パーシステントホモロジーとその応用 (2)

2024-03-24
純粋数学者のためのデータサイエンス入門
九州大学 マス・フォア・インダストリ研究所
池 祐一

目次

1回目
1. 位相的データ解析の考え方とパーシステントホモロジー
2. 位相的データ解析の応用例 パート1（データ解析）

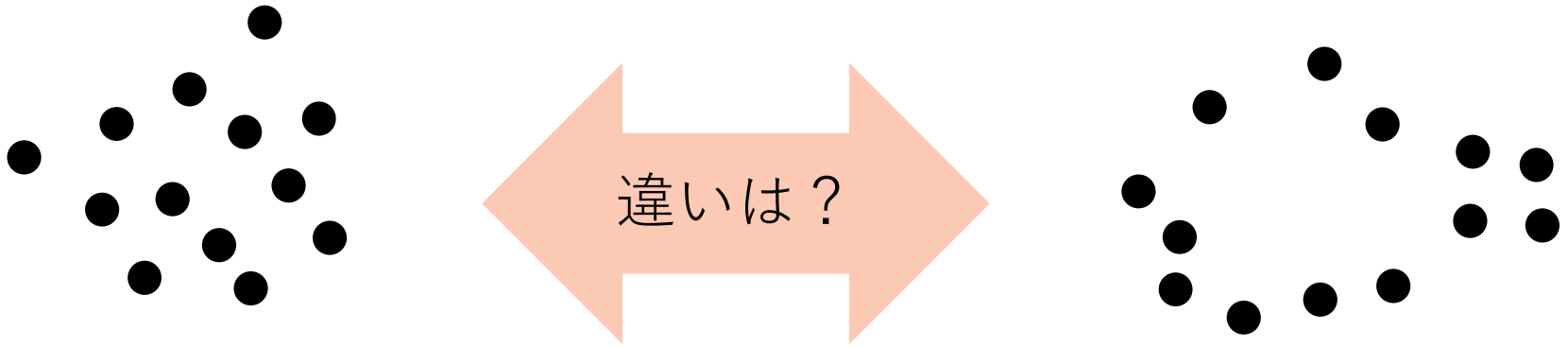
2回目
1. 機械学習について少しだけ
2. 位相的データ解析の応用例 パート2（機械学習との組み合わせ）

前回の復習

パーシステントホモロジー，応用例

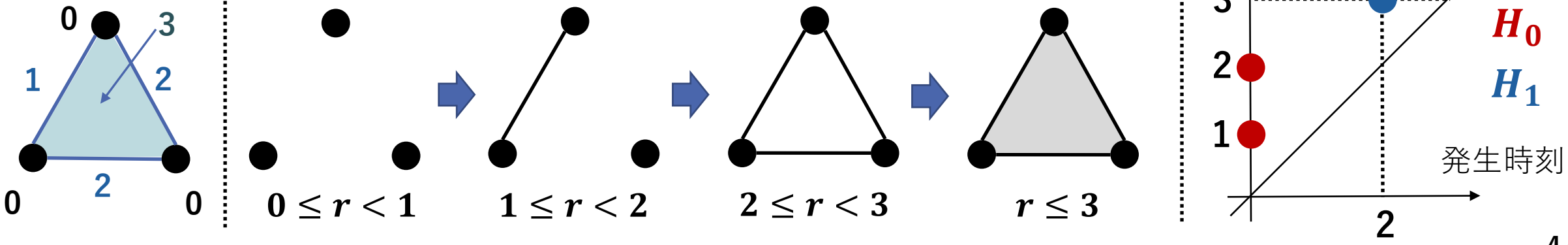
位相的データ解析とは（復習）

■**データの「トポロジー」の情報**を取り出して使う手法



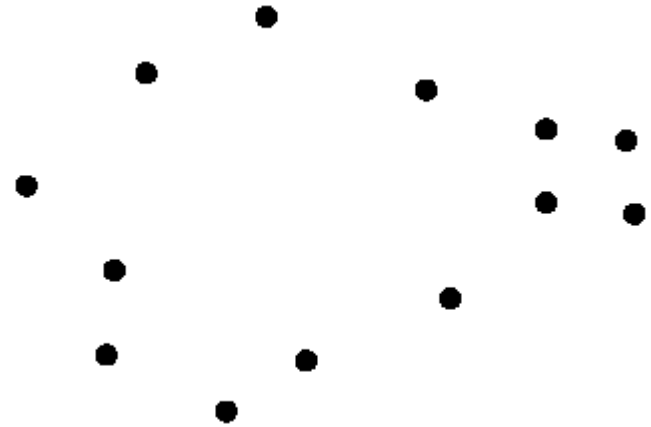
■単体的複体のフィルトレーションを作り，各ホモロジー類の
発生時刻・消滅時刻をプロットして**パーシステンス図（PD）**を作る

