Topologiczna Analiza Danych

Wojciech Kołowski

Styczeń 2020

1. Analiza Danych
2. Topologia
3. Homotopia
4. Kompleksy
5. Homologia
6. Topologiczna Analiza Danych

Tradycyjne metody analizy danych

- Analiza danych często polega na dopasowaniu do danych jakiegoś uprzednio ustalonego kształtu albo zbadaniu ich struktury.
- Regresja liniowa to dopasowanie prostej.
- Grupowanie danych to dopasowanie kilku plam o różnym położeniu.
- W grupowaniu hierarchicznym dodatkowo jesteśmy zainteresowani grupowaniem grup w grupy wyższego rzędu.
- PCA i inne metody redukcji wymiarowości mają na celu dopasowanie danych do jak najmniejszej podprzestrzeni tak, żeby zachować jak najwięcej informacji.
- Topologiczna analiza danych (TDA) to technika uogólniająca wszystkie powyższe podejścia za jednym zamachem.

Wady tradycyjnych metod analizy danych

- Powyższe podejścia mają wiele różnych wad.
- Zasób predefiniowanych kształtów zazwyczaj jest ubogi (zna ktoś algorytm sprawdzający, czy dane nie mają przypadkiem kształtu butelki Kleina?).
- Algorytmy grupowania często wymagają zadania liczby grup albo mają inne ograniczenia, np. grupy muszą mieć taki sam kształt albo gęstość.
- Grupowanie hierarchiczne może znaleźć hierarchie tam gdzie ich nie ma.
- Topologiczna analiza danych spieszy nam na ratunek i poprawia wszystkie te słabości.

Kształt danych jako ich podsumowanie

- Topologia to nauka o kształtach, zaś topologiczna analiza danych to dziedzina stosująca topologię do badania kształtu danych.
- Zamiast dopasowywać predefiniowany kształt (jak prosta), po prostu znajdujemy faktyczny kształt.
- Kształt danych informuje nas o grupach występujących w danych i ich wzajemnych związkach.
- Poznanie kształtu danych ujawnia ich prawdziwy wymiar oraz umożliwia ich skompresowanie.
- Kształt danych podsumowuje ich najważniejsze właściwości i jest punktem wyjścia do dalszej analizy.

Fantastyczne kształty i jak je znaleźć

- Rodzą się zatem różne pytania.
- Jak znaleźć kształt danych?
- Jakie są w ogóle możliwe kształty?
- Co to jest kształt?
- ???
- Na te pytania odpowie nam geometria, a konkretniej topologia, a konkretniej topologia algebraiczna, a konkretniej metoda zwana z ang. persistent homology.

Krótka historia geometrii

- Historycznie topologia wyrosła z analizy matematycznej, ale filozoficznie bliżej jej do geometrii.
- Grecka geometria wyrosła z problemów praktycznych (wszak geometria to nic innego jak mierzenie Ziemii), ale szybko wyewoluowała do stanu nieco bardziej Platonicznego.
- Grecy myśleli, że wiadomo co to przestrzeń i badali różne rzeczy takie jak punkty, proste, kąty, okręgi, trójkąty etc.
- Po odkryciach Łobaczewskiego/Gaussa (geometria nieeuklidesowa), Einsteina (czasoprzestrzeń, do tego zaginana przez masę) i próbach formalnego okiełznania analizy matematycznej okazało się, że jednak nie wiadomo co to przestrzeń.

Zgniły geometryczny liberalizm

- Obecnie w geometrii panuje liberalizm – każdy może mieć taką przestrzeń, jaką sobie zdefiniuje.
- Jeżeli chcemy mieć odległość, to potrzebna nam przestrzeń metryczna.
- Jeżeli chcemy mieć kąty, to potrzebna nam przestrzeń liniowa z iloczynem skalarnym.
- Jeżeli nie chcemy faworyzować żadnych punktów, to potrzebna nam przestrzeń afiniczna – nie ma tam początku układu współrzędnych.
- Jeżeli chcemy rachunek różniczkowy, to potrzebna nam rozmaitość (manifold), czyli przestrzeń, która lokalnie jest euklidesowa, ale globalnie niekoniecznie.
- A kto powiedział, że w przestrzeni muszą być jakieś punkty? Patrz: pointless topology.

