

Дрынкин Роберт Робертович

Москва, 2019

## Содержание

1	Введение	3
2	Ликбез по топологическому анализу данных	3
2.1	Базовые определения	3
2.2	Группы гомологий	4
2.3	Устойчивые гомологии	5
2.4	Еще один взгляд на устойчивые гомологии	5
2.5	Полезные утверждения	6
3	Мозговые данные	6
3.1	Какие есть данные	6
3.2	Статья о распознавании лабиринта	6
4	Первый подход к данным	7
4.1	Проверка гипотезы о наличии информации о местоположении в данных	7
4.2	Поиск информативных нейронов	7
5	Второй подход к данным	8
5.1	Проверка статьи	8
5.2	Что пошло не так?	8
6	Результаты и выводы	9
6.1	Вопросы Владимира Сотскова	9

## Список иллюстраций

1	85% квантиль активности 20 лучших нейронов с точки зрения взаимной информации с позицией мыши	12
2	t-SNE вложение активности 20 лучших нейронов	13
3	85% квантиль активности лучших нейронов с точки зрения взаимной информации с позицией мыши, у которых нет больших не связанных областей активности	13

Аннотация:

Топологический анализ данных предоставляет инструментарий для исследования топологических пространств численными методами. В работе эти методы применяются к задаче извлечения топологических характеристик физического пространства по активности нейронов места в гиппокампе мыши. По данным нейрональной активности и видеозаписи положения мыши в лабиринте среди всех наблюдаемых нейронов выделены наиболее правдоподобные кандидаты в нейроны места. Далее, с полученным списком нейронов места и без видеозаписи положения мыши, удалось проверить описанный ранее теоретический метод и границы его применимости. В теории метод основан на теореме Александрова о нерве покрытия и позволяет описать форму физического пространства по активности нейронов места. При этом в теории предполагается, что нейроны места активируются в выпуклой области пространства. На практике оказалось, что это не обязательно так, и из-за этого теорема Александрова не применима.

## 1 Введение

Числа Бетти топологического пространства – это инвариант, индексирующий количество  $i$ -мерных пустот в пространстве. Нулевое число Бетти – это число компонент связности топологического пространства. У графов первое число индексирует количество независимых циклов в графе. У дерева первое число Бетти равно 0, у окружности – 1, и так далее. Топологический анализ данных позволяет работать анализировать информацию методами, заимствованными из топологии. При этом в отличие от классической алгебраической топологии, методы прикладной топологии позволяют отсекаать шумы в данных. В работе методы анализа данных и топологического анализа данных применяются к задачам, происходящим из нейробиологии. Известно, что в гиппокампе млекопитающих существуют отдельные нейроны, которые активируются, когда животное находится в определенной области физического пространства. В работах [3] и [4] описан теоретический алгоритм, как восстановить топологию физического пространства по данным об активностях набора нейронов места. Этот алгоритм опирается на теорему Александрова о нерве покрытия, одну из основных теорем гомотопической топологии. Мы проверяем теоретический алгоритм и границы его применимости на реальных данных мозговой активности мыши, полученных Владимиром Сотсковым и его коллегами из Института перспективных исследований мозга МГУ. Для этого сначала, с помощью обычного анализа данных, проверили, что есть связь между нейрональной активностью мыши и ее положением в лабиринте. Затем отобрали 20 нейронов, которые имеют наибольшую взаимную информацию с положением мыши. Попробовали применить описанный в статьях алгоритм, чтобы посчитать числа Бетти, однако это не вышло, так как часть отобранных нейронов имела 2 большие несвязные области активности. После того как оставили нейроны только с одной связной областью активности, получили результаты соответствующие статьям.

## 2 Ликбез по топологическому анализу данных

В этой секции будут введены основные определения касающиеся топологии в целом, симплициальных комплексов, их групп гомологий. Обо все этом можно почитать подробнее в книгах [1] и [2]. Тут же будут введены понятия фильтрации симплициальных комплексов и устойчивых гомологий, об этом можно почитать в книге [2].

### 2.1 Базовые определения

Определение 1 (Топологическое пространство)

Топологическое пространство – это множество  $X$ , с выделенным набором подмножеств  $T \subseteq 2^X$ , называемых открытыми, которые удовлетворяют следующим условиям

1.  $\emptyset, X \in T$
2. объединение любого набора открытых множеств является открытым
3. пересечение конечного числа открытых множеств является открытым

$T$  называют топологией пространства  $X$

Определение 2 (Непрерывное отображение)

Пусть даны 2 топологических пространства  $(X, T_X), (Y, T_Y)$ , тогда отображение  $f : X \rightarrow Y$  непрерывно, если  $\forall U \in T_Y f^{-1}(U) \in T_X$ .