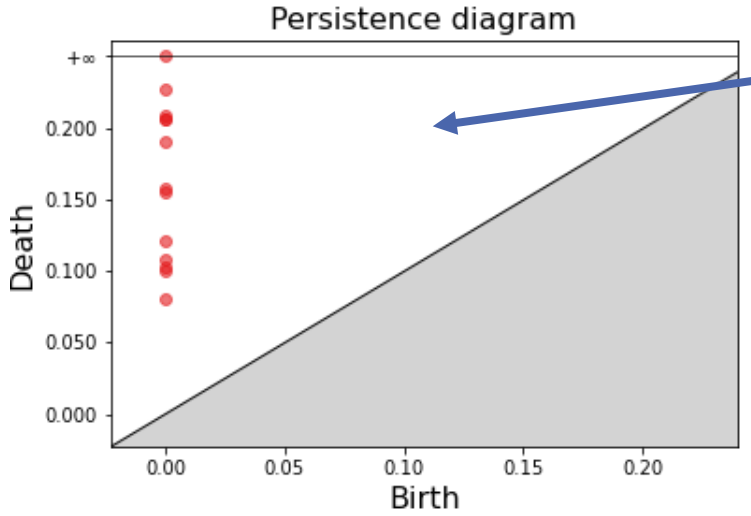
■PDからデータの「トポロジー」を理解できる



パーシステンス図のうれしさ（復習）

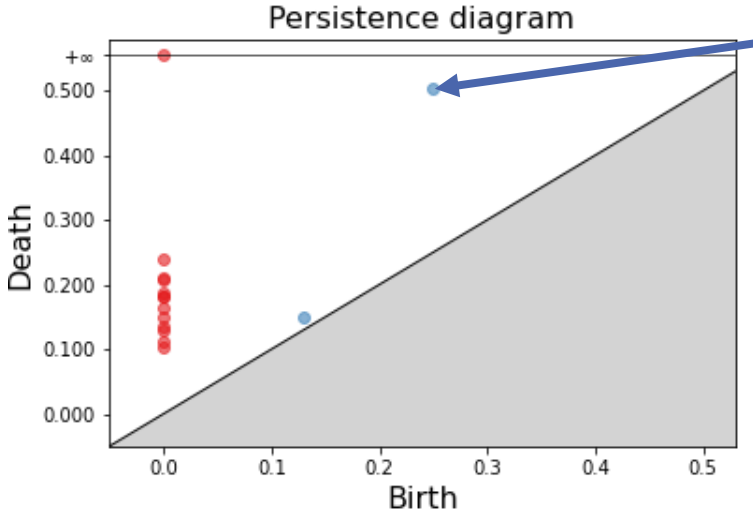
「トポロジー」が異なるデータをPDで区別できる





穴がない！

H_0



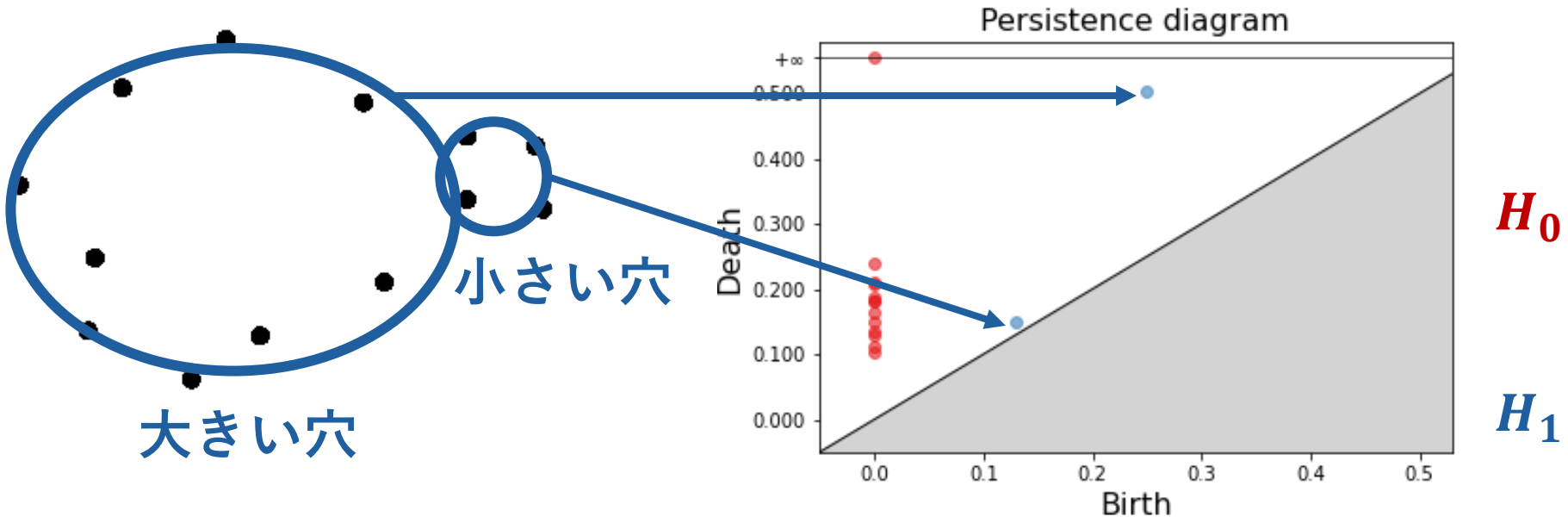
穴がある！

H_1

パーシステンス図のうれしさ（復習）

各点が，**データのどのような「トポロジー」を意味するかわかる**

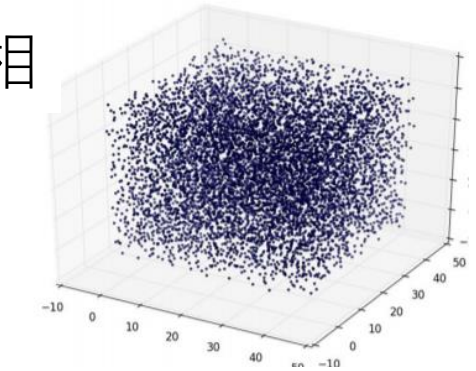
⇒データの新しい**記述子の役割**



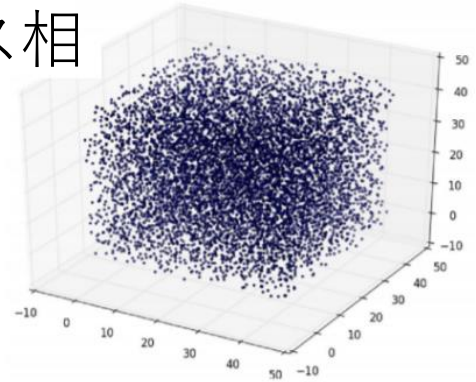
物質科学への応用（復習）

SiO_2の液相とガラス相の違いをPDを使って理解

液相

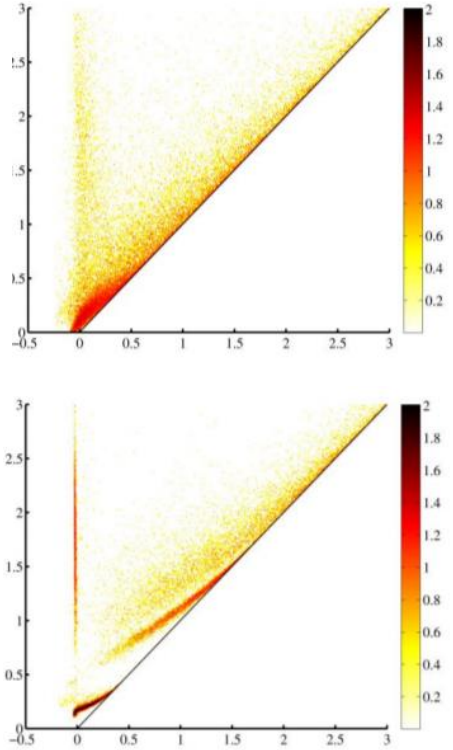


ガラス相



PD

H_1



※点が多いのでPDを密度で表示

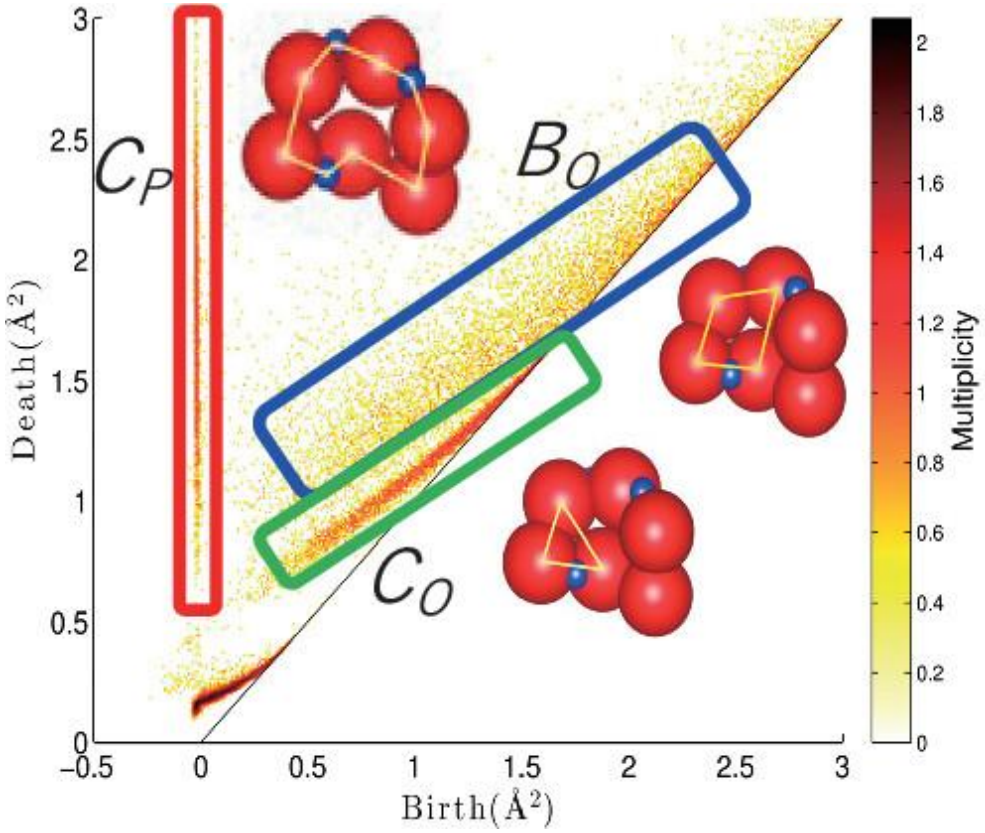
※SiとOで球の半径を変えて計算

パーシステンス図に変換すると差が明らかに見える

Figures from K. Fukumizu, Persistence Weighted Gaussian Kernel for Topological Data Analysis

物質科学への応用（復習）

SiO_2のガラス相のPDに特徴的な曲線が原子配置点群の**どの形に対応するかを出力可能（逆解析）**

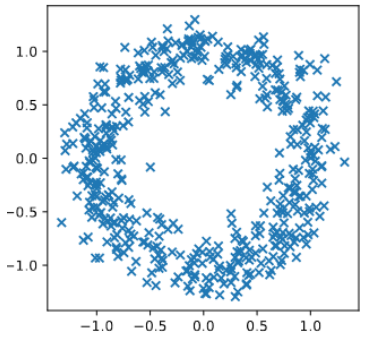


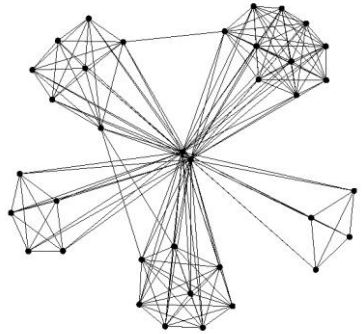
PD内の曲線C_O, B_O, C_Pが**中距離秩序**に対応すると信じられている

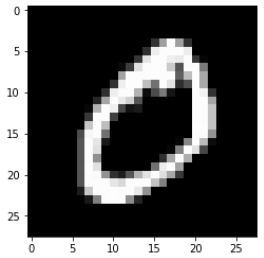
Figure from Obayashi et al., Persistent homology analysis for materials research and persistent homology software: Homcloud

TDAの典型的な使い方（復習）

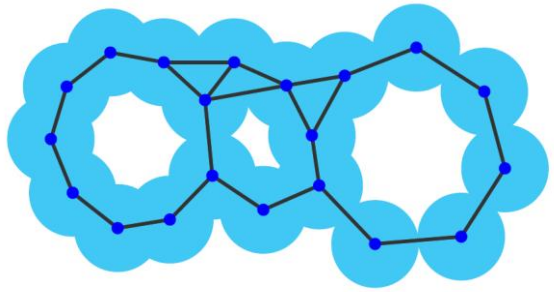
データ





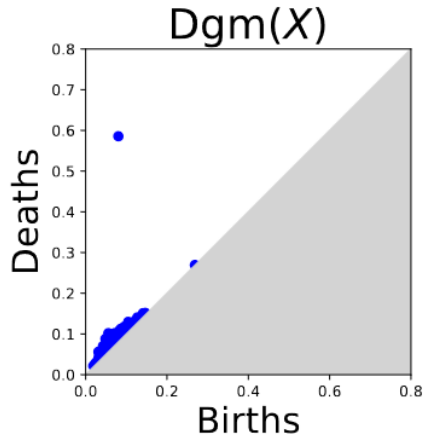


フィルトレーション

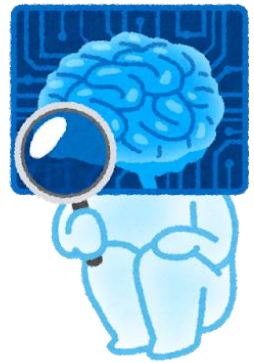


Software
Ripser, GUDHI,
HomCloud, …

パーシステンス図



専門家



機械学習

今日はまずここから

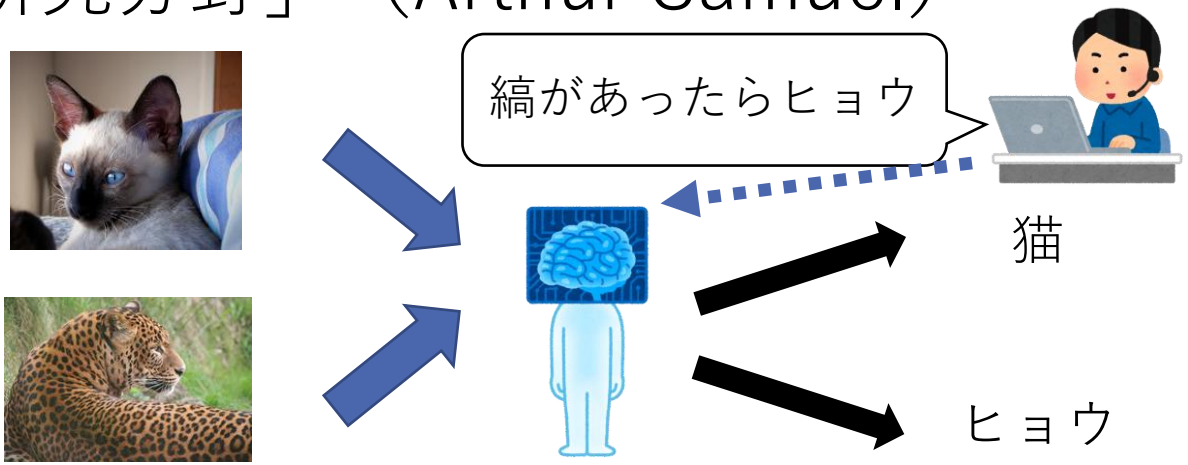
機械学習について少しだけ

機械学習は何のために行うのか，教師あり機械学習

機械学習とは

■人間と同等の情報処理を計算機で実現するための手法

■「計算機に処理方法を明示的にプログラミングせずに，学習する能力を与える研究分野」（Arthur Samuel）



明示的にプログラミング

機械学習

処理方法を自分で会得

教師あり機械学習：設定

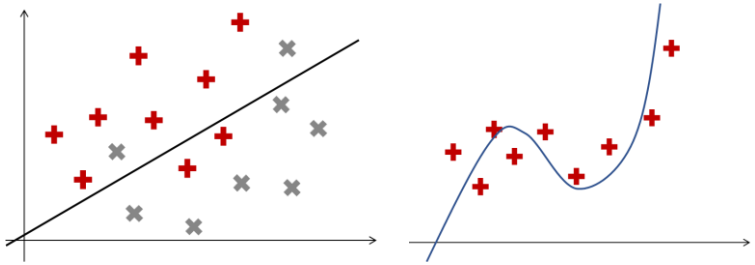
教師あり機械学習：答えつきのデータが使える

例：

	数学	理科（答え）
Aさん	80	78
Bさん	40	61
Cさん	100	90
…	…	…

タスクの例

- ■**分類**：出力が有限個の中のどれか
  - ■手書き数字の画像から数字を判定するなど
- ■**回帰**：出力が連続値
  - ■数学の点数から理科の点数を予測するなど



タスクに応じて**良い写像を生成**することが目的

■**パラメータ付けられた写像族**$(f_\theta)_\theta$から写像を見つける

例：$f_{W,b}(x) = Wx + b$ 線形モデル，ニューラルネットワーク

教師あり学習：ニューラルネットワーク

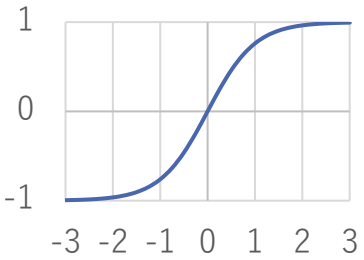
パラメータθが入った関数$f_\theta(x)$のひとつがニューラルネットワーク

■線形変換と非線形変換を繰り返すことで出来る関数

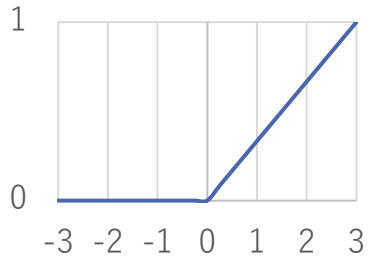
■線形変換：$x \mapsto Wx$

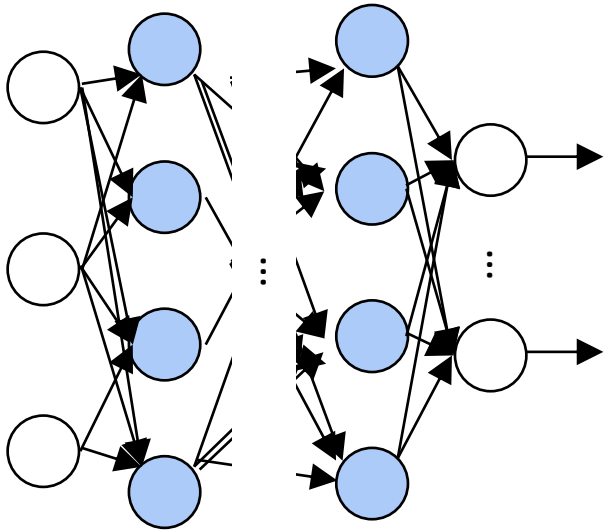
■非線形変換：シグモイド関数やReLU関数を使う

$$h(u) = \frac{1}{1+e^{-u}}$$



$$h(u) = \max\{u, 0\}$$





■ニューラルネットワーク

$x \mapsto W_1 x \mapsto h_1(W_1 x) \mapsto W_2 h_1(W_1 x) \mapsto \cdots \mapsto h_L(W_L \cdots W_2 h_1(W_1 x) \cdots)$