Topologia – cóż to za zwierzątko?

- Topologia to ten dział geometrii, który ma dość elastyczne podejście do pojęcia przestrzeni – jeżeli naszą przestrzeń trochę pociągniemy albo pozginamy, dalej jest to ta sama przestrzeń. Nie wolno nam natomiast przestrzeni ciąć ani rwać – wtedy przestrzeń się zmienia.
- Obrazowo: odcinek to prosta, a okrąg to kwadrat. Dla topologa kubek niczym nie różni się od pączka z dziurą, bo kubek można w sposób ciągły przekształcić w pączek. `https://en.wikipedia.org/wiki/File:Mug_and_Torus_morph.gif`
- Krowa natomiast niczym nie różni się od sfery, co obrazuje poniższy obrazek: `https://en.wikipedia.org/wiki/File:Spot_the_cow.gif`

Topologia algebraiczna

- W nowoczesnej topologii (i ogólniej w nowoczesnej matematyce) jest taki pomysł, żeby badać obiekty (np. przestrzenie topologiczne) przypisując im jakieś inne obiekty (np. grupy).
- Motywacja jest taka, że przypisane obiekty są prostsze w obsłudze, a zawierają wystarczająco informacji dla naszego celu (np. pozwalają odróżniać nieizomorficzne przestrzenie).
- Dzięki temu w jednej dziedzinie (topologia) można wykorzystać wiedzę z innej (teoria grup).
- Tym właśnie zajmuje się topologia algebraiczna: jak użyć algebry do badania przestrzeni topologicznych.
- Znów panuje liberalizm – nie ma jakiejś jednej nadrzędnej teorii i każdy może sobie wymyślić swoją. Najlepiej mi znaną jest teoria homotopii, ale dla topologicznej analizy danych ważniejsze będą przeróżne teorie homologii.

Przykład topologii algebraicznej – dziury

- Zobaczmy na przykładzie, o co chodzi w topologii algebraicznej.
- Jednym z niezmienników, które pozwalają rozróżniać przestrzenie, jest struktura jednowymiarowych dziur w danej przestrzeni.
- Dla przykładu, okrąg nie jest tym samym co płaszczyzna, odcinek ani sfera, bo okrąg ma jednowymiarową dziurę, zaś płaszczyzna, odcinek ani sfera nie mają jednowymiarowych dziur.
- Co to jest dziura, jak możemy ją opisać i wykryć?
- Są różne teorie, między innymi teoria homotopii oraz teoria homologii. Przyjrzyjmy się obu.
- Ostrzeżenie: poniższe slajdy dotyczące teorii homotopii są w zasadzie nie na temat, tzn. nie mają związku z topologiczną analizą danych.

Teoria homotopii – intuicja

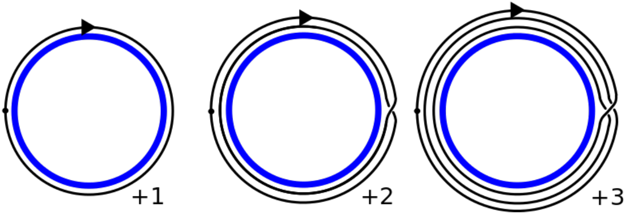
- Kluczowe jest pojęcie pętli i ich homotopii.
- Pętla w punkcie x to ścieżka, która zaczyna się i kończy w punkcie x.
- Dwie pętle (o początku i końcu w tym samym punkcie) są homotopiczne, jeżeli można w sposób ciągły przekształcić jedną na drugą.
- Obrazek (ale dla ścieżek, a nie dla pętli): `https://en.wikipedia.org/wiki/File:HomotopySmall.gif`
- W każdym punkcie x jest też pętla, którą pójście polega tak na prawdę na staniu w miejscu. Zwie się ona pętla trywialną.
- Jeżeli jakaś pętla nie jest homotopiczna z pętlą trywialną, to znaczy, że biegnie dookoła dziury.
- Jak to wszystko formalnie opisać?

