



Санкт-Петербургский  
Государственный  
Политехнический  
Университет

Институт компьютерных  
наук и технологий

## курс: Решение прикладных задач методами машинного обучения

### Лекция 9

Технологии решения прикладных задач МО.

---

9 ноября 2022 г.

## Что обсуждали на прошлой лекции

- Формально процесс обучение может быть представлен с помощью модел решения обратных алгоритмических задач , т.е. построение алгоритма, который из произвольного входного набора данных умеет «вычислять» определенное решение. Например;  
Дано: некоторое смысловое понятие ( например, «дерево»). Этому понятию соответствует **счетное множество объектов реальности**. Само понятие характеризуется некоторым объемом - мощностью множества объектов, которые кодируются этим понятие ( в данном случае это все деревья на Земле). Это множество образует класс эквивалентности «дерево». Класс эквивалентности является **топологическим пространством** – каждый его элемент это «открытое множество, то есть имеет «окрестностью» весь класс эквивалентности.
  - необходимо «обучить» оператор так, чтобы в наборе произвольных входных данных можно было найти ( обнаружить) класс эквивалентности «дерево», сопоставить ему смысловое понятие и объяснить почему сделан такой выбор.
- Итак, процессу обучения можно сопоставить **оператор отображения** множества входных данных на множество выходных данных.
- В рамках такой модели можно сформулировать много прикладных задач машинного обучения много, но все они имеют общие этапы, а именно:
  - понимание сути решаемой задачи и доступных для ее анализа данных
  - предобработка данных и «изобретение» признаков, которые составляют основу процесса обучения
  - построение модели «объекта» как совокупности признаков
  - решение проблем оптимизации и переобучения модели
  - оценивание качества результата
  - внедрение и эксплуатация модели

## Еще раз о сути задачи обучения ...

Суть процесса можно описать так построение оператора, **непрерывного отображения** счетного множества входного данных («облако данных»: пиксели, воксели, симплексы)) в **конечное множество «осмысленных» понятий**, характеризующих **существенные свойства рассматриваемых объектов** реальности.

Чтобы формализовать задачу обучения надо определить, что есть

1. «объект» машинного обучения – например «смысловое понятие», которое можно сопоставить входным данным
2. «предмет» машинного обучения – оператор отображения

## Обучение как субъективное «естествознание»

В результате обучения «картина мира» определяется не столько свойствами самого этого мира, сколько характеристиками субъекта познания (человека), а именно его «концептуальными взглядами».

Выводы:

Исключить субъективное начало из процесса обучения полностью **невозможно**.

**Причинность**, которая с точки зрения физики, есть объективное начало для всех научных знаний, (в физике, например, причина понимается «механистически» как **внешняя сила или вероятность** – в квантовой механике) **в процессе обучения** обретает новое

Статистическое (повторение),

алгебраическое (операционное)

топологическое (форма)

содержание

## О «физике» процесса обучения

Надо построить процесс выделения из данных их «смыслового описания». Существующие различные подходы к построению «отображения». Для этого можно использовать :

- **Статистические** (корреляции между данными, которые образуют «неделимые» символы)
- **Алгебраические** (факторизация множеств входных данных на классы эквивалентности с сохранением предикатов)
- **Топологические** ( использование дескрипторов топологических инвариантов, например, диаграмм персистентности изменения параметров масштаба)

методы

# Обучение как основа согласованного соучастия объективного и субъективного начал

**Прикладная задача обучения:** построение обладающей потенциальной полнотой **совокупности реакций** субъекта на водные сигналы **в пределах существующих** физических или логических ограничений (сверх-задача обучения гораздо шире)

**Тезис 1.** Любая способная к «обученную» система является «само организованной активной системой», которая реализуют «согласованные действия», используя для этого **физические (причинные), так и «информационные» (знания) ресурсы** у той среде, где система в данный момент существует .

**«Первый закон самоорганизации»** (применим к микро и макро масштабам) :  
Необходимым условием сохранения целостности «само организованной активной системой» в неравновесных и динамически изменяющихся условиях является **согласованное поведение** всех элементов такой системы  
Вопрос: за счет чего происходит «согласование» ? . (структурогенез)

# Структурогенез как процесс ....

- В физике структурогенез - процесс формообразования, обусловленный predetermined последовательностью изменения структурных характеристик строения объекта с учетом влияния среды.