Определение 3 (Гомеоморфизм)

Отображение топологических пространств  $f : X \rightarrow Y$  – называется гомеоморфизмом, если  $f$  и  $f^{-1}$  – непрерывны.

Определение 4 (Индукцированная топология)

Пусть дано топологическое пространство  $(X, T_X)$ , тогда на любом подмножестве  $Y \subseteq X$  можно ввести индуцированную с  $X$  топологию  $T_Y = \{U \cap Y | U \in T_X\}$ .

В дальнейшем, если не оговорено противного, на всех множествах, вкладываемых в  $\mathbb{R}^n$  будем считать, что топология индуцирована с  $\mathbb{R}^n$ .

Определение 5 (Гомотопные отображения)

Отображения  $f, g : X \rightarrow Y$  – называются гомотопными (обозначается  $f \simeq g$ ), если существует непрерывное отображение  $F : X \times I \rightarrow Y$ , такое, что  $F(x, 0) = f(x)$ ,  $F(x, 1) = g(x)$ .

Гомотопия является отношением эквивалентности на отображениях.

Определение 6 (Гомотопическая эквивалентность)

Топологические пространства  $X, Y$  – гомотопически эквивалентны, если существуют  $f : X \rightarrow Y, g : Y \rightarrow X$ , такие, что  $f \circ g \simeq id_X, g \circ f \simeq id_Y$ .

Гомотопическая эквивалентность является отношением эквивалентности на топологических пространствах, гомотопическим типом  $X$  называется класс пространств гомотопически эквивалентных  $X$ .

Определение 7 (Симплекс)

$n$ -мерный симплекс  $\sigma^n$  – это выпуклая оболочка упорядоченного набора из  $n + 1$  точки  $v_0 \dots v_n \in \mathbb{R}^n$ , не лежащих в одной  $(n - 1)$  мерной плоскости.

Определение 8 (Грань симплекса)

$m$ -мерной гранью симплекса  $\sigma^n = [v_0, \dots, v_n]$ , называется симплекс на вершинах  $[v_{i_1}, \dots, v_{i_m}]$ .

Определение 9 (Симплициальный комплекс)

Симплициальный комплекс – это набор симплексов  $\sigma_1, \dots, \sigma_n$ , таких что  $\forall \sigma_i, \sigma_j$  либо  $\sigma_i \cap \sigma_j = \emptyset$ , либо  $\sigma_i \cap \sigma_j$  является гранью  $\sigma_i$  и  $\sigma_j$ .

Определение 10 (Триангуляция)

Триангуляция пространства  $X$  – это симплициальный комплекс  $S_X$  гомеоморфный  $X$ . Говорят, что на  $X$  задана структура симплициального комплекса, имея в виду, что задана его триангуляция.

## 2.2 Группы гомологий

Определение 11 (Симплициальная цепь)

Пусть  $X$  – симплициальный комплекс,  $A$  – абелева группа, тогда  $C_n(X, A)$  – абелева группа, порожденная  $n$ -мерными симплексами  $X$ , ее элементы называют симплициальными цепями, они имеют вид конечной формальной суммы  $\sum_i k_i \sigma_i^n, k_i \in A$ .

Определение 12 (Граничный гомоморфизм)

Граничный гомоморфизм  $C_n(X)$  это отображение  $\partial_n : C_n(X) \rightarrow C_{n-1}(X)$ . Покажем, как оно действует на базисных элементах:

$$\partial_n(\sigma_i) = \sum_j (-1)^j \sigma_i|_{[v_0 \dots \hat{v}_j \dots v_n]}$$

Где  $\sigma_i|_{[v_0 \dots \hat{v}_j \dots v_n]}$  – это  $n - 1$  мерная грань симплекса  $\sigma_i$ , полученная опусканием вершины  $v_j$ .

Утверждение 1 Не сложно показать, что композиция  $C_n(X) \xrightarrow{\partial_n} C_{n-1}(X) \xrightarrow{\partial_{n-1}} C_{n-2}(X)$  является нулевым отображением.

Таким образом у нас имеется точная последовательность абелевых групп  $\{C_i, \partial_i\}$ , называемая цепным комплексом:

$$\dots \rightarrow C_{i+1} \xrightarrow{\partial_{i+1}} C_i \xrightarrow{\partial_i} C_{i-1} \rightarrow \dots \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

В силу утверждения 1  $\text{Im} \partial_{n+1} \subseteq \text{Ker} \partial_n$ .

### Определение 13 (Группы гомологий)

В такой ситуации определим  $n$ -ую группу гомологий симплициального комплекса, как

$$H_n = \text{Ker } \partial_n / \text{Im } \partial_{n+1}$$

Утверждение 2 Гомологии являются функтором из категории топологических пространств в категорию абелевых групп, то есть непрерывное отображение  $f : X \rightarrow Y$  индуцирует гомоморфизм  $\hat{f} : H_n(X) \rightarrow H_n(Y)$ , при это если  $f$  – гомотопическая эквивалентность, то  $\hat{f}$  – изоморфизм.