Teoria homotopii – grupa podstawowa 1

- Jeżeli mamy przestrzeń X i punkt x, to wszystkie pętle w x tworzą grupę (uwaga: przyjmujemy, że pętle są równe, jeżeli są homotopiczne).
- Działanie to sklejanie pętli: mając dwie pętle, możemy najpierw pójść jedną, a potem drugą, co w wyniku też daje pętlę.
- Element neutralny to pętla trywialna.
- Odwrotność pętli to pójście nią w przeciwnym kierunku.
- Grupa wszystkich pętli o początku i końcu w punkcie x z działaniem jak powyżej zwie się grupą podstawową, $\pi_1(X, x)$.
- Co więcej, każda funkcja ciągła $f : X \to Y$, która zachowuje wyróżniony punkt (czyli $f(x) = y$), daje homomorfizm grup $\tilde{f} : \pi_1(X, x) \to \pi_1(Y, y)$.

Teoria homotopii – grupa podstawowa 2

Grupa podstawowa sfery S^2 jest trywialna, bo każdą pętlę można przekształcić w pętle trywialną, co demonstruje obrazek.



Grupa podstawowa okręgu to $\mathbb{Z}$, bo możemy stać w miejscu (0), pójść k razy zgodnie z ruchem wskazówek zegara $(+k)$, albo przeciwnie do ruchu wskazówek zegara $(-k)$

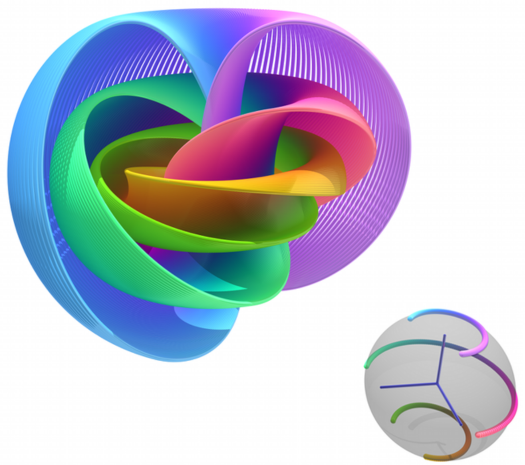
Teoria homotopii – grupa podstawowa 3

- Dla odcinka $[0,1]$ oraz prostej $\mathbb{R}$, dysku i kwadratu (pełnego w środku) grupa podstawowa jest trywialna, podobnie jak dla sfery.
- Stąd wniosek, że okrąg nie jest izomorficzny ze sferą/prostą/odcinkiem/dyskiem/kwadratem etc.
- Lista grup podstawowych różnych przestrzeni: `http://mathonline.wikidot.com/list-of-fundamental-groups-of-common-spaces`

Teoria homotopii – problemy obliczeniowe

- Teoria homotopii nie jest bez wad.
- Wada obliczeniowa jest taka, że wyższe grupy homotopii (czyli grupy opisujące dziury o wymiarze ≥ 1) trudno jest obliczyć – znane są algorytmy to robiące, ale w większości ciekawych przypadków nikt nigdy nie doczekał się, aż skończyły pracę.
- Wobec powyższego ciężko liczyć na to, że teoria homotopii nada się do analizy danych.

Teoria homotopii – problemy konceptualne



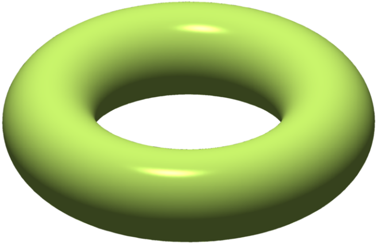
Nieco ciekawsza jest "wada" konceptualna: (wyższe) grupy homotopii są bardzo nietrywialne i zaskakujące, np. dwuwymiarowa sfera S^2 ma trójwymiarową dziurę, znaną jako fibracja Hopfa.

Droga do homologii

- Bardziej przyjaznymi teoriami, zarówno pod względem algorytmicznym, jak i braku trójwymiarowych dziur w dwuwymiarowych obiektach, są przeróżne teorie homologii.
- Za chwilę przyjrzymy się czemuś, co z ang. nazywa się simplicial homology, czyli teorii homologii dla simpleksów i kompleksów.
- Teoria używana w topologicznej analizie danych zwie się natomiast z ang. persistent homology i zobaczymy ją głównie na obrazkach.
- Zanim jednak do tego dojdzie, musimy odpowiedzieć sobie na parę podstawowych pytań.