## РОСТ и РАЗВИТИЕ

Все живое растет и развивается.

**Рост** – увеличение в размерах.

**Развитие** – приобретение новых свойств.

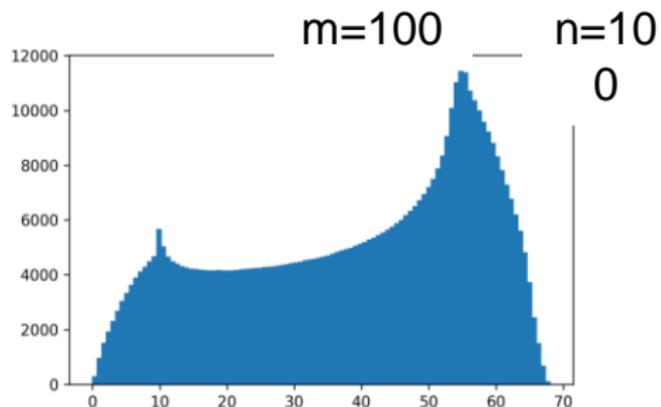
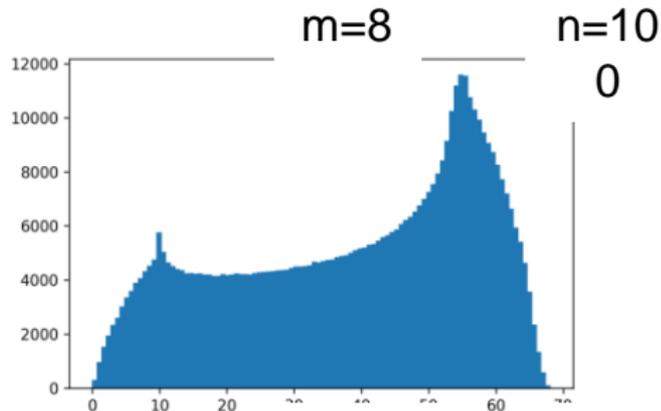
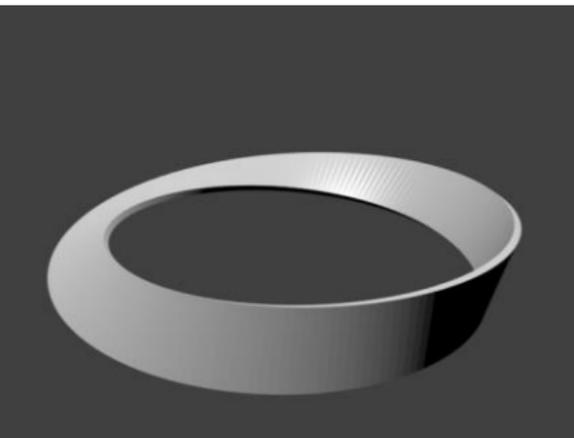
Будут ли такие свойства у камня, стекла, книги, пепала?

Является ли МО структурогенезом?

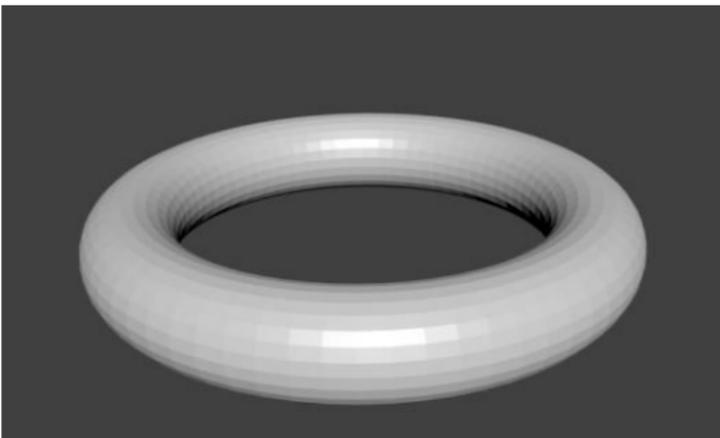
## Рассмотрим подходы к МО, основанные на топологии

- Топология — раздел математики, который изучает **свойства фигур**, сохраняющиеся при непрерывных деформациях — растяжении, сжатии или изгибании.
- В топологии есть понятие гомеоморфность – топологическая эквивалентность – то есть переход одной фигуры в другую без разрывов и склеек.
- Пример: Окружность и граница квадрата гомеоморфны.