Следствие 1 Любые 2 триангуляции одного пространства имеют одинаковые группы гомологий.

### Определение 14 (Числа Бетти)

$\beta_k = \text{rank}(H_k)$  – кое число Бетти

Как уже говорилось выше,  $k$ -ое число Бетти индексирует  $k$ -мерные пустоты в пространстве.

## 2.3 Устойчивые гомологии

### Определение 15 (Фильтрация симплициальных комплексов)

Фильтрация симплициальных комплексов – это последовательность вложенных друг в друга симплициальных комплексов.

$$S_1 \subset \dots \subset S_n$$

Пусть  $S_i = S_{i-1} \cup \delta_i$ , где  $\delta_i$  – множество симплексов, тогда говорят, что симплексы  $\delta_i$  появились в момент времени  $i$ .

### Определение 16 (Устойчивые гомологии)

Пусть дана фильтрация  $K = \{S_0 \subset \dots \subset S_n\}$ ,  $C^i$  – цепной комплекс  $i$ -ого симплициального комплекса, обозначим:  $Z_k^i = \text{Ker}(C_k^i \xrightarrow{\partial_k} C_{k-1}^i)$ ,  $B_k^i = \text{Im}(C_{k+1}^i \xrightarrow{\partial_{k+1}} C_k^i)$

Тогда  $p$ -устойчивыми  $k$ ми гомологиями называется:

$$H_k^{i,p} = Z_k^i / (Z_k^i \cap B_k^{i+p})$$

Не трудно заметить, что циклы гомологичные в  $S_i$ , гомологичны и в  $S_{i+p}$ . Пусть  $H_k^i = H_k(S_i)$ , можно рассмотреть естественный гомоморфизм  $\eta_k^{i,p} : H_k^i \rightarrow H_k^{i+p}$ , отображающий класс гомологичных циклов, в класс содержащий их, тогда  $p$ -устойчивые гомологии можно определить по другому:

$$H_k^{i,p} = \text{Im } \eta_k^{i,p}$$

Допустим не граничный  $k$ -цикл  $z$  появился в момент времени  $i$ , при добавлении симплекса  $\sigma$ . Класс циклов  $[z]$ , гомологичных  $z$ , является элементом группы  $H_k^i$ . Допустим в момент времени  $j > i$ , появился симплекс  $\tau$ , превративший цикл  $z$  в границу, то есть  $z \in B_k^j$ , в этот момент уменьшился ранг  $k$ ых гомологий. Говорят, что цикл  $[z]$  существовал с момента времени  $i$ , до момента  $j$ , его время жизни  $j - i - 1$ . Говорят, что цикл  $z$  был создан симплексом  $\sigma$ , и уничтожен симплексом  $\tau$ ,  $\sigma$  – создатель  $z$ ,  $\tau$  – уничтожитель  $z$ .

## 2.4 Еще один взгляд на устойчивые гомологии

Пусть дана фильтрация  $F_0 \subset F_1 \subset \dots$ , рассмотрим симплициальные цепи этой фильтрации над кольцом  $\mathbb{K}$ :

$$\begin{array}{ccccccc} & & \vdots & & \vdots & & \\ & & \uparrow f_1 & & \uparrow f_1 & & \\ 0 & \xleftarrow{\partial_0} & C_0(F_1, \mathbb{K}) & \xleftarrow{\partial_1} & C_1(F_1, \mathbb{K}) & \xleftarrow{\partial_2} & \dots \\ & & \uparrow f_0 & & \uparrow f_0 & & \\ 0 & \xleftarrow{\partial_0} & C_0(F_0, \mathbb{K}) & \xleftarrow{\partial_1} & C_1(F_0, \mathbb{K}) & \xleftarrow{\partial_2} & \dots \end{array}$$

$$C_j(F_\bullet, \mathbb{K}) = \bigoplus_{i=0}^{\infty} C_j(F_i, \mathbb{K}) - \text{градуированный } \mathbb{K} \text{ модуль}$$

Оказывается  $C_j(F_\bullet, \mathbb{K})$  является также градуированным  $\mathbb{K}[x]$  модулем:

$$\forall \sigma \in C_j(F_i, \mathbb{K}) : x^l \cdot \sigma = (f_{i+l-1} \circ \dots \circ f_{i+1} \circ f_i)(\sigma)$$

$\partial_j : C_j(F_\bullet, \mathbb{K}) \rightarrow C_{j-1}(F_\bullet, \mathbb{K})$  – гомоморфизмы  $\mathbb{K}[x]$  модулей.