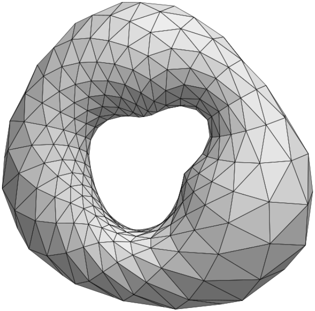
Jak reprezentować przestrzenie w komputerze? 1

- Przestrzenie mogą być skomplikowane.
- Mogą mieć nieprzeliczalnie wiele punktów, a komputery mają przecież skończoną pamięć.
- Mogą być gładkie i ciągłe, a komputery operują przecież na dyskretnych danych.
- Mogą rozciągać się w nieskończoność (zauważ, że to inna właściwość, niż posiadanie nieskończonej ilości punktów).
- Przykładem upierdliwej przestrzeni mającej dwie pierwsze właściwości jest torus.



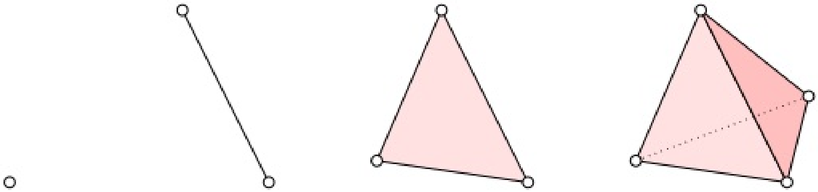
Jak reprezentować przestrzenie w komputerze? 2

- Sytuacja jest podobna jak z liczbami rzeczywistymi. Jak je reprezentować? Nie da się, więc konieczne jest przybliżenie, czyli liczby zmiennoprzecinkowe.
- Analogicznie dla przestrzeni: każdą niepatologiczną przestrzeń można przybliżyć za pomocą pewnej ilości prostych figur geometrycznych, jak trójkąty, kwadraty etc. Poetycko zwie się to triangulacją.



Simpleksy

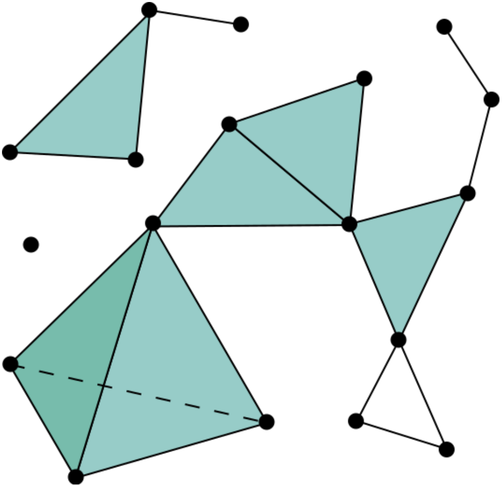
- Torus jest dwuwymiarowy, więc do jego triangulacji wystarczą same trójkąty, ale przestrzenie więcej niż dwuwymiarowe wymagają czegoś więcej.
- Na ratunek przychodzą nam simpleksy. Simpleks to uogólnienie trójkąta na dowolną liczbę wymiarów.
- 0-simpleks to punkt, 1-simpleks to odcinek, 2-simpleks to trójkąt, 3-simpleks to czworościan, 4-simpleks to... wyobraźnio do boju.



Kompleksy

- Kompleks (ang. simplicial complex) to byt zrobiony z simpleksów.
- Jeżeli σ i τ są simpleksami w kompleksie C, to ich przecięcie $\sigma \cap \tau$ też musi być simpleksem w C.
- Jeżeli σ jest simpleksem w kompleksie C, zaś τ jest krawędzią/ścianą (jak to nazwać w wyższym wymiarze?) σ, to τ jest simpleksem w C.
- Formalniej kompleks możemy reprezentować za pomocą rodziny zbiorów, np. $\{\{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}, \{d\}, \{a, d\}\}$ to trójkąt z odcinkiem ad przyklejonym do a (sprawdź to).

Kompleksy – przykład



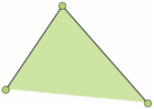
Ćwiczenie: opisz formalnie (jako zbiór) kompleks z obrazka.