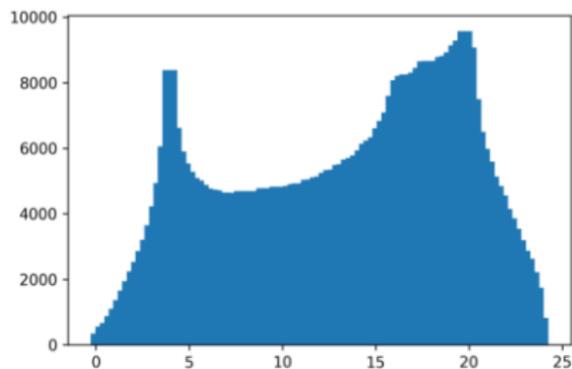
# Пример – лист Мебиуса



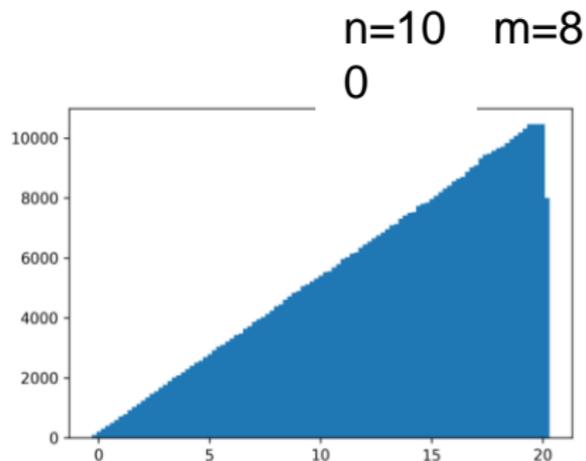
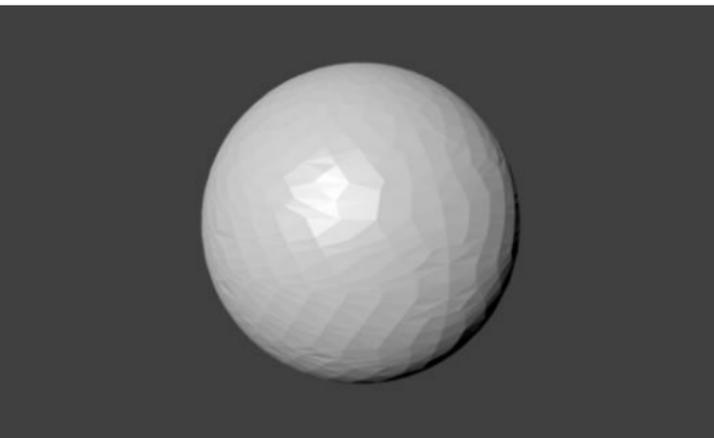
# Инварианты «похожести» лист-Мебиуса—тор