※上の図のようにあらわされることもある

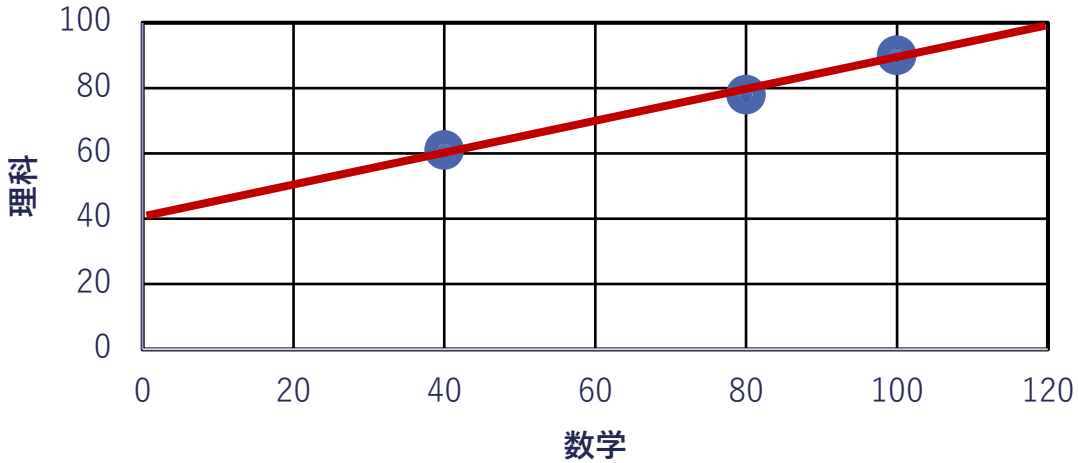
■パラメータは**行列$W_1, \ldots, W_L$の成分全部**．最近は数が5000億以上のことも

教師あり学習：訓練

Q. パラメータはどうやって設定するか？

■線形回帰では$\sum_i \left(y_i - (ax_i + b)\right)^2$を最小にするように決定（**最小二乗法**）

	数学	理科（答え）
Aさん	80	78
Bさん	40	61
Cさん	100	90
…	…	…



A. 答えに対する当てはまりの悪さを測る量をなるべく小さくする

■$f_\theta(x)$のyへの当てはまりの悪さをあらわす**損失関数** $l(y, f_\theta(x))$を設計

■以下の訓練誤差をなるべく小さくするようにθを決定：

$$\mathcal{L}(\theta) \coloneqq \frac{1}{n}\sum_i l\left(y_i, f_\theta(x_i)\right)$$

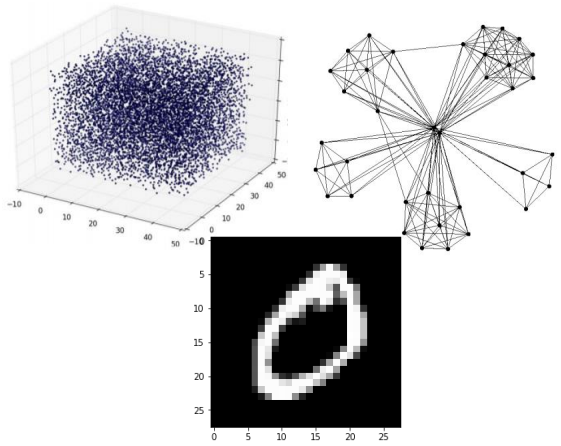
■$\mathcal{L}$が微分可能なら**勾配降下法**を使って小さくする: $\theta_{k+1} = \theta_k - \alpha_k \nabla \mathcal{L}(\theta_k)$

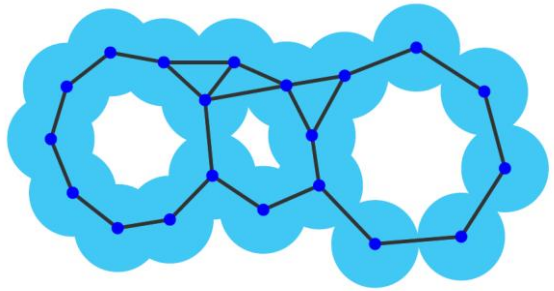
位相的データ解析の応用例 パート2

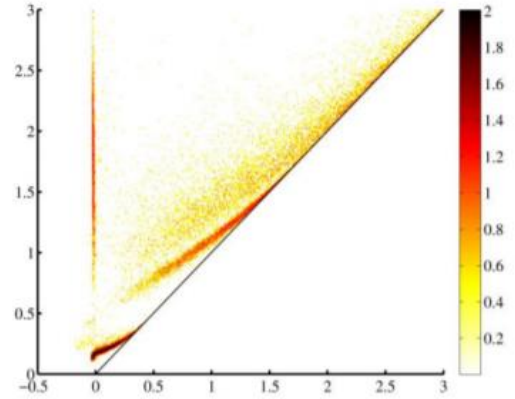
機械学習との組み合わせ，記述子以外の使い方

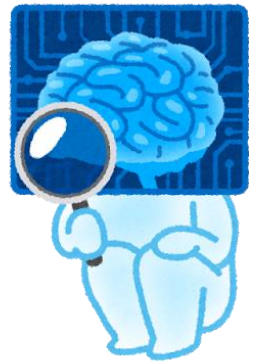
PHを機械学習の入力とするパイプライン

データ　　フィルトレーション　　パーシステンス図　　機械学習









例
- シミュレーションで生成された点群
- センサーによる振動
- グラフ・画像

Software
Ripser, GUDHI, HomCloud, …

機械学習に入力するためにベクトル化

Figures from K. Fukumizu, Persistence Weighted Gaussian Kernel for Topological Data Analysis

パーシステンス図のベクトル化 1

パーシステンス図の**ベクトル化手法**が多く提案されている

■PHと機械学習を組み合わせる際に有用

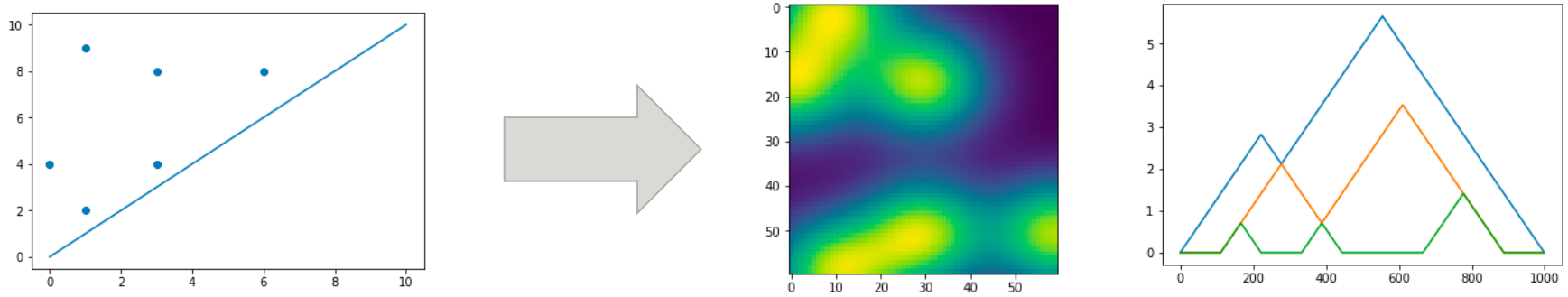
■**パーシステンスイメージ** (PI) : ガウス関数によるカーネル密度推定：

$$\rho(D)(z) \coloneqq \sum_{u \in D} w(u) \cdot \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|z-u\|^2}{2\sigma^2}\right), \qquad w\colon \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$$

領域を分割してそこで$\rho(D)$を積分することで有限次元ベクトルを得る

■**パーシステンスランドスケープ** (PL) : 三角状の関数を用いて定義

$$\lambda_k(D)(t) \coloneqq k - \max \min\{t - b_i, d_i - t\}_+, \qquad D = \{(b_i, d_i)\}_i$$



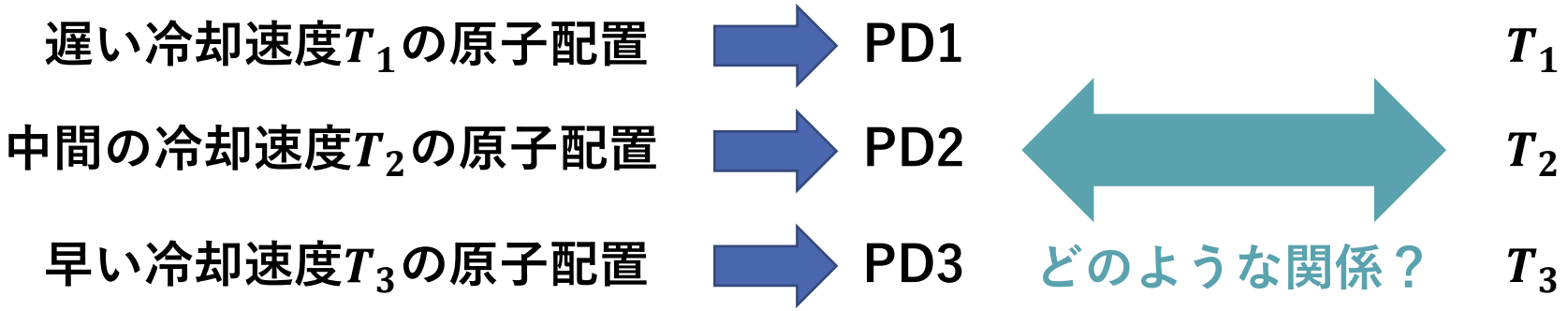
■他にも**Weighted Gaussian kernel**などいろいろある

金属ガラスの解析（再）

Hirata et al., Structural changes during glass formation extracted by computational homology with machine learning, 2020

