

# NumPy и матрицы



**Константин Башевой**

Аналитик-разработчик, Яндекс



**Константин Башевой**  
Аналитик-разработчик  
Яндекс

Помогаю аналитикам с инфраструктурой  
Собираю инструменты обработки данных  
Рассказываю как это весело

Последние 10 лет:

Rambler&Co

Ростелеком

Яндекс

# Что сегодня будет

# Программа на сегодня

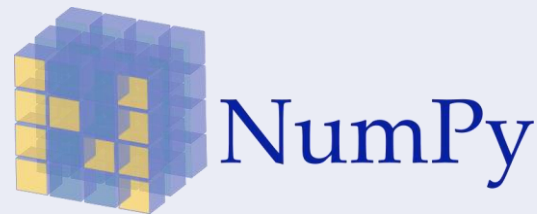
4



Метрики расстояний  
и векторы



Как это работает  
на примере  
фильмов

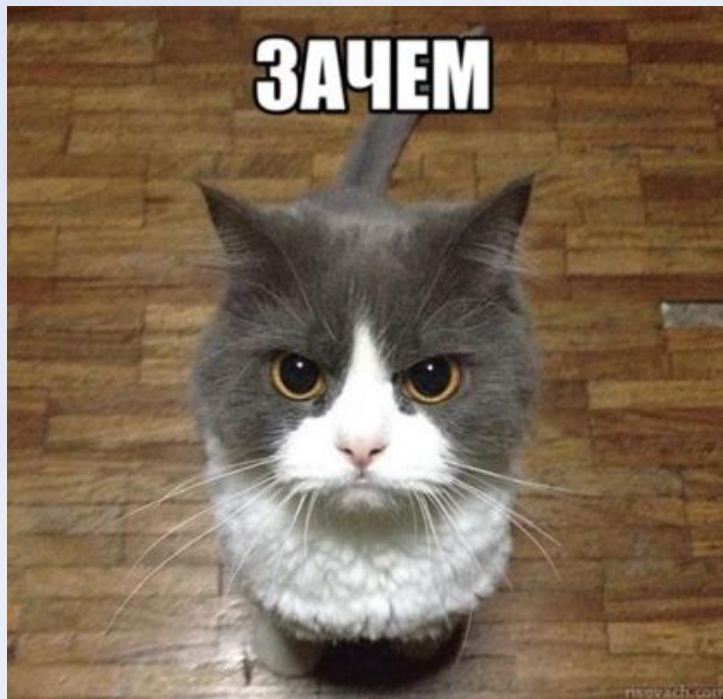


Практика в  
NumPy

# Метрики расстояний

# Зачем нам матрицы

6



# Как сравнивать объекты между собой

7

Москва



Сочи

1 631км

Владивосто



к

6 417км

# Как сравнивать объекты между собой

8

Москва



Сочи

1 631км

Владивосто



к

6 417км

Векторы



# Как сравнивать объекты между собой

9

Москва



Сочи

1 631км

Векторы

Москвы

Владивосто

к

6 417км

- Владивосток южнее Сочи
- Длина Сочи порядка 145км
- Парки и скверы занимают половину площади

# Применимы ли векторы к фильмам?

Какие из этих фильмов близки?

11



## Как будем сравнивать

12

Признак	Джон Уик	В бой идут одни старики	Сладкий ноябрь
Боевик / военный			
Мелодрама			
Трагический			
Новинка			
Отечественный			
Есть Киану Ривз			

## Как будем сравнивать

13

Признак	Джон Уик	В бой идут одни старики	Сладкий ноябрь
Боевик / военный	да	да	нет
Мелодрама	нет	нет	да
Трагический	нет	да	да
Новинка	да	нет	нет
Отечественный	нет	да	нет
Есть Киану Ривз	да	нет	да

## Выберем столбцы с признаками

14

Признаки	Боевик / военный	Мелодрам а	Трагическ ий	Новинка	Отечестве нный	Есть Киану Ривз
Джон Уик	1	0	0	2014	0	1
В бой идут одни старики	1	0	1	1973	1	0
Сладкий ноябрь	0	1	1	2001	0	1

# Рисуем векторы

15

боевик /  
военный



## Что круто в векторах

16

- ✓ Между векторами легко посчитать угол
- ✓ Векторы могут быть любой размерности
- ✓ Можно использовать алгоритмы кластеризации



## Пример для интернет-магазина

17

User ID	Смартфон	Smart TV	Ноутбук	Шкаф	Холодильник	Посуда
1	2	1	0	0	0	0
2	1	1	2	1	0	0
3	2	0	1	0	0	0
4	1	1	2	1	0	1
5	0	0	1	2	0	0
6	0	0	0	0	0	5
7	1	0	0	0	0	0
8	0	1	1	0	0	0
9	0	0	0	1	1	3
10	1	0	0	2	1	4