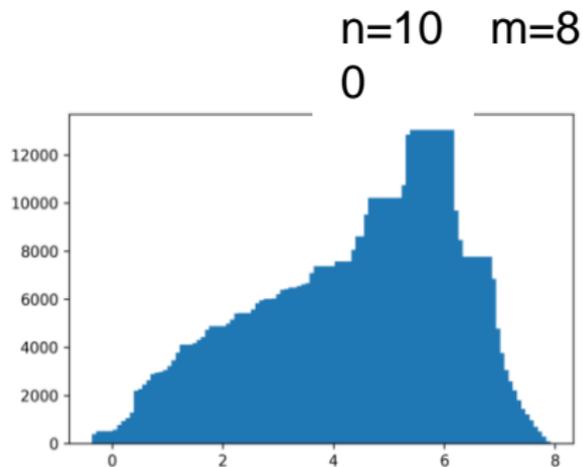
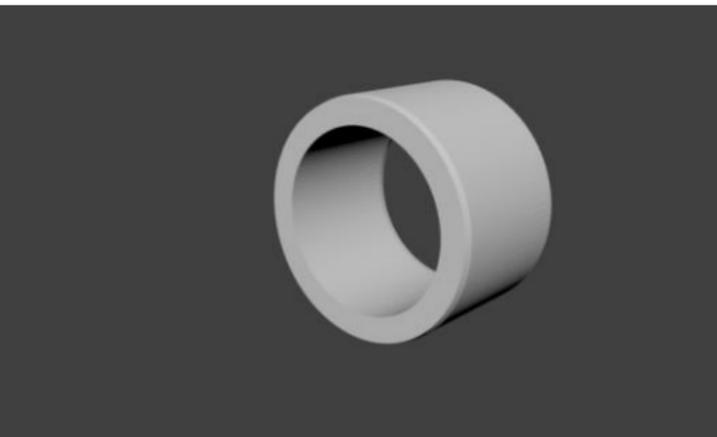
$n=10$   $m=8$



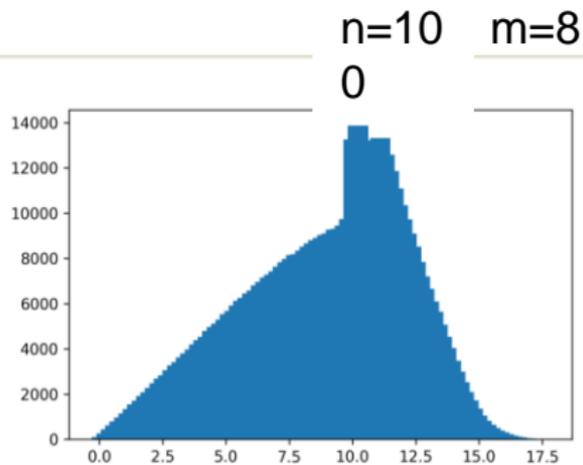
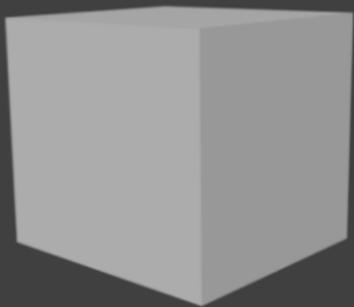
# Робастность меры схожести



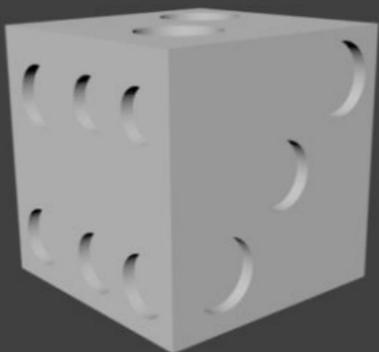
# Робастность меры (1)



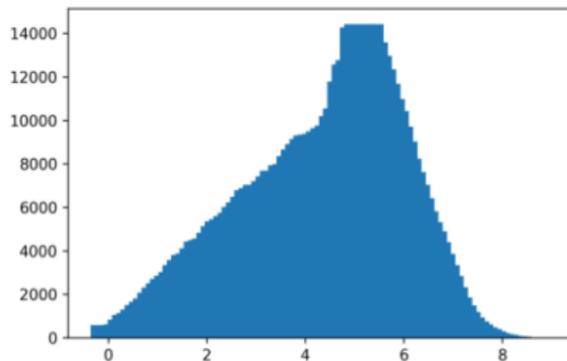
## Робастность меры (2)