Pd 80%とSi 20%からなる金属ガラスの解析

- ■高速冷却のシミュレーションで原子配置を用意
- ■複数の冷却速度でデータを生成
- ■原子配置のPDと冷却速度の関係を調べた（ベクトル化+線形回帰）

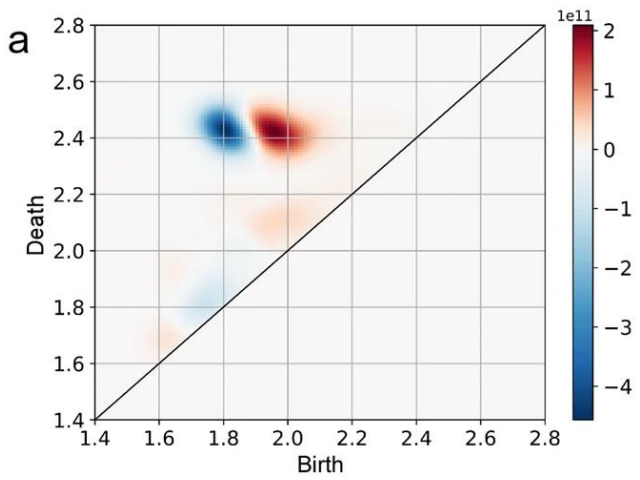


1. PDを**パーシステンスイメージ**に変換する
2. 冷却温度とパーシステンスイメージに対して**線形回帰**を行う

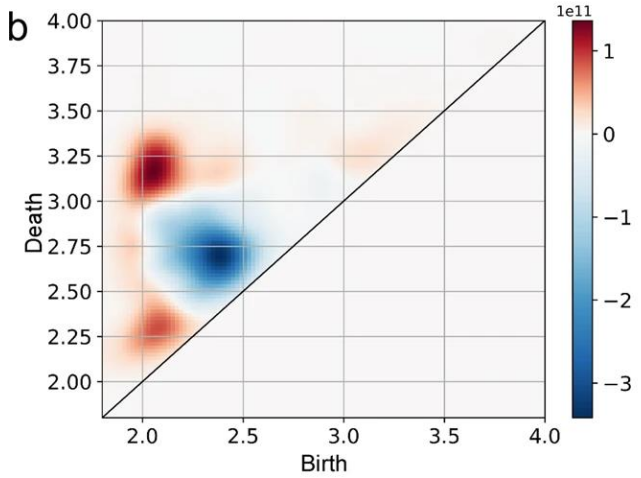
金属ガラスの解析（再）

1. PDを**パーシステンスイメージ**に変換する
2. 冷却温度とパーシステンスイメージに対して**線形回帰**を行う

Pd原子配置の2次元の穴に関するPD



Si原子配置の1次元の穴に関するPD



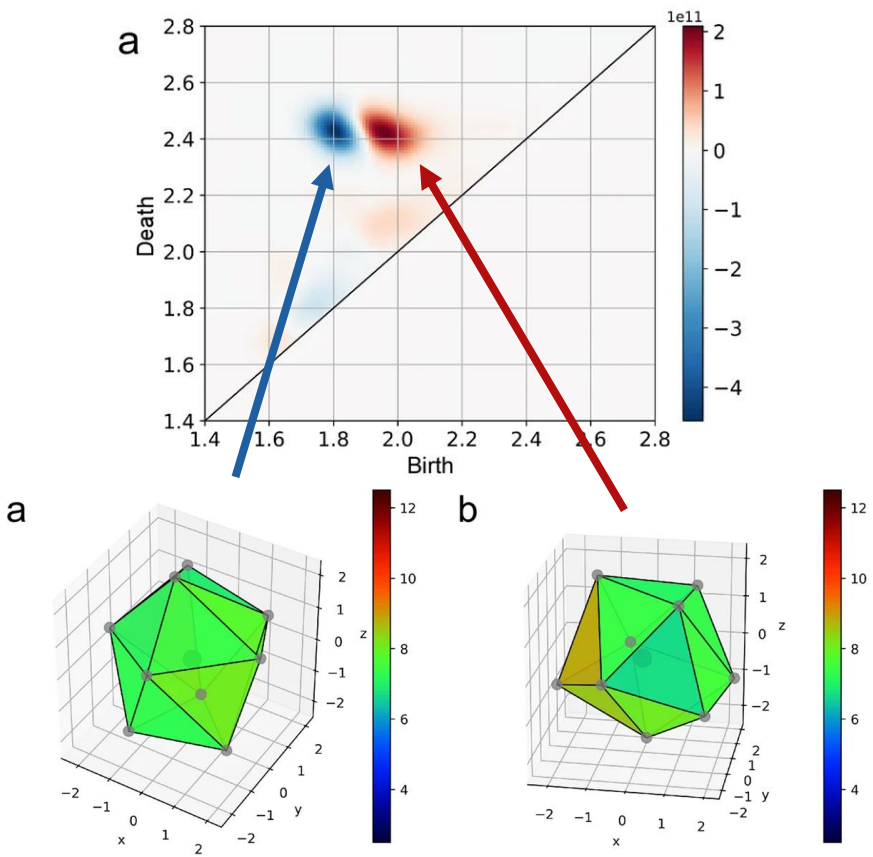
回帰係数が正か負かで，その範囲のPDの点が冷却温度に
正に効くか負に効くか分かる

■冷却温度が早いほど，赤い領域の点が多く，青い領域の点が少ない

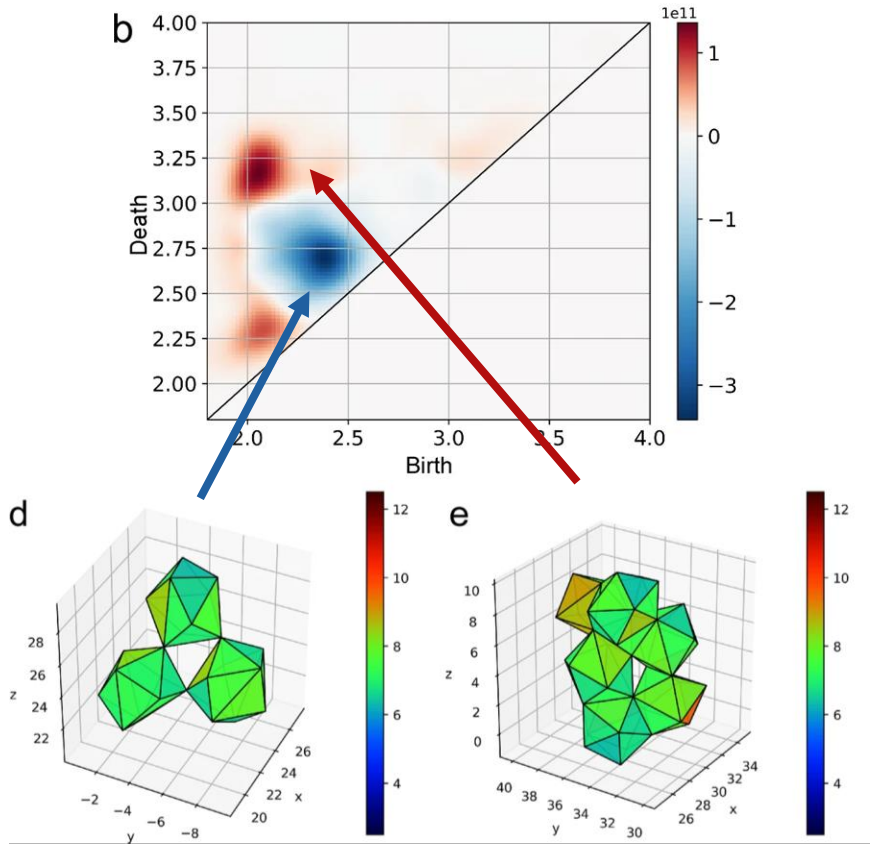
金属ガラスの解析（再）

逆解析により具体的にどの原子配置の構造が冷却速度に関係するかもわかる

Pd原子配置の2次元の穴に関するPD



Si原子配置の1次元の穴に関するPD



パーシステンス図のベクトル化 2

PDのベクトル化は**機械学習で学習する**ことも可能

■**Hofer et al., Deep Learning with Topological Signatures, 2017**

■ ガウス関数の値をPDの各点にわたって和をとりベクトル化

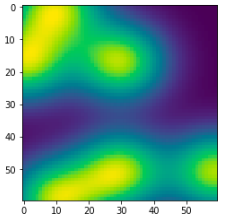
■ **中心と分散のパラメータを**タスクに応じて**学習させる**

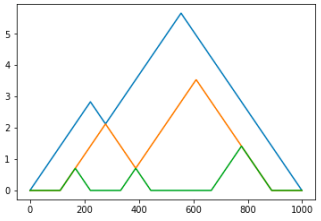
■**Carrière et al., PersLay, 2020:** DeepSets（点群に対するNN）に基づく

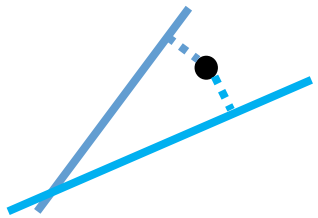
$$\mathrm{PersLay}(D) \coloneqq \mathrm{op}\left(\{w(p)\cdot\phi(p)\}_{p\in D}\right)$$

■ $\phi\colon \mathbb{R}^2 \to \mathbb{R}^q$ 点変換写像，たとえば

1. $\phi_\Gamma\colon p \mapsto \left[\exp\left(-\frac{|p-\boldsymbol{c_k}|^2}{2\sigma^2}\right)\right]_{k=1}^{q}$ ガウス関数
2. $\phi_\Lambda\colon p \mapsto [\max\{0, y-|\boldsymbol{t_k}-x|\}]_{k=1}^{q}$ 三角形の関数
3. $\phi_L\colon p \mapsto [\langle p, \boldsymbol{e_k}\rangle + \boldsymbol{b_k}]_{k=1}^{q}$ 直線からの符号付き距離







■ $w\colon \mathbb{R}^2 \to \mathbb{R}$ 重み関数，

■ op: 任意の置換不変な操作（和・最大・最小・k番目に大きい値など）

■ タスクに応じて**ϕが持つパラメータを学習させる**

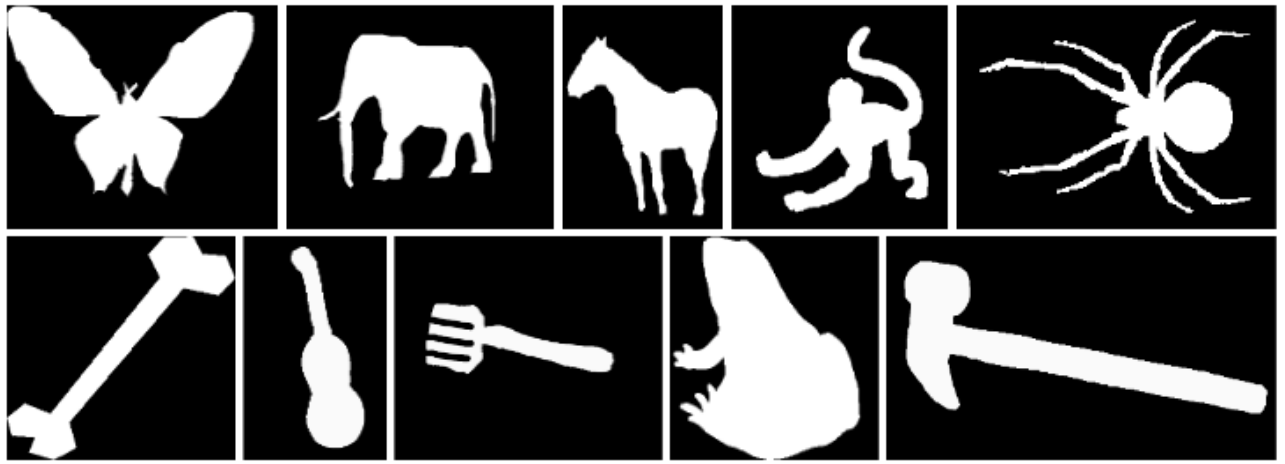
2次元画像分類への応用

Hofer et al., Deep Learning with Topological Signatures, 2017

■白黒画像の分類に**PHとベクトル化学習**を用いる

Animal
20クラス

MPEG-7
70クラス

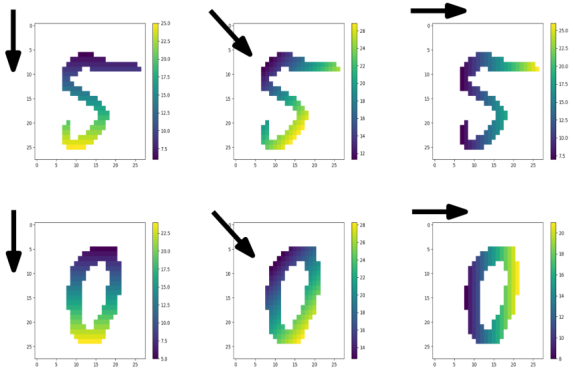


	MPEG-7	Animal
‡Skeleton paths	86.7	67.9
‡Class segment sets	90.9	69.7
†ICS	96.6	78.4
†BCF	97.2	83.4
Ours	91.8	69.5