Kompleksy - antyprzykład

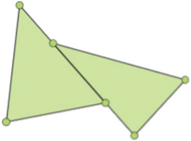
Not a Simplicial Complex

- Every face of a simplex in a complex is in the complex

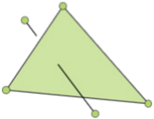
Edge is missing



- Non-empty intersection of two simplices is a face of each of them



Sharing half an edge



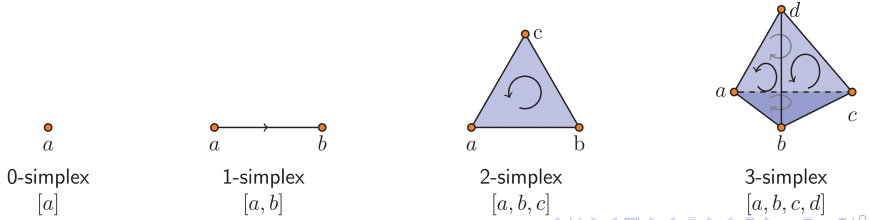
Intersection not a vertex

Teoria dziur, znowu

- Jest intuicyjnie jasne, że trójkąt (pełny w środku) nie zawiera żadnej dziury, ale jego brzeg (czyli trzy odcinki) zawiera dziurę.
- Jak sformalizować tę intuicję?
- A jak rozszerzyć ją na dziury wyższego wymiaru?
- Podstawową rzeczą, jaką można zrobić z dziurą i nie zrobić sobie przy tym krzywdy, to chodzenie dookoła niej.
- Oczywiście można też chodzić dookoła rzeczy, które nie są dziurami.
- Te dwa podstawowe fakty już niedługo posłużą nam do zdefiniowania grup homologii, ale najpierw kilka niezbędnych definicji.

Orientacja 1

- Niech C będzie kompleksem. O każdym simpleksie $\sigma \in C$ możemy myśleć, że ma jakąś orientację.
- Zorientowane simpleksy możemy reprezentować za pomocą ciągów wierzchołków.
- Dla odcinków orientacja to po prostu kierunek. ab idzie z lewa na prawo, a ba z prawa na lewo.
- Dla trójkątów orientacja to wybór definicji dla "zgodnie z ruchem wskazówek zegara". abc to zgodnie, zaś acb – przeciwnie.



Orientacja 2

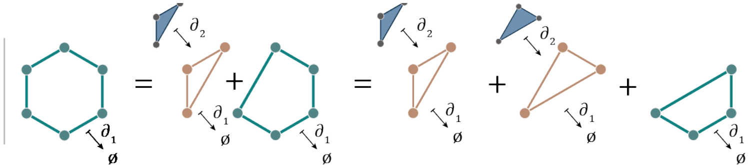
- Niestety orientację wierzchołka albo czworościanu ciężej sobie wyobrazić, będziemy więc musieli radzić sobie symbolicznie.
- Ponieważ są tylko dwie orientacje, to zmianę orientacji możemy reprezentować za pomocą znaku $-$, a zatem $ab = -ba$.
- Jeżeli zmienimy kolejność wierzchołków w naszym zorientowanym simpleksie, to orientacja ulega zmianie, a zatem $abc = -bac = bca = -cba = cab = -acb$.

Łańcuchy

- Łańcuch to formalna suma simpleksów tego samego wymiaru (czyli dodawanie nic nie robi, po prostu jest).
- Przykład 1: $ab + bc + ca$ to łańcuch zrobiony z trzech boków trójkąta, czyli trójkąt "pusty w środku".
- Przykład 2: $a + b + c$ to łańcuch zrobiony z wierzchołków trójkąta.
- Łańcuchy simpleksów o wymiarze k tworzą grupę C_k. Element neutralny to pusty łańcuch (nic w nim nie ma), działanie to formalne dodawanie, a branie elementu odwrotnego to zmiana orientacji każdego simpleksu w łańcuchu.