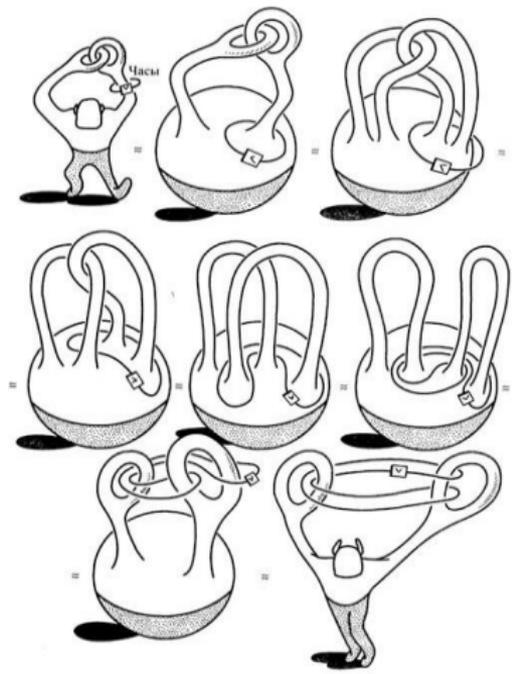
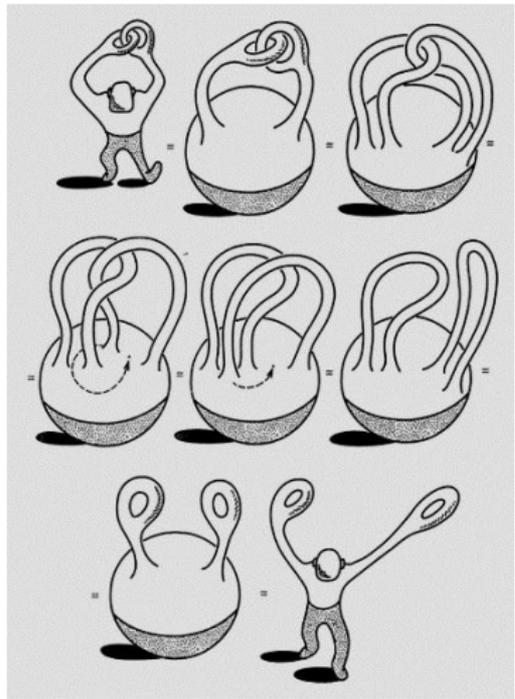
# Робастность + текстура