■**白い画素の格子点が頂点**と見て，それらを**縦横でつなぎ単体複体**を作る

■円周上に32個等角度に点を取り，それらの方向に増加する**1次関数の劣位集合フィルトレーション**のPHを計算

■出来た32個の0次PDをベクトル化してNNに入力



0次PDのみでも表現を学習することで良い精度

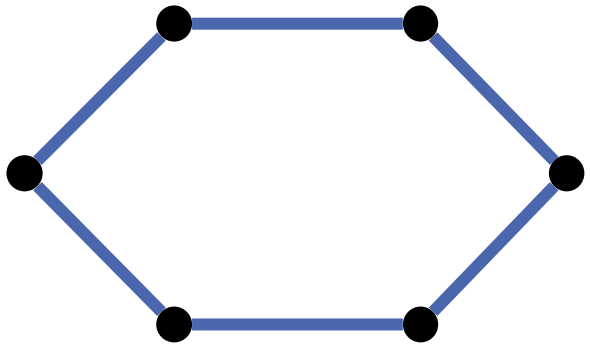
グラフ分類への応用

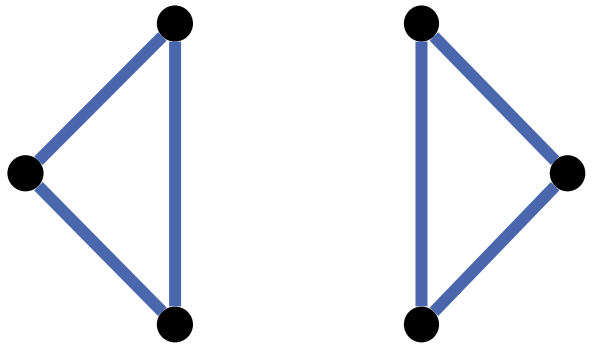
たくさんのグラフを分類したい（例：タンパク質分類）

PHを使うと近傍情報だけでなく全体的な構造を分類に使える

すべての点につながっている辺は2本

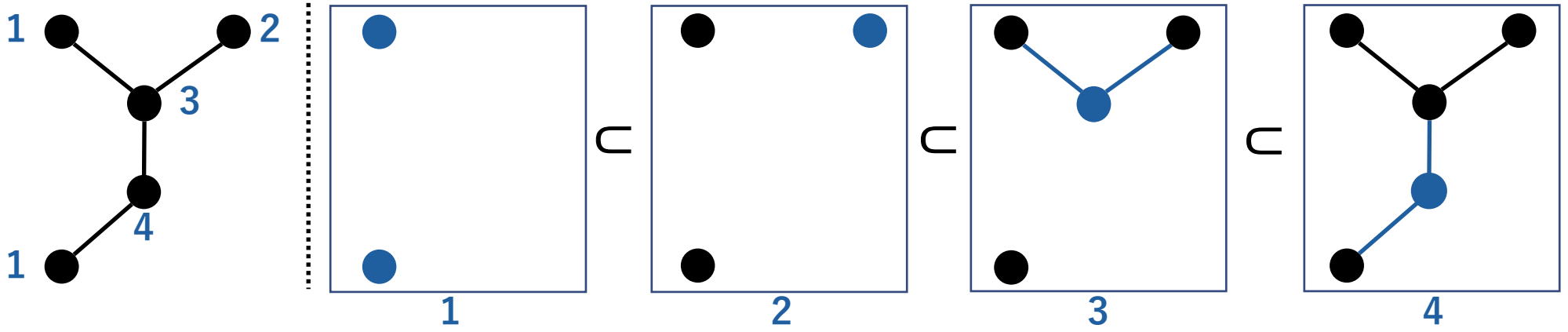
連結成分は1つ
ループは1つ





連結成分は2つ
ループは2つ

頂点上の関数を与え**劣位集合フィルトレーションのPH**を考える



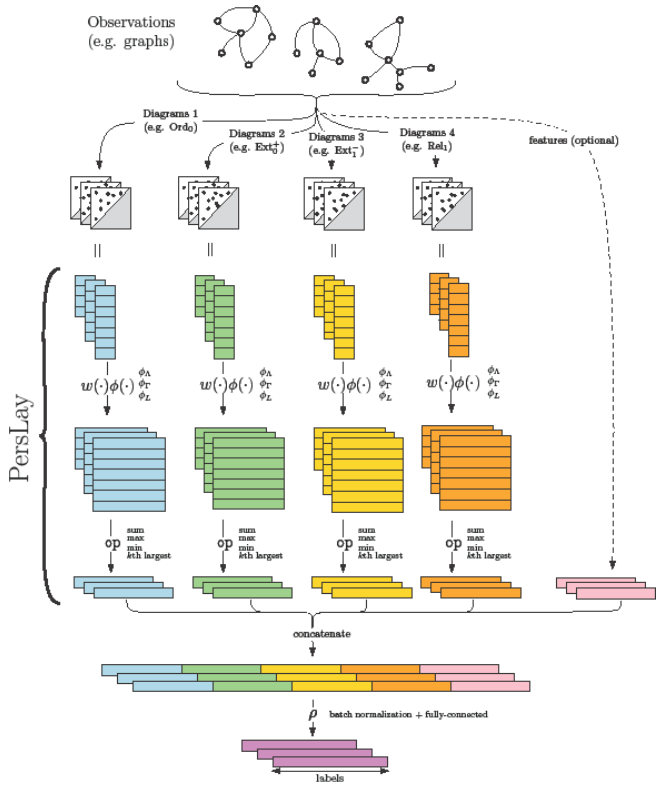
グラフ分類への応用

Carrière, Chazal, I. Lacombe, Royer, Umeda, PersLay, 2020

単純なネットワークにベクトル化学習したPDを入力

■頂点上の関数はグラフラプラシアンの固有関数を用いて構成

⇝PDしか使わずに（当時）最先端の分類器と同等の精度を達成



Dataset	SV[1]	RetGK* [2]	FGSD [3]	GCNN [4]	GIN [5]	PersLay Mean	PersLay Max
REDDIT5K	—	56.1	47.8	52.9	57.0	55.6	56.5
REDDIT12K	—	48.7	—	46.6	—	47.7	49.1
COLLAB	—	81.0	80.0	79.6	80.1	76.4	78.0
IMDB-B	72.9	71.9	73.6	73.1	74.3	71.2	72.6
IMDB-M	50.3	47.7	52.4	50.3	52.1	48.8	52.2
COX2*	78.4	80.1	—	—	—	80.9	81.6
DHFR*	78.4	81.5	—	—	—	80.3	80.9
MUTAG*	88.3	90.3	92.1	86.7	89.0	89.8	91.5
PROTEINS*	72.6	75.8	73.4	76.3	75.9	74.8	75.9
NCI1*	71.6	84.5	79.8	78.4	82.7	73.5	74.0
NCI109*	70.5	—	78.8	—	—	69.5	70.1

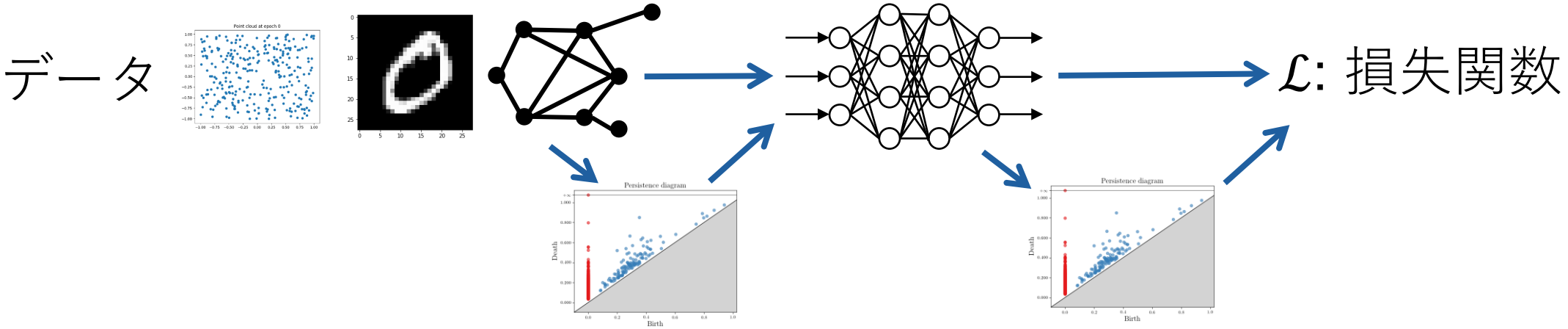
※10-foldの交差検証

トポロジー情報を上手く使うことは有効

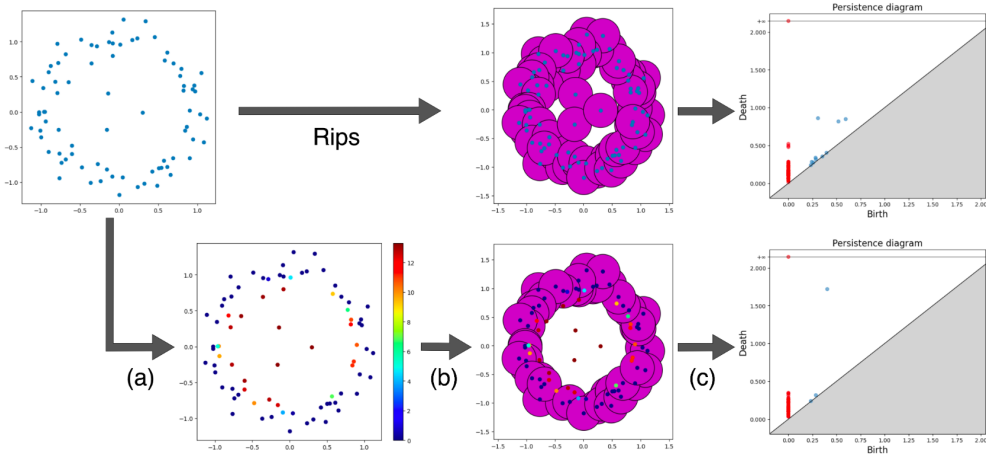
データの記述子以外のTDAの使い方

データの記述子としてPDを使う以外の使用例を2つ紹介

1. PHを損失関数に組み込んでトポロジー的に学習する

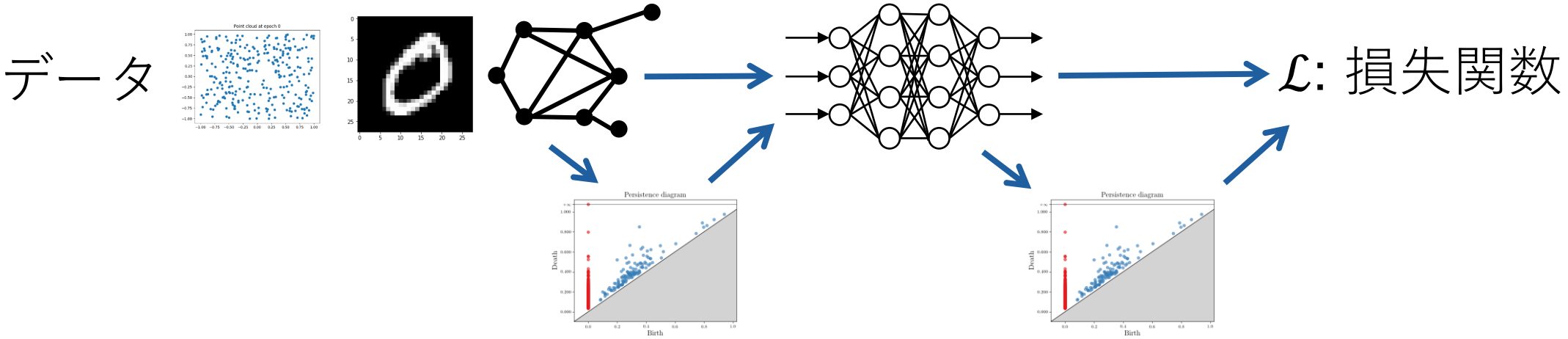


2. フィルトレーションをタスクに合わせて学習する

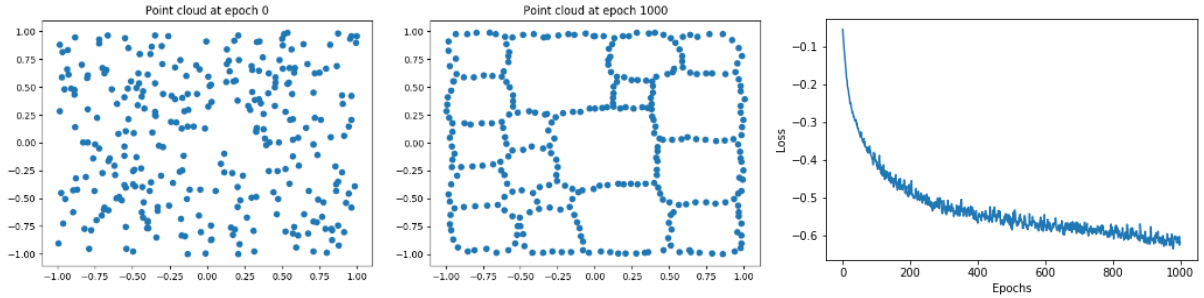


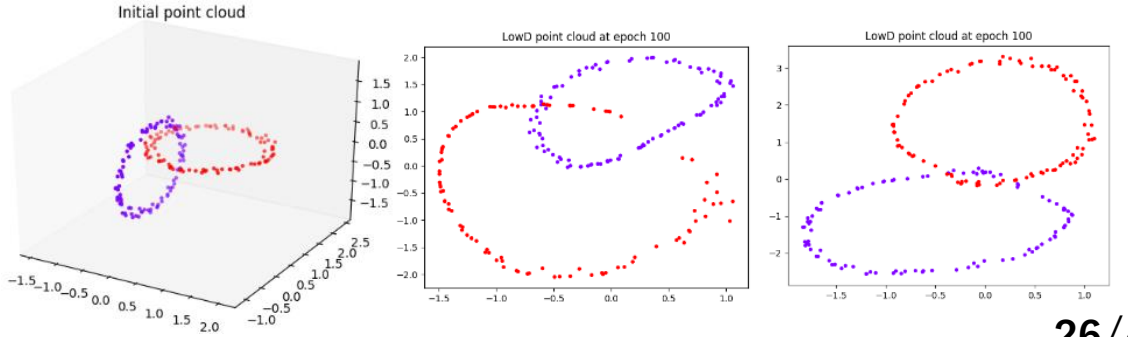
PHと機械学習：トポロジー的学習

近年は**PH的損失関数**を使ってデータやネットワークをコントロールする研究が多くなってきている



- **A Topology Layer for Machine Learning, 2020**: 点群の変形に応用
- **Topological Autoencoders, 2020**: 位相的性質も保つ次元圧縮



