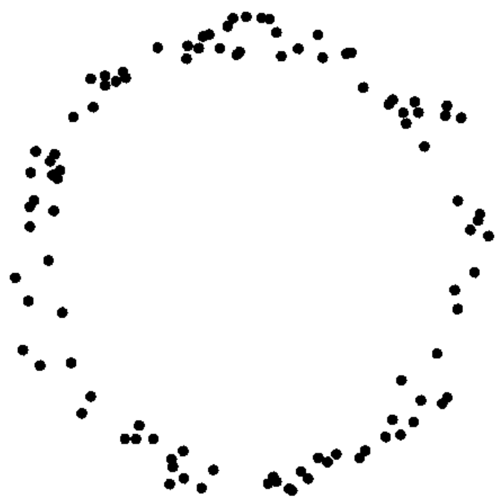
- train a (multiclass) classifier on PDs extracted from the training shapes
- apply classifier to PDs extracted from query shape

Accuracies (%) using TDA descriptors (kernels on barcodes):

	k_{PSS}	k_{PWG}	k_{SW}
Human	68.5 ± 2.0	64.2 ± 1.2	$\mathbf{74.0} \pm 0.2$
Airplane	55.4 ± 2.4	61.3 ± 2.9	$\mathbf{72.6} \pm 0.2$
Ant	86.3 ± 1.0	87.4 ± 0.5	$\mathbf{92.3} \pm 0.2$
FourLeg	67.0 ± 2.5	64.0 ± 0.6	$\mathbf{73.0} \pm 0.4$
Octopus	77.6 ± 1.0	78.6 ± 1.3	$\mathbf{85.2} \pm 0.5$
Bird	67.6 ± 1.8	$\mathbf{72.0} \pm 1.2$	67.0 ± 0.5
Fish	76.1 ± 1.6	$\mathbf{79.6} \pm 0.5$	75.0 ± 0.4

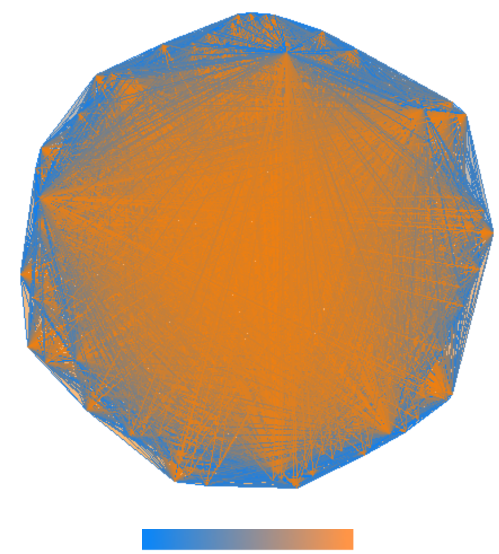
Recap'