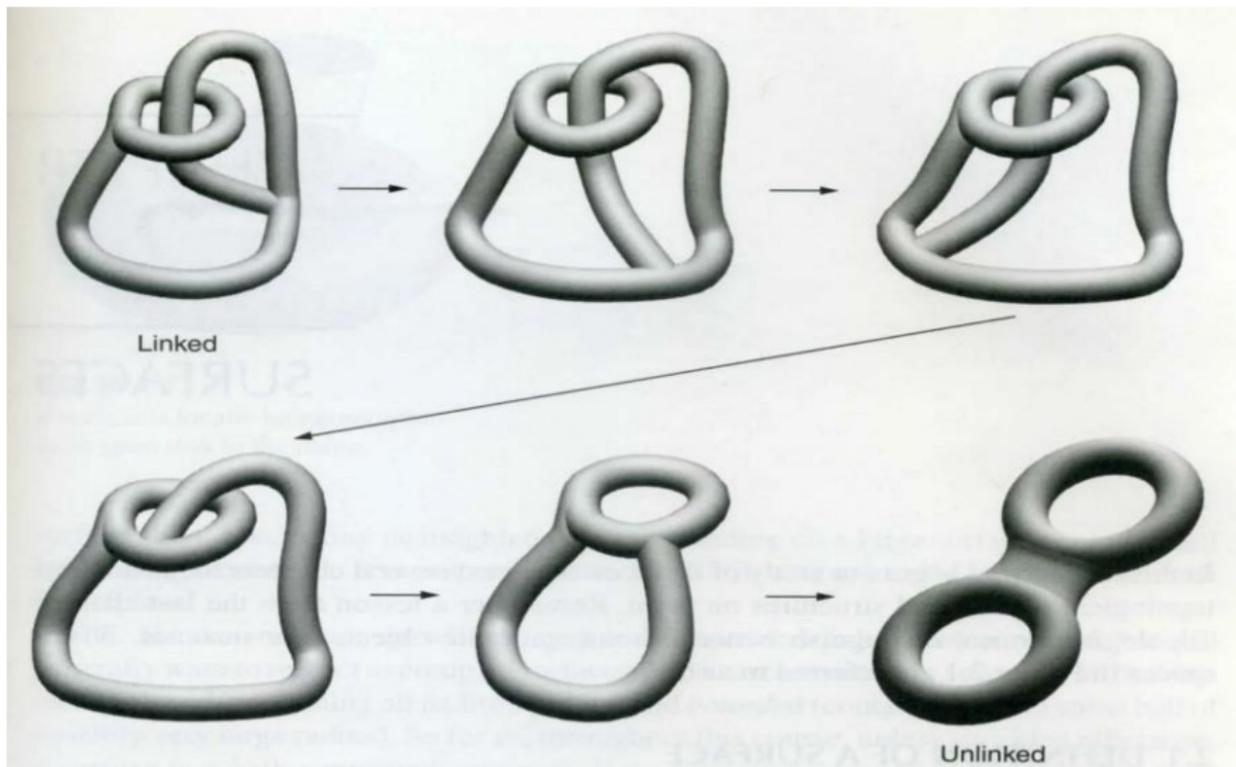
$n=10$   $m=8$



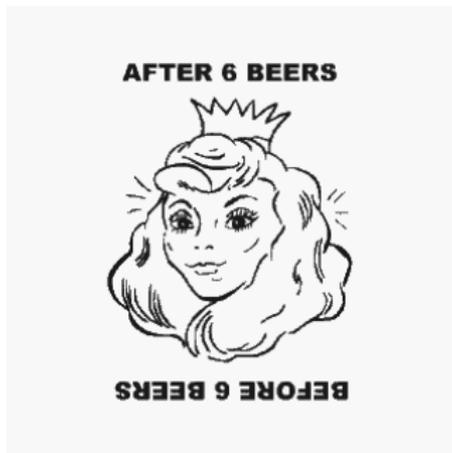
# Топологические пространства с «ограничениями»



# «Сила» топологической эквивалентности



# Субъективность восприятия : проекционная семантика формы



# Основной вопрос формирования «экосистемы машинного обучения»

**Что необходимо:** **целостное** восприятие данных, порождаемых процессами реального мира («**знание – сила**»).

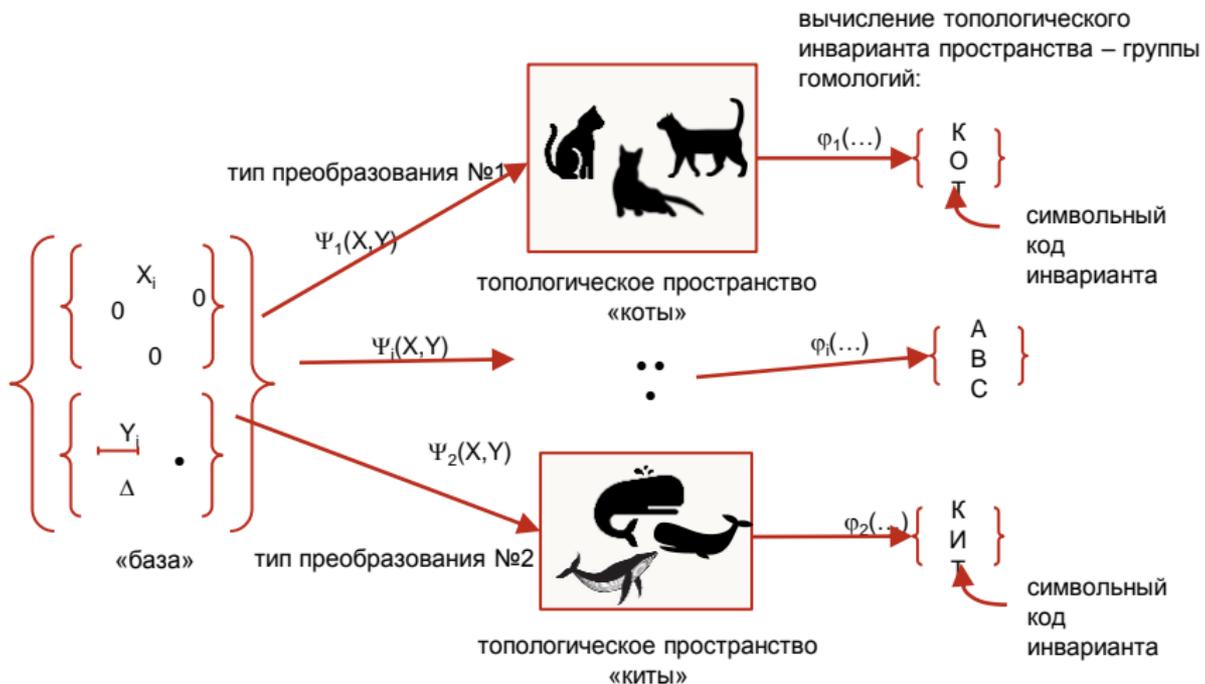
**Как это сделать:** «необходимо понимать ... то, чего в **целостном** виде нет в отдельных деталях конструкции, используя для предыдущий опыт и теоретические знания».