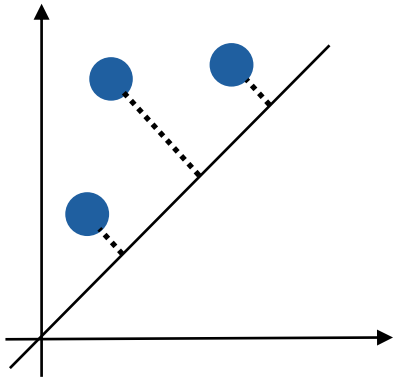


おもちゃの例 (1/2)

点群を**ループの個数が多くなる**ように変形する

■1次（ループに関する）のPD $D_1(X)$を考えて，

$$\mathcal{L}(X) = -\sum_{p \in D_1(X)} \|p - \pi_\Delta(p)\|_\infty^2 + d(X, C)$$



とする．ここでπ_Δは対角線への射影でCは正方形．

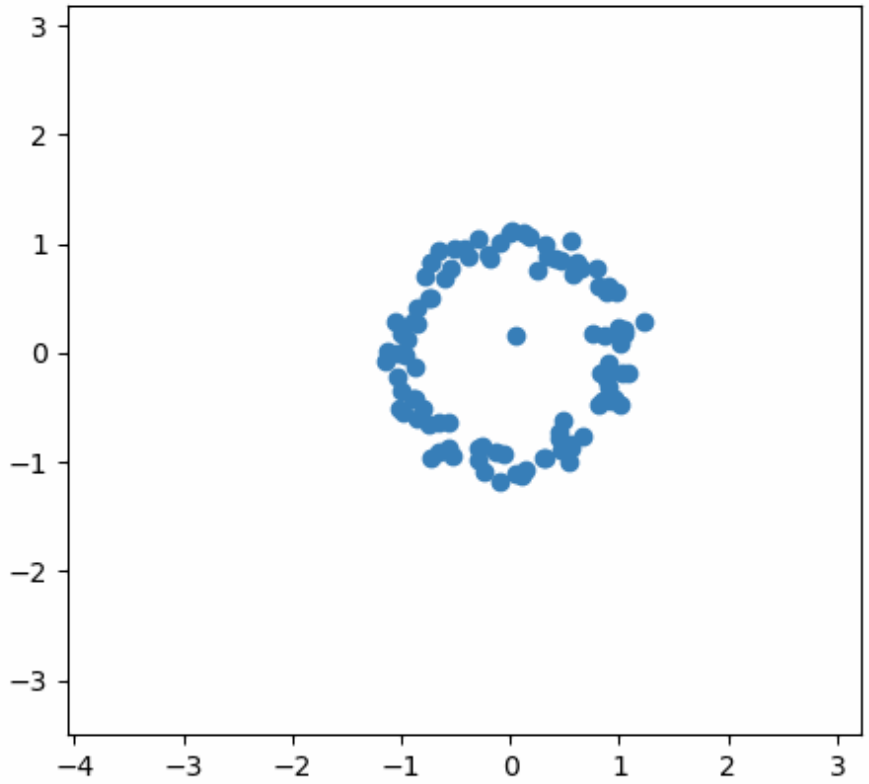
■**勾配降下法** $x_{k+1} = x_k - \alpha_k \nabla \mathcal{L}(x_k)$ により最適化

以下で見るように，パーシステンス図を与える写像は，ほとんど至るところで微分可能で，勾配降下法を使って最適化できる

おもちゃの例 (2/2)

点群を**ループの個数が多くなる**ように変形する

■$\mathcal{L}(X)$を勾配降下法で最適化



パラメータ付けられたフィルトレーション

復習：単体的複体 K のフィルトレーション

$\leftrightarrow$ 関数 $f\colon K \to \mathbb{R}$ s.t. $\sigma \subset \tau \Rightarrow f(\sigma) \leq f(\tau)$

$\leftrightarrow$ ベクトル $f \in \mathbb{R}^K$ s.t. $\sigma \subset \tau \Rightarrow f_\sigma \leq f_\tau$

$$\mathrm{Filt}_K := \{F \in \mathbb{R}^K \mid \sigma \subset \tau \Rightarrow f_\sigma \leq f_\tau\}$$

定義 $U \subset \mathbb{R}^d$で**パラメータ付けられたフィルトレーション**とは
関数$F\colon U \to \mathrm{Filt}_K$のこと

例

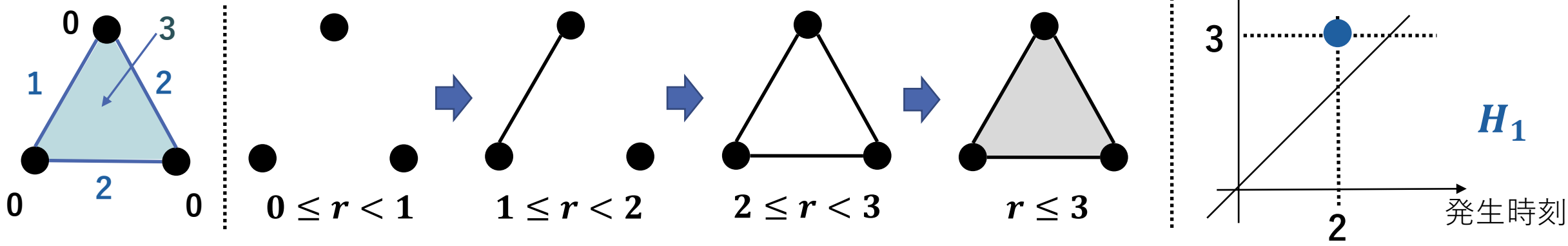
- **劣位集合フィルトレーション** $F\colon \mathbb{R}^{|V(K)|} \to \mathbb{R}^{|K|},\ F_\sigma(f) := \max_{i\in\sigma} f_i$
- **VRフィルトレーション** $F\colon (\mathbb{R}^d)^N \to \mathbb{R}^{|\Delta_N|},\ F_\sigma(x) := {}^1\!/_2 \max_{i,j\in\sigma} \|x_i - x_j\|$
- **MLのパラメータ** $f_\theta\colon \mathbb{R}^d \to \mathbb{R}^D\ (\theta \in \Theta), P \subset \mathbb{R}^d$: 有限集合

$$F\colon \Theta \to \mathbb{R}^{|\Delta_N|},\ F_\sigma(\theta) := {}^1\!/_2 \max_{i,j\in\sigma} \|f_\theta(x_i) - f_\theta(x_j)\|$$

パーシステンス図の対応の微分可能性

Q. パーシステンス図を与える写像を微分することはできるか？

A. 多くの点では（ほとんど至るところで）微分できる



1. **生成・消滅単体の組**$\{(\sigma_{b_i}, \sigma_{d_i})\}_i$を見つける（組合せ的）

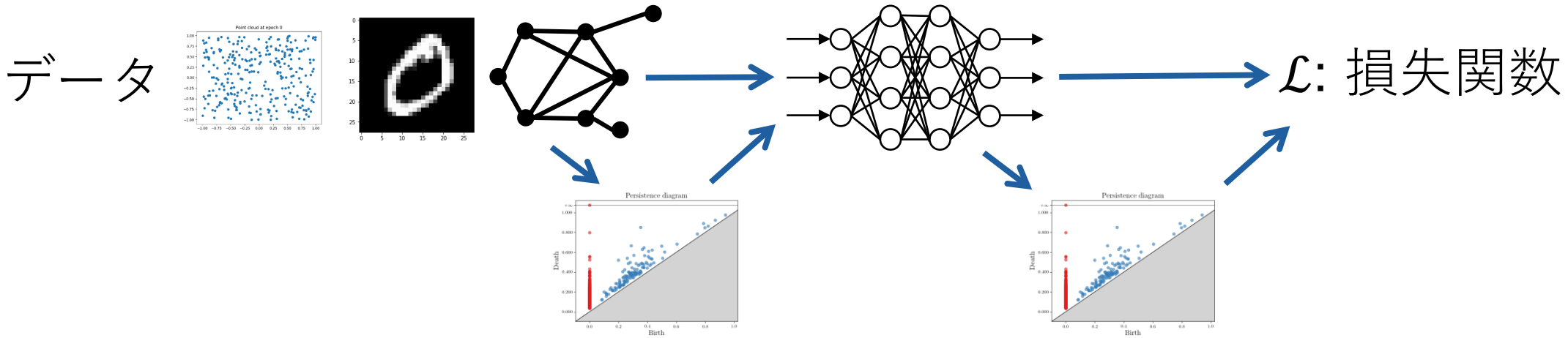
   e.g.　　：生成単体　　：消滅単体

2. 各組に**フィルトレーションの値を対応**させる$\{(f(\sigma_{b_i}), f(\sigma_{d_i}))\}_i$ e.g. (2,3)

■Fが滑らかにパラメータ付けられている ⇒ 微分$x \mapsto \{(\nabla_x F(\sigma_{b_i}), \nabla_x F(\sigma_{d_i}))\}_i$
を単体の順序が変わらない範囲で考えることが可能

PH的損失関数の収束

■**PH的損失関数に関する最適化の保証**は個別の関数について考えられており，理論保証がない場合もあった



■**Carrière, Chazal, Glisse, I., Kannan, and Umeda, Optimizing persistent homology based functions, 2021**

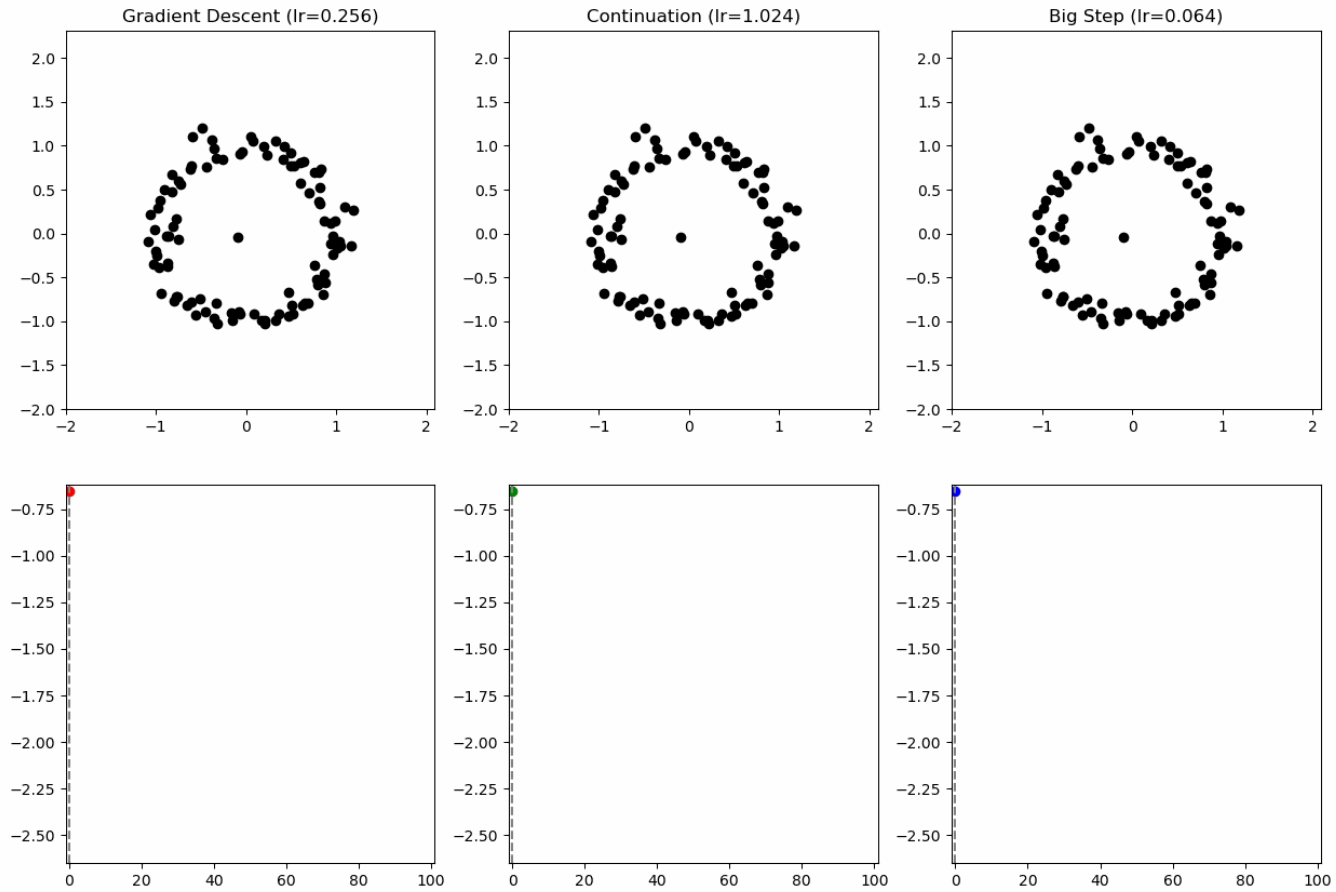
■広いクラスのPH的損失関数について**確率的劣勾配降下法がほとんど確実に収束**することを証明した