Brzegi i cykle

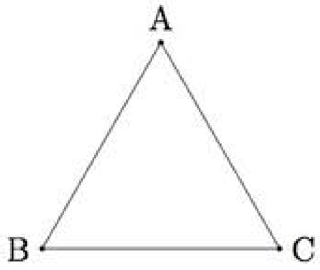
- Operator $\partial_k : C_k \to C_{k-1}$ przypisuje łańcuchowi jego brzeg.
- Przykład: $\partial abc = bc - ac + ab = ab + bc + ca$, czyli brzegiem pełnego w środku trójkąta jest pusty w środku trójkąt.
- Cykl to łańcuch, który nie ma brzegu.
- Przykład: $ab + bc + ca$ to cykl, bo $\partial(ab + bc + ca) = (b - a) + (c - b) + (a - c) = 0$.
- Fakt: $\partial_k \partial_{k+1} \sigma = 0$, czyli brzegi nie mają brzegów. Wobec tego brzegi są cyklami.
- Jeżeli dodamy simpleksy o przeciwnej orientacji, to się skrócą. Dzięki temu możemy rozkładać cykle na prostsze cykle.



Grupy homologii

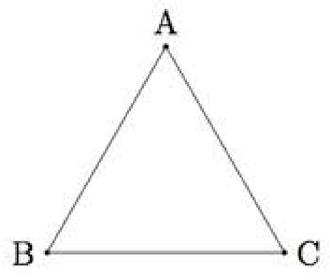
- k-tą grupę homologii kompleksu C możemy zdefiniować tak: $H_k(C) = \ker \partial_k / \mathrm{im}\ \partial_{k+1}$
- $\ker \partial_k$ to grupa k-łańcuchów, które nie mają brzegu, czyli grupa k-cykli.
- $\mathrm{im}\ \partial_{k+1}$ to grupa k-łańcuchów, które są wynikiem działania ∂_{k+1} na $(k+1)$-łańcuchach, czyli grupa k-brzegów.
- Mówiąc po ludzku: k-ta grupa homologii kompleksu C to grupa k-cykli, ktore nie są k-brzegami.
- Po co nam było to wszystko? Otóż jeżeli cykl nie jest brzegiem, to znaczy, że okrąża on jakąś dziurę. Wobec tego $H_k(C)$ opisuje wszystkie sposoby, na jakie możemy w kompleksie C chodzić dookoła k-wymiarowych dziur.

Grupy homologii – przykład 1



- Każdy wierzchołek jest 0-cyklem, bo nie ma brzegu. Stąd $\ker \partial_0 = \mathbb{Z}^3$ (każdy cykl jest zrobiony z kombinacji wierzchołków).
- Z drugiej strony mamy $\partial_1 ab = b - a$, $\partial_1 bc = c - b$, $\partial_1 ca = a - c$. Widać, że $\partial_1 ca = -(\partial_1 ab + \partial_1 bc)$, a zatem są tylko dwa niezależne brzegi i stąd $\operatorname{im} \partial_1 = \mathbb{Z}^2$.
- Wobec tego $H_0(abc) = \mathbb{Z}^3/\mathbb{Z}^2 = \mathbb{Z}$.

Grupy homologii – przykład 2



- Jest tylko jeden 1-cykl, mianowicie $ab + bc + ca$, gdyż $\partial_1(ab + bc + ca) = (b - a) + (c - b) + (a - c) = 0$. Stąd $\ker \partial_1 = \mathbb{Z}$.
- Z drugiej strony mamy $\operatorname{im} \partial_2 = 1$ (grupa trywialna), gdyż żaden 1-łańcuch nie jest brzegiem żadnego 2-simpleksu, czyli trójkąta. Wynika to z faktu, że na obrazku nie ma trójkąta (przypominam, że trójkąt jest pełny w środku).
- Wobec tego $H_1(abc) = \mathbb{Z}/1 = \mathbb{Z}$.