## Можно ли найти ответ на фундаментальный вопрос:

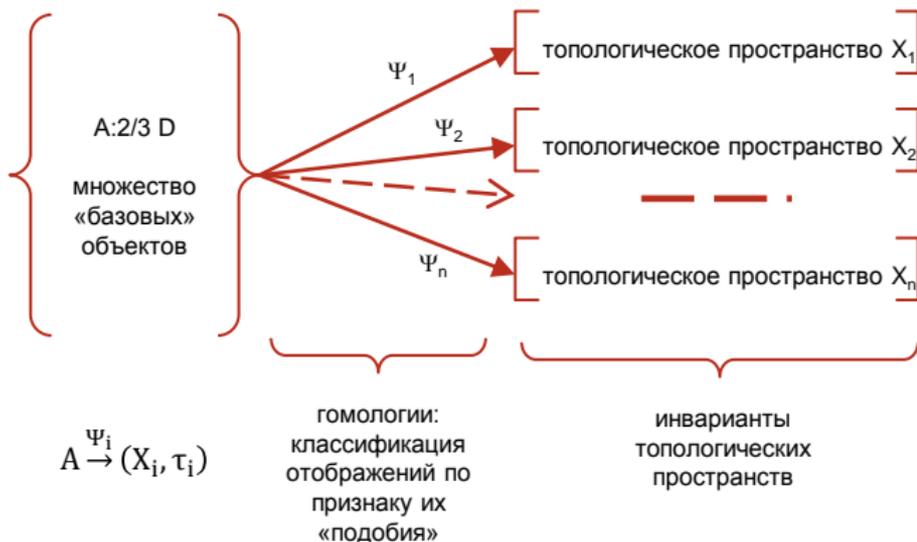
- как в одном топологическом пространстве  $(X, \tau)$  отличить одну конфигурацию точек (пикселей, вокселей, симплексов, ...) от другой конфигурации
  - можно ли конкретному множеству топологических структур, имеющих границы, сопоставить содержательное понятие (цепной комплекс, граница, ..., «кот», «кит»)
  - может ли граница совпадать с самим объектом  $\left( \begin{array}{l} \partial\partial = 0 \\ \partial: Z_n = Z_{n-1} \end{array} \right)$
- образует ли множество содержательных понятий топологическое (структурированное) пространство (понятие как открытое множество)

# Конструирование базы топологии



# «генераторы» топологического пространства

К «базе» применяются различные виды отображений  $\{ \Psi_i \}$ , задающих топологию



## Структурогенез как результат «топологических ограничений»

Процесс формирования структуры (формы) - структурогенеза описывается через понятие **супервентности** (англ. Supervenience) — отношение детерминированности (полной определенности) состояния одной системы состоянием другой системы. Реальность это совокупность «среды» и «объектов».

Считается, что набор свойств объекта **супервентен** относительно набора свойств среды в том случае, если существование различия между двумя объектами невозможно без существования такого же различия между двумя фактами в свойствах среды.

Именно, через свойство супервентности возможно описание **объективной зависимости ментальных явлений от физических явлений**,

## Выводы, которые важны для практики

феномен «supervenience» цифровых технологий проявляется в том, что имеют место

- **Детерминированность (отсутствие различий)** в информационных свойствах вычислителя при отсутствии различий в структуре и физических свойствах компонент;
- **Топологичность (непрерывность преобразований)** компьютерной программы в коды машинных команд при отсутствии различий в аппаратной конфигурации компьютера;
- **Связанность (отсутствие различий) в результатах выполнения программы** при отсутствии различий в структуре входных данных