■ フィルトレーションがdefinableで損失関数が局所リプシッツならOK

cf. Davis et al., Stochastic subgradient method converges on tame functions, 2020

PH的関数最適化ライブラリの宣伝

共同研究者の西川直輝さん（東大修士学生）が**PH的関数の最適化ライブラリ**を作成



- いろいろ遊べます
- 近年開発された勾配降下法以外の効率的な手法も実装

https://github.com/git-westriver/benchmark_ph_optimization

応用例 (1/2)：画像の劣位集合

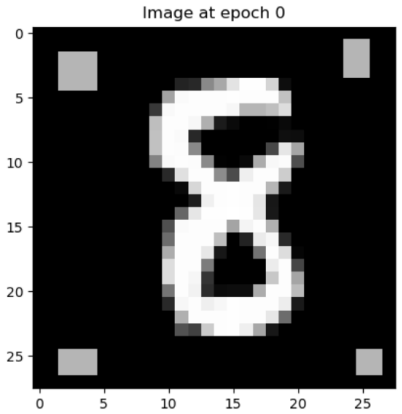
画像を変形して染みをのせた手書き数字画像から染みを削除したい

- 損失関数として$\mathcal{L}_1(I) = \sum_{p \in D_0(I)} \|p\|_2^2$を用いて勾配法を適用すると，連結成分を減らすように働き中央の画像のようになる
- ピクセル値が0か1になるように

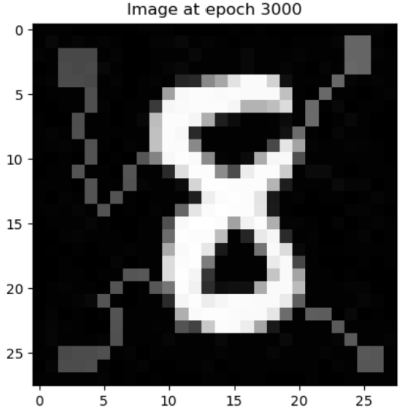
$$\mathcal{L}_2(I) = \sum_{p \in D_0(I)} \|p\|_2^2 + \sum_i \max\{|x_i|, |1 - x_i|\}$$

を用いると結果は右の画像のようになる

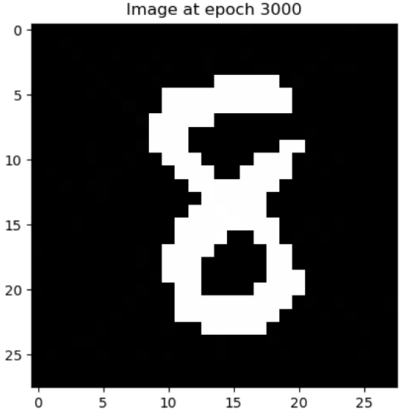
元画像



$\mathcal{L}_1$の結果



$\mathcal{L}_2$の結果



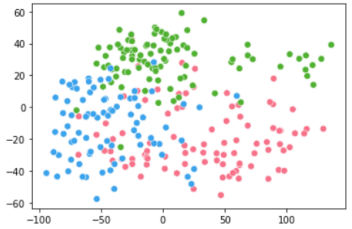
応用例 (2/2)：埋め込みのPH的統制

Vandaele et al., Topologically Regularized Data Embeddings, 2022: データ埋め込み $E: \mathbb{X} \to \mathbb{R}^d$ に応用

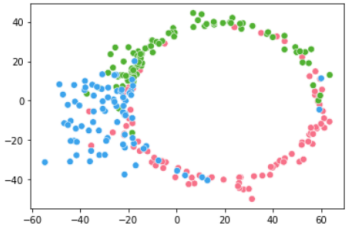
■EのPDの点を$d_k - b_k$が降順になるように並べて

$$\mathcal{L}_{\mathrm{top}}(E) := \mu \sum_{k=i,d_k<\infty}^{k=j} (d_k - b_k),\ \mu \in \{\pm 1\} \text{ を考える}$$

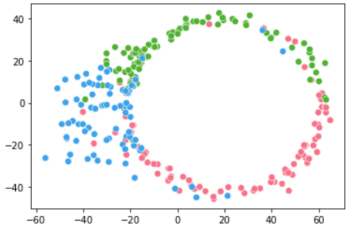
■$\mathcal{L}_{\mathrm{emb}}(E, \mathbb{X}) + \lambda_{\mathrm{top}} \mathcal{L}_{\mathrm{top}}(E)$を最適化することで$E$のトポロジーを統制



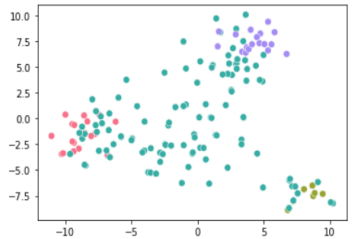
(a) Ordinary PCA embedding.



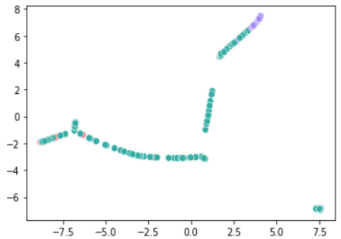
(b) Top. optimized embedding.



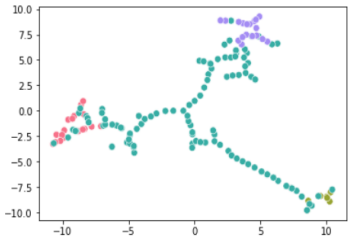
(c) Top. regularized embedding.



(a) Ordinary UMAP embedding.



(b) Top. optimized embedding.



(c) Top. regularized embedding.

$\left\| x - \frac{1}{|E|} \sum_{y \in E} y \right\| \geq \tau$
を満たす点のPDを用いる

PH損失関数の別の応用：フィルター学習

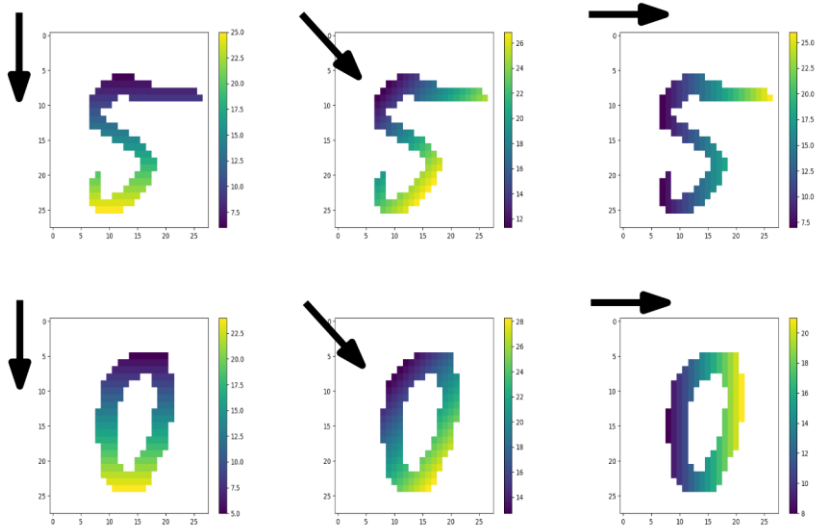
分類に用いる**PHのフィルターで最適なものを学習**する

■手書き数字画像を劣位集合フィルトレーションの0次PDで分類する問題

■方向から定まる1次関数の族をとり

$$\mathcal{L}(f) = \sum_{l} \frac{\sum_{y_i=y_j=l} d(D_0(I_i, f), D_0(I_j, f))}{\sum_{y_i=l} d(D_0(I_i; f), D_0(I_j; f))}$$

を最適化する．最適化前と後での（RFによる）分類精度は下の表



Dataset	Baseline	Before	After	Difference	Dataset	Baseline	Before	After	Difference
vs01	100.0	61.3	99.0	**+37.6**	vs26	99.7	98.8	98.2	-0.6
vs02	99.4	98.8	97.2	-1.6	vs28	99.1	96.8	96.8	0.0
vs06	99.4	87.3	98.2	**+10.9**	vs29	99.1	91.6	98.6	**+7.0**
vs09	99.4	86.8	98.3	**+11.5**	vs34	99.8	99.4	99.1	-0.3
vs16	99.7	89.0	97.3	**+8.3**	vs36	99.7	99.3	99.3	-0.1
vs19	99.6	84.8	98.0	**+13.2**	vs37	98.9	94.9	97.5	**+2.6**
vs24	99.4	98.7	98.7	0.0	vs57	99.7	90.5	97.2	**+6.7**
vs25	99.4	80.6	97.2	**+16.6**	vs79	99.1	85.3	96.9	**+11.5**

グラフのフィルトレーション学習

Hofer et al., Graph Filtration Learning, 2020

- ■フィルター関数をタスクとデータに応じて学習
- ■頂点上の関数をグラフニューラルネットワーク (GNN) として実現
  - ■$V \ni v \mapsto \mathrm{GNN}(G, l(v))$, $l: V \to \mathbb{R}^n$はノードの表現
  - ■実験では隠れ層64次元の1-GINを用いて，後ろに2層のNNを使っている
- ■出来上がったPDのベクトル化を分類に使用

Method	REDDIT-BINARY	REDDIT-MULTI-5K	IMDB-BINARY	IMDB-MULTI
	Initial node features: uninformative		*Initial node features:* $l(v) = \deg(v)$	
PH-only	$90.3_{\pm 2.6}$	$55.7_{\pm 2.1}$	$68.9_{\pm 3.5}$	$46.1_{\pm 4.2}$
1-GIN (GFL)	$90.2_{\pm 2.8}$	$55.7_{\pm 2.9}$	$74.5_{\pm 4.6}$	$49.7_{\pm 2.9}$
1-GIN (SUM) (Xu et al., 2019)	$81.2_{\pm 5.4}$	$51.0_{\pm 2.2}$	$73.5_{\pm 3.8}$	$50.3_{\pm 2.6}$
1-GIN (SP) (Zhang et al., 2018a)	$76.8_{\pm 3.6}$	$48.5_{\pm 1.8}$	$73.0_{\pm 4.0}$	$50.5_{\pm 2.1}$
Baseline (Zaheer et al., 2017)	$77.5_{\pm 4.2}$	$45.7_{\pm 1.4}$	$72.7_{\pm 4.6}$	$49.9_{\pm 4.0}$
State-of-the-Art (NN)				
DCNN (Wang et al., 2018)	n/a	n/a	49.1	33.5
PatchySAN (Niepert et al., 2016)	86.3	49.1	71.0	45.2
DGCNN (Zhang et al., 2018a)	n/a	n/a	70.0	47.8
1-2-3-GNN (Morris et al., 2019)	n/a	n/a	74.2	49.5
5-GIN (SUM) (Xu et al., 2019)	88.9	54.0	74.0	48.8