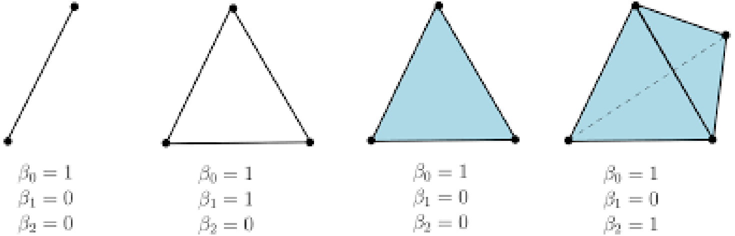
Grupy homologii – interpretacja przykładu

- Jak zinterpretować powyższe wyniki? Gdzie są nasze dziury?
- Dla H_1 sprawa jest prosta – jest jedna dziura, którą można chodzić k razy zgodnie z ruchem wskazówek zegara $(+k)$, k razy przeciwnie do ruchu wskazówek zegara $(-k)$, albo w ogóle nigdzie nie iść (0). Grupa takiego chodzenia dookoła dziury jest więc izomorficzna z $\mathbb{Z}$, czyli dokładnie tak, jak nam wyszło.
- Dla H_0 z wyobraźnią jest trudniej. Wystarczy nam zatem wiedzieć jedynie, że $H_0(C)$ reprezentuje ilość spójnych składowych kompleksu C.

Liczby Bettiego

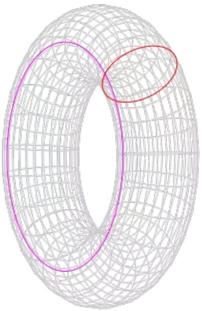
- Grupy homologii byłyby dość upierdliwe do zaimplementowania w większości mainstreamowych języków programowania.
- Nie mówiąc o tym, że w przykładzie ciągle pojawiała się grupa $\mathbb{Z}$ i nie był to przypadek – przydałaby się jakaś prostsza reprezentacja dziur.
- Wobec tego definiujemy liczby Bettiego: dla kompleksu C mamy $\beta_k(C) = \text{rank}(H_k(C))$.
- Po ludzku: k-ta liczba Bettiego to liczba generatorów k-tej grupy homologii.
- Intuicja: β_0 to liczba spójnych składowych, β_1 to liczba dziur jednowymiarowych, β_2 to liczba dziur dwuwymiarowych etc.

Liczby Bettiego – przykład 1



- Odcinek jest spójny ($\beta_0 = 1$) i nie ma dziur ($\beta_k = 0, k \geq 1$).
- Pusty trójkąt jest spójny ($\beta_0 = 1$) i ma jednowymiarową dziurę ($\beta_1 = 1$), ale nie ma więcejwymiarowych dziur ($\beta_k = 0, k \geq 2$).
- Pełny trójkąt jest spójny ($\beta_0 = 1$) i nie ma dziur ($\beta_k = 0, k \geq 1$).
- Pusty czworościan (czyli same ściany, bez wnętrza) jest spójny ($\beta_0 = 1$), nie ma 1-wymiarowych dziur ($\beta_1 = 0$), ma 2-wymiariową dziurę ($\beta_2 = 1$) i nie ma więcejwymiariowych dziur ($\beta_k = 0, \geq 3$).

Liczby Bettiego – przykład 2



Liczby Bettiego dla torusa: $(\beta_0, \beta_1, \beta_2) = (1, 2, 1)$ i $\beta_k = 0$ dla $k \geq 3$. Jest tak dlatego, że torus jest spójny ($\beta_0 = 1$), ma dwie niezależne od siebie jednowymiarowe dziury ($\beta_1 = 2$; zaznaczone na rysunku – nie da się w sposób ciągły przekształcić jednej w drugą) oraz jedną dwuwymiarową dziurę ($\beta_2 = 1$; pamiętajmy, że torus jest pusty w środku).

Liczby Bettiego – ćwiczenia

Ile wynoszą liczby Bettiego dla:

- Zbiorów punktów?
- Grafów?
- Simpleksów?
- Brzegów simpleksów?
- k-wymiarowych sześcianów?
- Brzegów k-wymiarowych sześcianów?

Liczby Bettiego – podsumowanie

- Ponieważ w n-wymiarowej przestrzeni nie może być $\geq n$-wymiarowych dziur, to liczby Bettiego wynoszą zero dla $k \geq$ wymiar przestrzeni.
- Dzięki temu możemy opisać przybliżony kształt przestrzeni za pomocą niezbyt długiego ciągu liczb.
- Fajnie, ale jak technicznie znaleźć ten przybliżony kształt danych?