Значит гомологии  $H_j(F_\bullet, \mathbb{K})$  – являются  $\mathbb{K}[x]$  модулями.

Если  $\mathbb{K}$  – поле, то

$$H_j(F_\bullet, \mathbb{K}) = \left( \bigoplus_{i=1}^r \Sigma^{\alpha_i} \mathbb{K}[x] \right) \oplus \left( \bigoplus_{j=1}^s \Sigma^{\beta_i} \mathbb{K}[x] / (x^{\eta_i}) \right)$$

Тогда можно рассмотреть что происходит в каждой градуировке (в каждый момент времени):  $\Sigma^{\alpha_i} \mathbb{K}[x]$  – цикл появившийся в момент времени  $\alpha_i$ , который никогда не исчезнет,  $\bigoplus_{j=1}^s \Sigma^{\beta_i} \mathbb{K}[x] / (x^{\eta_i})$  – цикл появившийся в момент времени  $\beta_i$  и исчезающий в момент времени  $\beta_i + \eta_i - 1$ .

## 2.5 Полезные утверждения

Определение 17 (Нерв покрытия)

Пусть  $X$  – топологическое пространство и  $\{U_i\}$  – его покрытие. Нервом покрытия называется симплициальный комплекс  $S$  на вершинах  $\{v_i\}$ , соответствующих множествам из покрытия, при этом:

$$[v_{i_1}, \dots, v_{i_k}] \in S \iff \bigcap_{j=1}^k U_{i_j} \neq \emptyset$$

Теорема 1 (Теорема Александрова / Нерв лемма)

Пусть  $X$  – триангулируемое топологическое пространство и  $\{U_i\}$  – его покрытие замкнутыми множествами, при этом все  $U_i$  и их пересечения стягиваемы, тогда нерв данного покрытия гомотопически эквивалентен  $X$ .

Доказательство может быть найдено в книге [1].

Следствие 2 Числа Бетти пространства и его нерва совпадают.

## 3 Мозговые данные

### 3.1 Какие есть данные

Есть запись мозговой активности мыши, которая 10 минут ходит по круговому лабиринту. 10 раз в секунду записывается состояние активности 590 нейронов. Также есть вручную отобранные биологами 24 лучших (наиболее информативных на глаз) нейронов.

### 3.2 Статья о распознавании лабиринта

В статьях [3] и [4] предлагается следующий подход к восстановлению чисел Бетти лабиринта по которому бегают мышь: сопоставим каждому нейрону места вершину комплекса, на множество вершин натянем симплекс, если они были одновременно активны, в силу теоремы Александрова числа Бетти полученного комплекса должны совпадать с числами Бетти лабиринта, по которому бегают мышь.

Тут возникают разные вопросы о том когда нейрон считать активным, как бороться с шумом и выбросами в данных и так далее, постараемся ответить на эти вопросы далее.

## 4 Первый подход к данным

Для начала, произвольными средствами, попробуем понять есть ли в данных хоть какая-то информация о местоположении мыши в лабиринте, чтобы понять, имеет ли смысл вообще проверять какие-то гипотезы основываясь на этих данных (так как, например, биологи могли просто не попасть в нейроны места). А так же постараемся каким нибудь более общим методом (чем вручную) отобрать наиболее информативные нейроны.

### 4.1 Проверка гипотезы о наличии информации о местоположении в данных

Для проверки гипотезы, попробуем научиться с хорошей точностью предсказывать местоположение мыши. Для этого:

1. Разделим окружность на 10 равных сегментов и будем решать задачу предсказания в каком из этих сегментов находится мышь. В качестве метрик будем использовать:
  - (а) 1-Точность: отношение правильных предсказаний ко всем
  - (б) 3-Точность: отношение количества предсказаний, когда правильный ответ, либо предсказан точно, либо это соседний участок, ко всем
2. Разделим все данные на 2 части: обучающую (первые 70% времени) и тестовую (оставшиеся 30%)
3. На обучающей части выборки выучим какой-нибудь алгоритм, в нашем случае градиентный бустинг на разрешающих деревьях (GBDT - Gradient Boosting Descision Trees [6]), а точнее одну из его реализаций – catboost<sup>1</sup> [7]
4. Применим обученный алгоритм на тестовой выборке и сравним точность его предсказаний с разумными эвристиками, не использующими активность нейронов

Таблица 1: Проверка наличия информации в нейронах о местоположении

Алгоритм	1-Точность	3-Точность
GBDT на всех данных	0.645	0.978
GBDT на 24 нейронах отобранных в ручную	0.470	0.957
GBDT на 20 нейронах отобранных автоматически	0.606	0.974
Случайный из распределения обучающей выборки	0.145	0.362
Случайный из распределения тестовой выборки	0.186	0.358
Лучший константный прогноз	0.368	0.521

В итоге получили следующие результаты: 1. Можно увидеть, что градиентный бустинг с почти 100% вероятностью может предсказывать в какой трети окружности находится мышь.