Sparse　　Dense

フィルトレーションを学習した方が精度が高まる場合がある

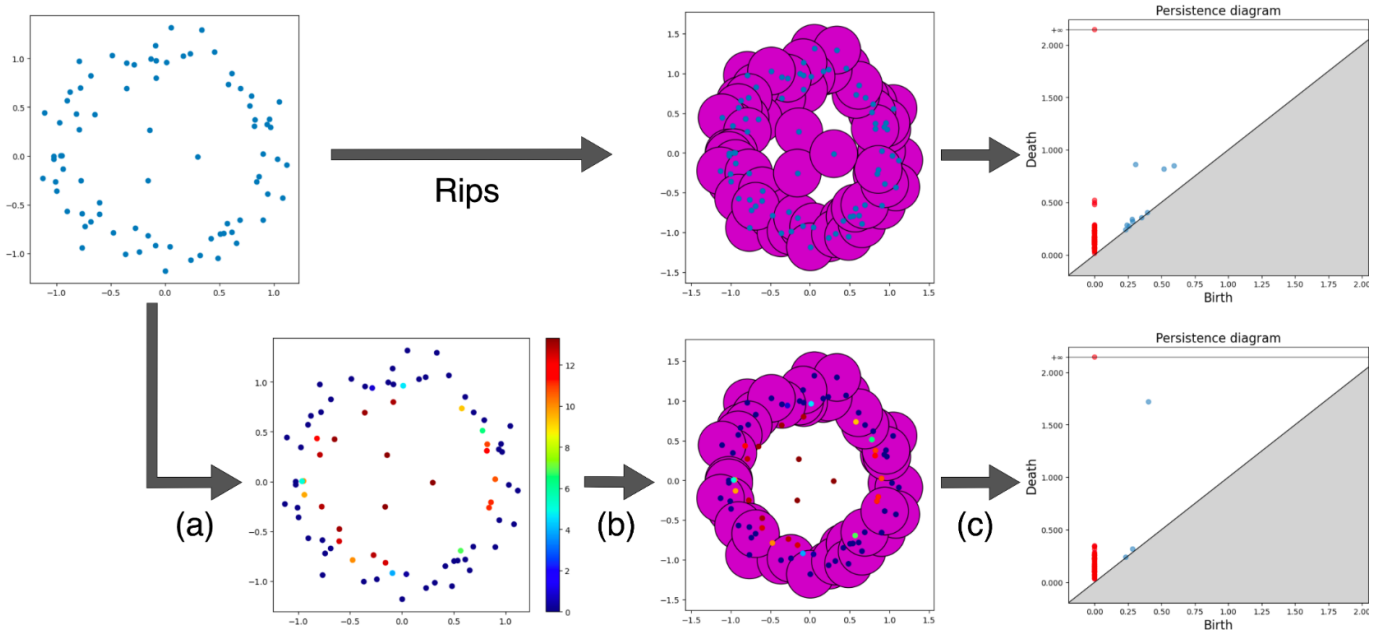
点群のフィルトレーション学習 (1/2)

Nishikawa, I., and Yamanishi, Adaptive Topological Feature via Persistent Homology: Filtration Learning for Point Clouds, 2023

■分類タスクに応じて**点群のフィルトレーションを学習**

■ 球の膨張の**開始時刻を点ごとに学習**

■ **等長変換不変な構造**を提案



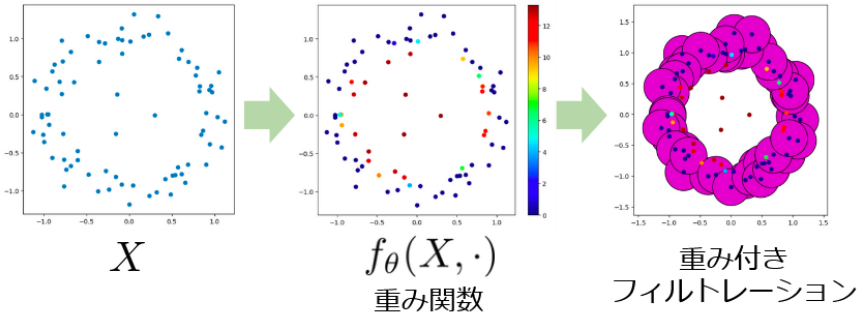
ユーザがフィルトレーションを指定しなくても，**データから自動的に学習**

タンパク質分類の精度

DistMatrixNet	Rips	DTM	Ours
65.0 ± 12.0	79.9 ± 3.0	78.0 ± 1.6	**81.9 ± 2.1**

点群のフィルトレーション学習 (2/2)

■各点xの球の**膨張の開始時刻**を点群Xと点xを入力とした
ニューラルネットワーク$f_\theta(X,x)$で定める

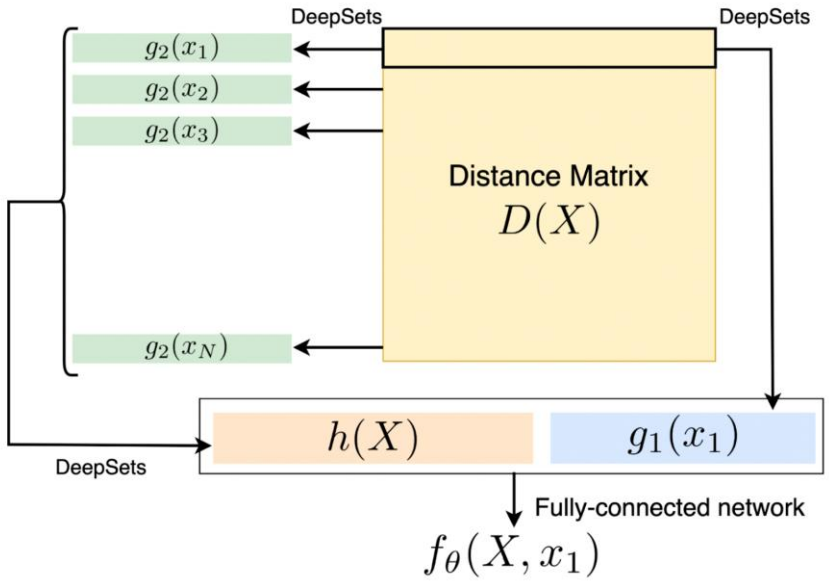


■$f_\theta(X,x)$が満たす条件

1. **点の順序について不変**
2. **点群全体Xの性質と各点xの性質の両方**を反映
3. **等長不変変換**：$f_\theta(TX,Tx)=f_\theta(X,x)$（$T$: 等長変換）

■**距離行列**と**DeepSets**を用いて
これらの条件を満たすように構成

■**PersLayでベクトル化**して分類に使用
全体を**タスクに合わせて学習**



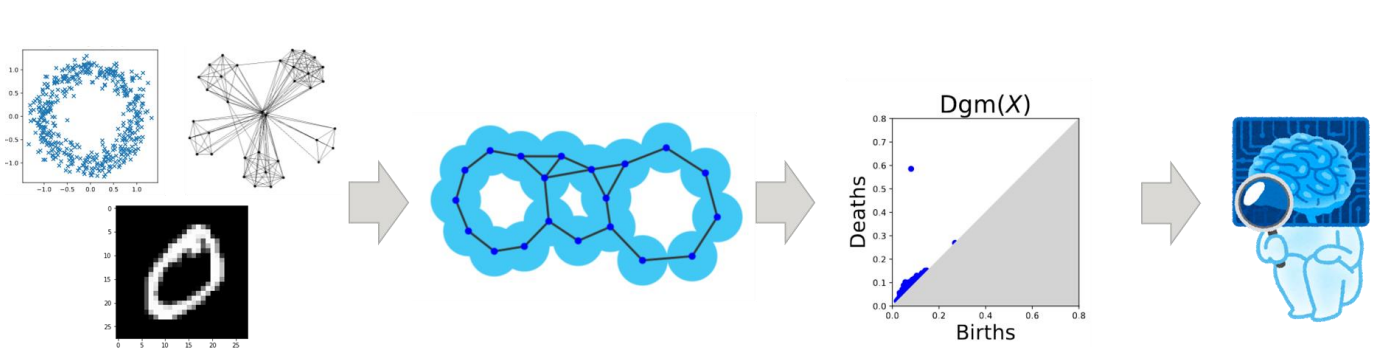
まとめ

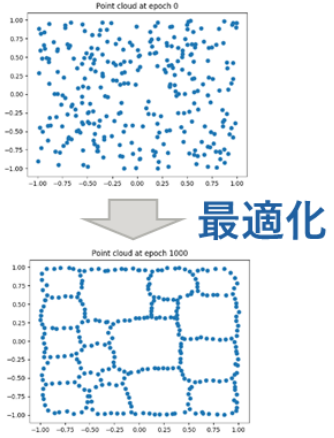
1. 機械学習について少しだけ

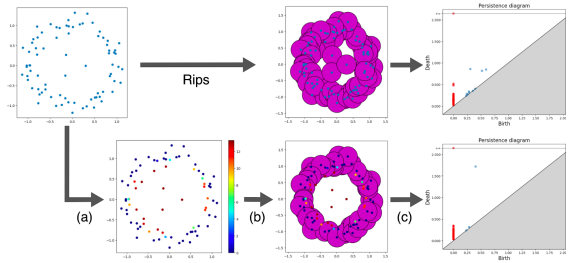
- 機械学習はタスクとデータから勝手に**望ましい関数を学習**させる手法
- モデルという関数族から，当てはまりの悪さをはかる**損失関数を最小化**することで関数を選択

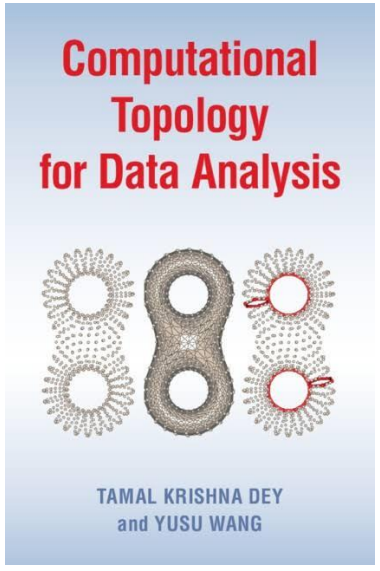
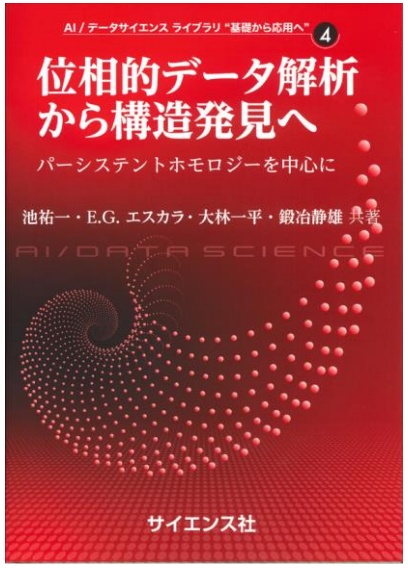
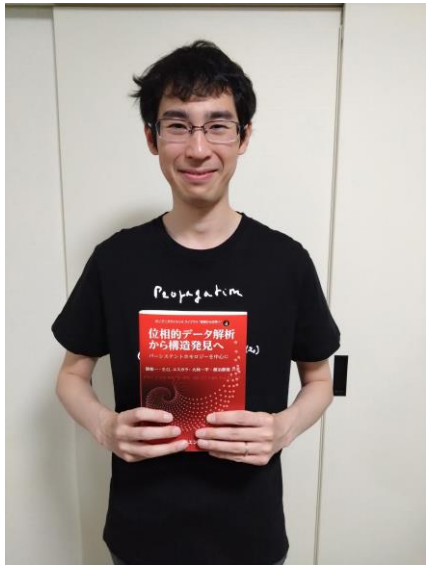
2. 位相的データ解析の応用例 パート2

- **PDを機械学習に入力する**ことでトポロジー情報を使って機械学習できる
- 最近はデータの記述子以外の使い方も現れていて，**PDの微分可能性**に基づいた**トポロジー的学習**や**フィルトレーション学習**などがある







参考文献

- T. Dey and Y. Wang, Computational Topology for Data Analysis, Cambridge University Press, 2022.
- 池祐一・E.G. エスカラ・大林一平・鍛冶静雄, 位相的データ解析から構造発見へ：パーシステントホモロジーを中心に, サイエンス社, 2023.