Dopasowywanie kompleksu do danych 1

- Załóżmy, że mamy N punktów danych pochodzących z k-wymiarowej przestrzeni X (nie musi być $X = \mathbb{R}^k$, ale zazwyczaj pewnie tak będzie).
- Żeby teoria homologii poszła w ruch, musimy dopasować do danych jakiś kompleks.
- Jeżeli nam się to uda, to będziemy mogli opisać jego przybliżony kształt (czyli strukturę dziur) za pomocą liczb Bettiego.
- Ponieważ chmura N punktów daje co najwyżej N-wymiarowy kompleks, to do opisu kształu danych wystarczy $\min(N, k)$ liczb.
- Jeżeli N i k są bardzo duże to możemy mieć problem, ale spodziewamy się, że najciekawsze i tak będzie niskowymiarowe przybliżenie naszego kształtu, więc nie musimy się przejmować.

Dopasowywanie kompleksu do danych 2

- Potrzebna będzie nam metryka $d : X \times X \to \mathbb{R}$.
- Pomysł na dopasowanie kompleksu jest następujący.
- Wybieramy pewien $\varepsilon \in \mathbb{R}$. Jeżeli mamy n punktów, które leżą w odległości $\leq 2\varepsilon$ każdy od każdego, to wrzucamy do naszego kompleksu odpowiadający im n-simpleks.
- Pytanie: jaki ε wybrać?
- Jeżeli nie wiesz jaki wybrać parametr, wybierz wszystkie: `https://sauln.github.io/blog/nerve-playground/`

Kształty w różnej skali

- Parametr ε możemy rozumieć jako skalę w której przyglądamy się naszym danym.
- Jeżeli ε jest bardzo małe, to widzimy gołe punkty bez żadnych powiązań.
- Jeżeli ε jest bardzo duże, to widzimy jeden wielgachny punkt bez żadnej wewnętrznej struktury.
- Dla pośrednich ε jesteśmy w stanie dostrzec różne ciekawe kształty.
- Każdy z tych kształtów powstaje w pewnej skali, a w miarę jak zwiększamy skalę, znika i staje się częścią jakiegoś większego kształtu.

Homologia persystentna

- Taka jest właśnie idea stojącą za teorią (techniką?) homologii persystentnej.
- Żeby dowiedzieć się czegoś o danych, sprawdzamy w jakiej skali rodzą się, a w jakiej umierają dziury różnego wymiaru.
- Im dłużej dana dziura żyje, tym lepiej oddaje ona prawdziwy kształt danych. Dziury żyjące krótko są jedynie przejawem szumu.

Topologiczna analiza danych – podsumowanie

- Na początku postawiliśmy sobie za cel znajdowanie kształtu danych i udało nam się zapoznać z podstawowymi ideami, które to umożliwiają.
- Pozostaje jeszcze kwestia tego, jak wypada to wszystko w praktyce – algorytmy, wyniki etc.
- Niestety 15 minut to za krótko, żeby opowiedzieć od podstaw do praktyki o tak skomplikowanym zagadnieniu jak topologiczna analiza danych.
- Jeżeli chcesz dowiedzieć się więcej, musisz zbadać temat samemu.
- Mam nadzieję, że dzięki tej prezentacji twoje poszukiwania będą mniej bolesne niż moje.

Przydatne materiały do czytania

- Krótkie wprowadzenie do topologicznej analizy danych: `https://jsseely.github.io/notes/TDA/`
- Polecam dokładnie je przejrzeć: zawiera linki do notatek wprowadzających do topologii, filmików ilustrujących homologię persystentną w nieco bardziej interaktywny sposób, prac naukowych z wynikami oraz listy oprogramowania, w tym do niewspomnianego podczas prezentacji algorytmu Mapper.

Przydatne materiały do oglądania

- O simpleksach, kompleksach i triangulacji: `https://www.youtube.com/watch?v=9vLAZkOk3IA`
- Polecam obejrzeć wszystkie filmiki z powyższej serii, stanowią wprowadzenie do homologii persystentnej o niebo bardziej interaktywne niż niniejsza prezentacja.
- Głównie o algorytmie Mapper: `https://www.youtube.com/watch?v=x3Hl85OBuc0`
- O Pythonowych narzędziach do topologii: `https://www.youtube.com/watch?v=AWoeBzJd7uQ`