### 4.2 Поиск информативных нейронов

Для начала посмотрим, насколько ухудшится результат, если обучить градиентный бустинг не на всех данных, а на 24 в ручную отобранных нейронах: таблица 1. Можно увидеть, что 1-точность сильно ухудшилась, почти приблизившись к наилучшему константному прогнозу, при этом 3-точность не так сильно уменьшилась.

Теперь попробуем как-нибудь автоматически набрать 20 нейронов, так чтобы результат не ухудшился (он улучшится). Для этого возьмем активность каждого нейрона, и сопоставим ей дискретную случайную величину, которая принимает значение  $i$ , если активность данного нейрона находится в  $[\frac{i}{10}; \frac{i+1}{10}]$  квантиле. Затем отсортируем нейроны по количеству взаимной информации между соответствующей ему случайной величиной и сегментом в котором находится мышь. И возьмем первые 20 из них (см. рис. 1). Мощность полученного множества нейронов, с отобранными биологами равна 4.

Отступление про абсолютные значения взаимных информаций: энтропия сегмента в котором находится мышь равна 3.11, у наиболее информативного нейрона взаимная информация равна 0.67, у 20-ого по информативности 0.487.

<sup>1</sup><https://catboost.ai>

Для начала попробуем “на глаз” проверить есть ли какая-нибудь связь между активностью нейронов и положением мыши, однако из-за того что мы отобрали 20 нейронов сделать это явно может быть затруднительно, попробуем придумать вложение  $\mathbb{R}^{20} \rightarrow \mathbb{R}^2$ , наилучшим образом сохраняющее расстоянием, и изобразим его на плоскости (что человеку анализировать уже существенно проще). Для этого воспользуемся алгоритмом t-SNE [3]. Вложим все данные на плоскость и раскрасим близкие по положению мыши точки похожими цветами, и нарисуем 2 графика: вложение данных первых 70% времени и последних 30%, и посмотрим на сколько эти графики похожи. Идея такова: в одних и тех же областях обоих графиков должны быть точки похожих цветов, это будет означать что похожая нейронная активность была в близких областях. Получили рис. 2, на нем видно, что наши предположения оправдались и действительно на обоих графиках одни и те же части покрашены в похожие цвета.

Теперь снова попробуем выучить градиентный бустинг на этих 20 нейронах, чтобы сравнить со вручную отобранными нейронами: получим и 1-точность и 3-точность близкую ко всем нейронам (существенно большую, чем на тех 24 отобранных вручную).

## 5 Второй подход к данным

### 5.1 Проверка статьи

Теперь попробуем сделать то, что описано в статье 3.2. Все описанное в ней работает, утверждается, что работает для нейронов “места”, такими нейронами будем считать эти 20. Они отобраны по количеству взаимной информации с позицией мыши.

Отождествим нейроны с вершинами симплициального комплекса. Положим  $s_k^t$  – симплекс, натянутый на вершины, соответствующие нейронам, уровень активности которых превышает  $(1 - k)$  квантиль их активности в момент времени  $t$ . Построим симплициальный комплекс  $S_k = \bigcup_t s_k^t$

Построим фильтрацию

$$S_{k_0} \subset (S_{k_0} \cup S_{k_1}) \subset \dots \subset \bigcup_{i=0}^K S_{k_i}, l_k = k_0 < \dots < k_K = r_k$$

Сосчитав устойчивые гомологии (с помощью python библиотеки Dionysus <sup>2</sup>) для такой фильтрации, взяв

$$l_k = 0.05 = k_0 < k_0 + 0.01 = k_1 < \dots < k_i < k_i + 0.01 < \dots < k_K = r_k = 0.5$$

получаем один нульмерный цикл, живущий все время, и несколько циклов больших размерностей (1-4), живущих существенно меньше (таблица 2)

Таблица 2: Циклы, отсортированные по времени жизни

Размерность цикла	0	2	2	2	4	3	3	2	2	3	...
Время появления цикла	0.05	0.07	0.10	0.13	0.27	0.19	0.22	0.12	0.10	0.14	...
Время уничтожения цикла	$\infty$	0.22	0.24	0.25	0.39	0.27	0.30	0.20	0.18	0.22	...
Время жизни цикла	0.45	0.15	0.14	0.12	0.12	0.08	0.08	0.08	0.08	0.08	...

Заметим, что в такой фильтрации получается очень много симплексов, которые за все время эксперимента живут очень мало (меньше 0.5% времени), скорее всего это просто шум. Попробуем с ним бороться следующим способом: будем рассматривать исходную фильтрацию, в которой оставим только симплексы живущие дольше  $l\%$  времени. Сосчитаем устойчивые гомологии для различных  $l$  (таблица 3). Из таблицы видно, что для различных  $l$  получается 2-3 1-мерных цикла, живущих существенно дольше других. Это не то, чего мы ожидали, далее попробуем разобраться почему так могло произойти.