input data



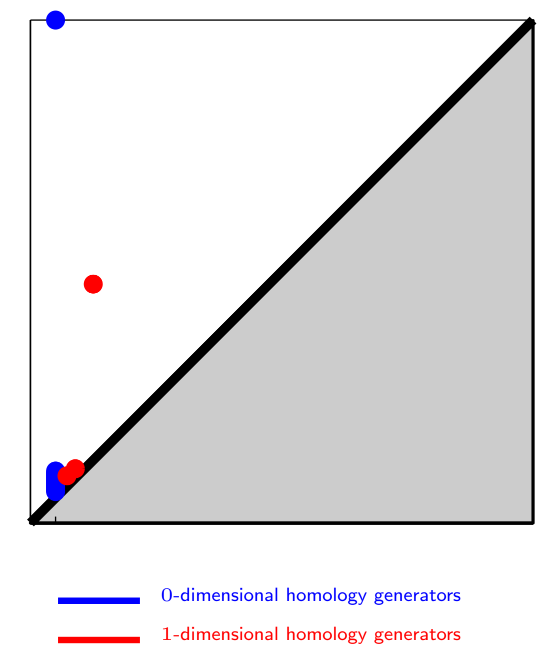
$\Longrightarrow$

domain / filter



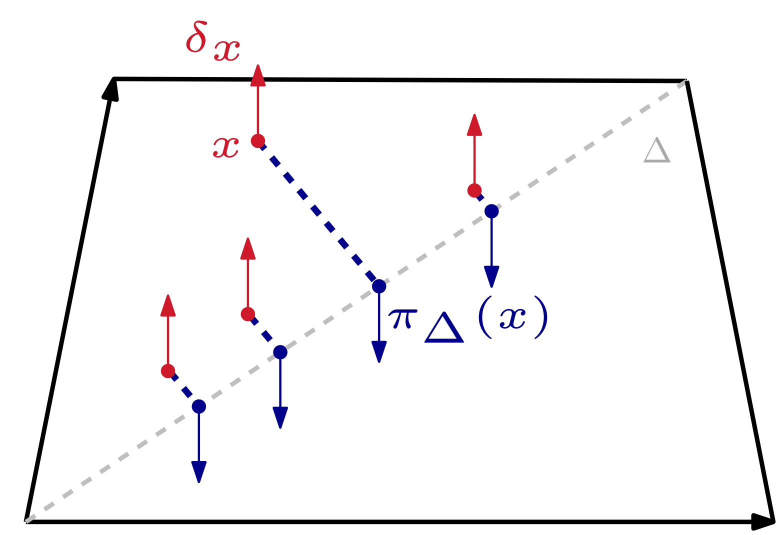
$\Longrightarrow$

persistence diagram



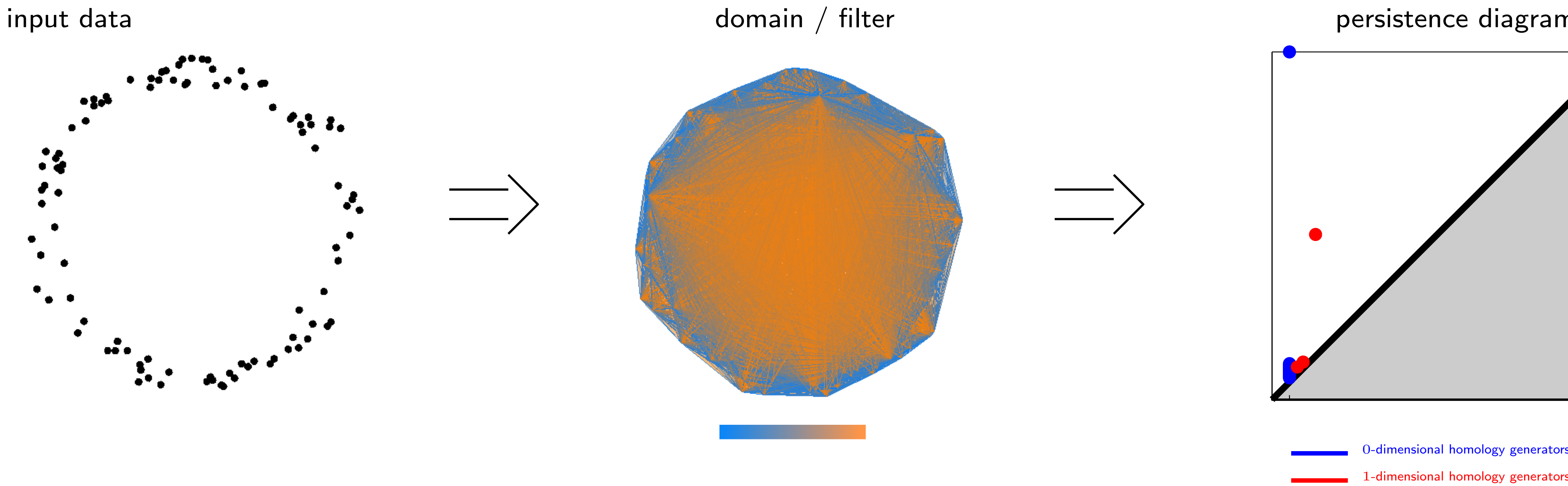
- kernels for persistence diagrams:
  - stable
  - additive, universal, etc.
  - easy to compute
  - distance-preserving (approximately)
  - (approximate) Gaussian kernel

$\Downarrow$



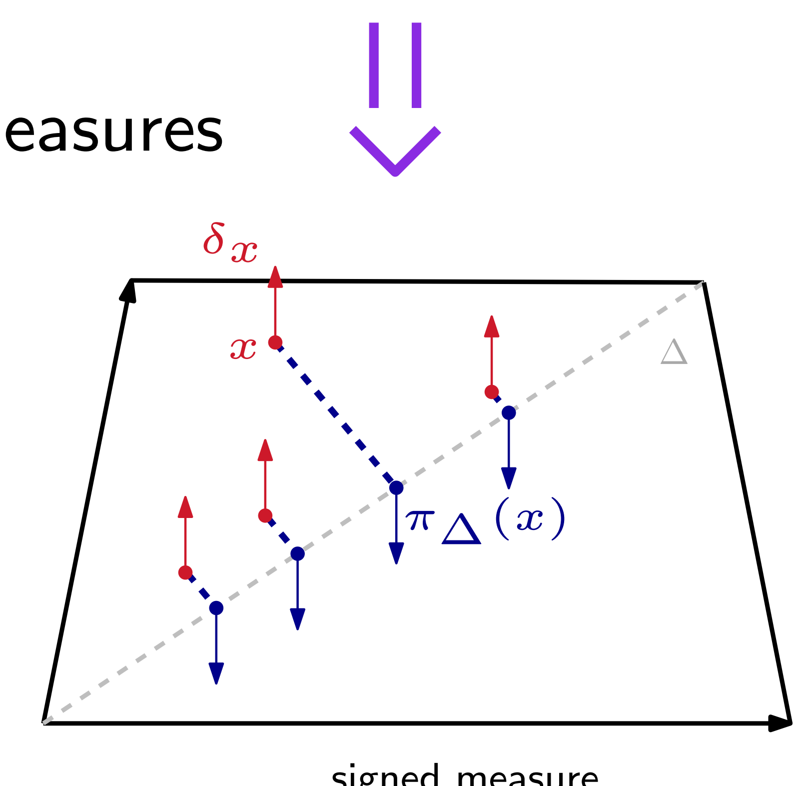
signed measure

Recap'



- theoretical framework:
  - quasi-isometric embedding of diagrams as signed measures
  - Wasserstein metric relaxation (sliced Wasserstein)

    $\rightarrow$ cnsd



Thank you!!