### 5.2 Что пошло не так?

Подумаем, почему так могло произойти, для этого вспомним теорема Александрова 1. Одним из ее условий была стягиваемость множеств, покрывающих пространство, заметим, что есть нейроны, которые активны в 2 больших нестягиваемых областях окружности. Гипотеза состоит в том, что эти нейроны нам все и портят.

<sup>2</sup><https://www.mrzv.org/software/dionysus/>



Проверим ее, для этого выкинем нейроны, у которых больше одной большой компоненты связности (рис. 3) и способом описанным выше сосчитаем устойчивые гомологии (таблица 4). Из полученной таблицы видно, что действительно получается один долгоживущий 1-мерный цикл, как и предполагалось в статье.

## 6 Результаты и выводы

Удалось выделить набор нейронов, которые имеют наибольшее количество информации о местоположении мыши (лучше, чем вручную). Однако не все из них имели связную область активности.

В целом, подход описанный в статье работает, но только для очень хороших нейронов места (со стягиваемой областью активности). Однако задача *unsupervised* нахождения таких нейронов выглядит крайне нетривиальной, а *supervised* (как в данной работе) не имеет большого смысла, так как мы предполагаем, что нам заранее координаты мыши, что делает задачу определения топологии пространства совсем тривиальной и без использования активности нейронов.

Возникла новая задача: попытаться отслеживать те “нейроны места”, у которых область загорания связная. Иначе нерв-лемма неприменима. С математической точки зрения, конечно, это понятно: если элементы покрытия нестягиваемы, то нерв-лемма не работает. Важно, что “неправильность” полученного результата можно увидеть из устойчивых гомологий, т.е. они не записывают эту неправильность в шум. По сути это означает, что все работает так как должно, и заявленные нам задачи имеют смысл.

### 6.1 Вопросы Владимира Сотскова

Q:

А что если взять и построить такие же активность не всех клеток, а только тех, кто активизируется, скажем, в половине трека (например, только в левой или в правой)?

A:

Были взяты 5 хороших нейронов активных в одной половине круга, у них получился один нульмерный цикл все остальные циклы живут очень мало.

Q:

Или, в качестве ещё одного контроля, взять активность вообще заведомо неспецифичных клеток? По идее, в этих случаях никакого топологического сигнала мы наблюдать не должны.

A:

Получаются только нульмерные циклы, причем чем больше таких нейронов брать, тем меньше циклов получается. (в итоге, если взять их достаточно много, то должна получиться просто точка, но если брать больше 60 нейронов оно считается долго, подсчет ещё не провели.

## Список литературы

- [1] Hatcher, A. (2002). Algebraic Topology. Cambridge: Cambridge University Press
- [2] Zomorodian, A. (2005). Contents. In Topology for Computing (Cambridge Monographs on Applied and Computational Mathematics, pp. Vii-X). Cambridge: Cambridge University Press.
- [3] Yuri I. Manin (2015). Neural codes and homotopy types: mathematical models of place field recognition arXiv:1501.00897.
- [4] Curto C, Itskov V (2008). Cell Groups Reveal Structure of Stimulus Space. PLoS Comput Biol 4(10): e1000205. doi:10.1371/journal.pcbi.1000205
- [5] van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research, 9(nov), 2579-2605.
- [6] Tianqi Chen and Carlos Guestrin (2016). XGBoost: A Scalable Tree Boosting System. arXiv:1603.02754
- [7] Anna Veronika Dorogush and Andrey Gulin and Gleb Gusev and Nikita Kazeev and Liudmila Ostroumova Prokhorenkova and Aleksandr Vorobev (2017). CatBoost: unbiased boosting with categorical features.

Таблица 3: Циклы, отсортированные по времени жизни

Минимальный процент вре- мени жизни цикла	Устойчивые гомологии									
0.82%	Размерность цикла	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	...
	Время появления цикла	0.05	0.30	0.31	0.05	0.05	0.05	0.05	0.05	...
	Время уничтожения цикла	$\infty$	$\infty$	$\infty$	0.18	0.18	0.13	0.13	0.12	...
	Время жизни цикла	0.45	0.20	0.19	0.13	0.13	0.08	0.08	0.07	...
0.65%	Размерность цикла	0.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	...
	Время появления цикла	0.05	0.23	0.28	0.09	0.21	0.05	0.05	0.05	...
	Время уничтожения цикла	$\infty$	$\infty$	0.47	0.17	0.28	0.11	0.09	0.09	...
	Время жизни цикла	0.45	0.27	0.19	0.08	0.07	0.06	0.04	0.04	...
0.49%	Размерность цикла	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	...
	Время появления цикла	0.05	0.13	0.11	0.19	0.34	0.07	0.05	0.12	...
	Время уничтожения цикла	$\infty$	$\infty$	0.39	0.41	$\infty$	0.22	0.18	0.19	...
	Время жизни цикла	0.45	0.37	0.28	0.22	0.16	0.15	0.13	0.07	...
0.33%	Размерность цикла	0.00	1.00	2.00	2.00	2.00	1.00	2.00	2.00	...
	Время появления цикла	0.05	0.11	0.30	0.31	0.26	0.07	0.34	0.39	...
	Время уничтожения цикла	$\infty$	$\infty$	$\infty$	$\infty$	0.45	0.24	$\infty$	$\infty$	...
	Время жизни цикла	0.45	0.39	0.20	0.19	0.19	0.17	0.16	0.11	...
0.16%	Размерность цикла	0	2	2	2	2	2	1	3	...
	Время появления цикла	0.05	0.16	0.19	0.20	0.23	0.17	0.06	0.36	...
	Время уничтожения цикла	$\infty$	0.42	0.42	0.39	0.42	0.34	0.20	$\infty$	...
	Время жизни цикла	0.45	0.26	0.23	0.19	0.19	0.17	0.14	0.14	...

Таблица 4: Циклы, отсортированные по времени жизни, с использованием нейронов со стягиваемой областью активности

Минимальный процент вре- мени жизни цикла	Устойчивые гомологии									
0.82%	Размерность цикла	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	...
	Время появления цикла	0.05	0.15	0.05	0.05	0.05	0.05	0.05	0.44	...
	Время уничтожения цикла	$\infty$	0.48	0.14	0.13	0.12	0.12	0.12	$\infty$	...
	Время жизни цикла	0.45	0.33	0.09	0.08	0.07	0.07	0.07	0.06	...
0.65%	Размерность цикла	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	...
	Время появления цикла	0.05	0.13	0.11	0.29	0.08	0.09	0.41	0.23	...
	Время уничтожения цикла	$\infty$	$\infty$	0.39	$\infty$	0.22	0.16	0.47	0.29	...
	Время жизни цикла	0.45	0.37	0.28	0.21	0.14	0.07	0.06	0.06	...
0.49%	Размерность цикла	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	...
	Время появления цикла	0.05	0.12	0.10	0.10	0.10	0.25	0.05	0.06	...
	Время уничтожения цикла	$\infty$	0.48	0.37	0.33	0.25	0.39	0.18	0.19	...
	Время жизни цикла	0.45	0.36	0.27	0.23	0.15	0.14	0.13	0.13	...
0.33%	Размерность цикла	0.00	1.00	1.00	2.00	1.00	2.00	2.00	1.00	...
	Время появления цикла	0.05	0.11	0.07	0.30	0.16	0.26	0.34	0.08	...
	Время уничтожения цикла	$\infty$	0.46	0.22	0.45	0.29	0.38	0.46	0.19	...
	Время жизни цикла	0.45	0.35	0.15	0.15	0.13	0.12	0.12	0.11	...
0.16%	Размерность цикла	0	2	1	2	1	1	2	1	...
	Время появления цикла	0.05	0.17	0.06	0.23	0.07	0.25	0.25	0.05	...
	Время уничтожения цикла	$\infty$	0.33	0.20	0.37	0.20	0.37	0.37	0.17	...
	Время жизни цикла	0.45	0.16	0.14	0.14	0.13	0.12	0.12	0.12	...

Рис. 1: 85% квантиль активности 20 лучших нейронов с точки зрения взаимной информации с позицией мыши

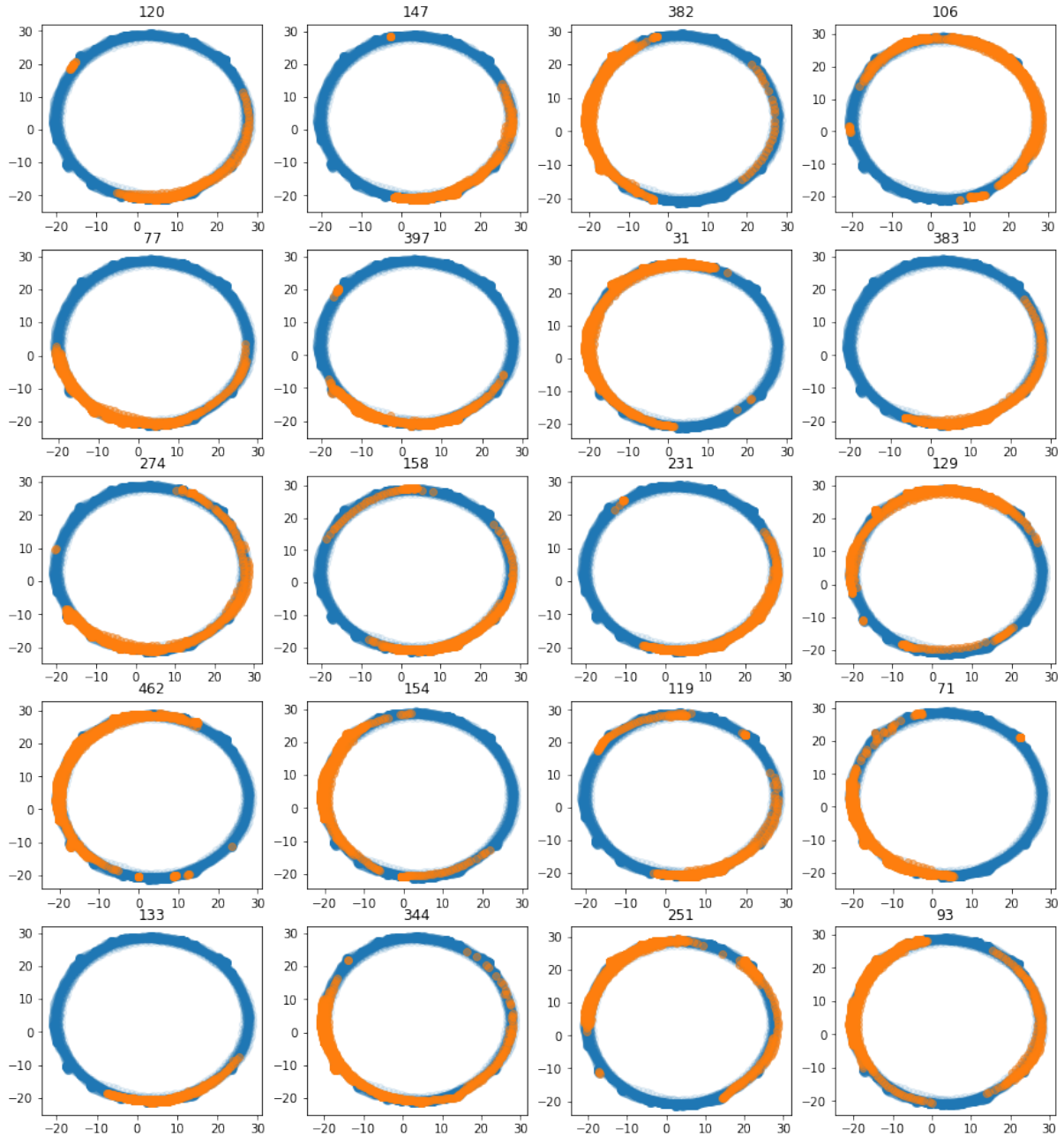


Рис. 2: t-SNE вложение активности 20 лучших нейронов

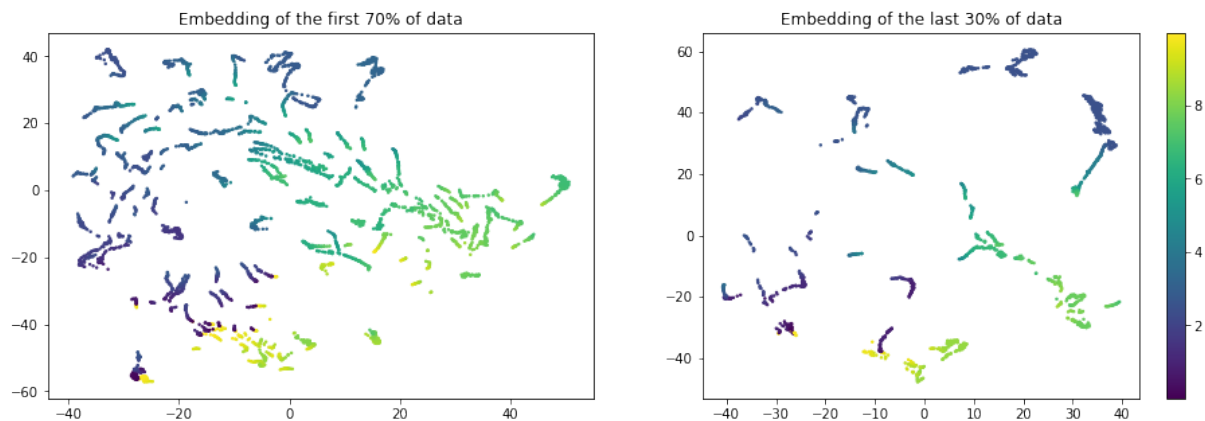


Рис. 3: 85% квантиль активности лучших нейронов с точки зрения взаимной информации с позицией мыши, у которых нет больших не связанных областей активности

