Современные методы анализа данных и машинного обучения

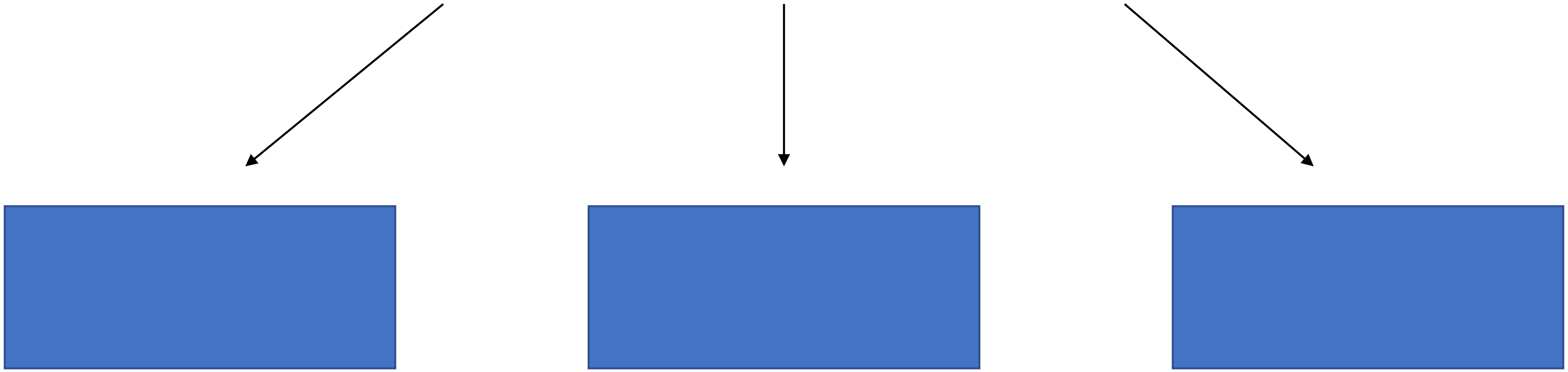
Тема 5. Лекция 5

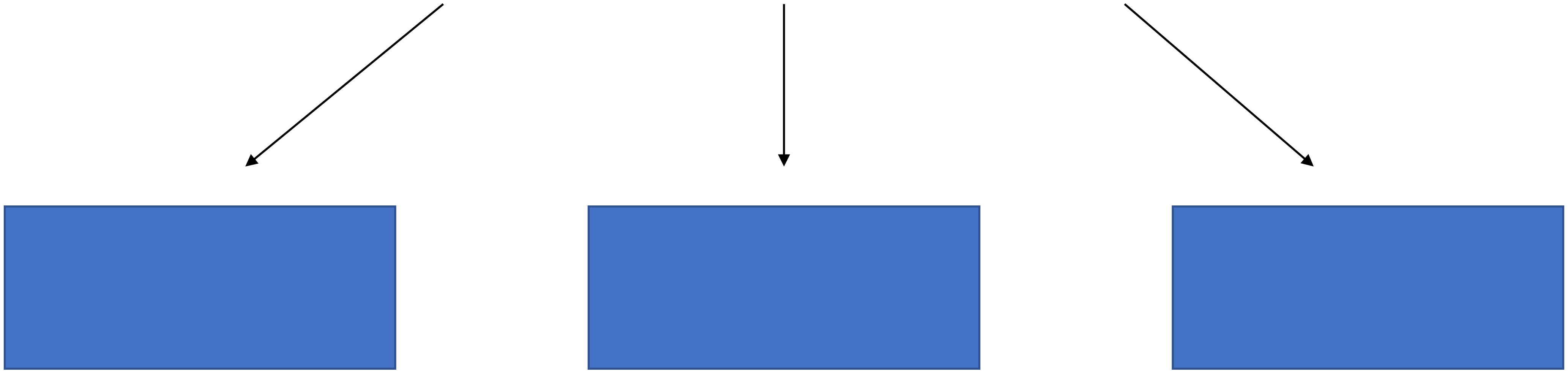
Математика для машинного обучения. Линейная алгебра

Юрий Саночкин

ysanochkin@hse.ru

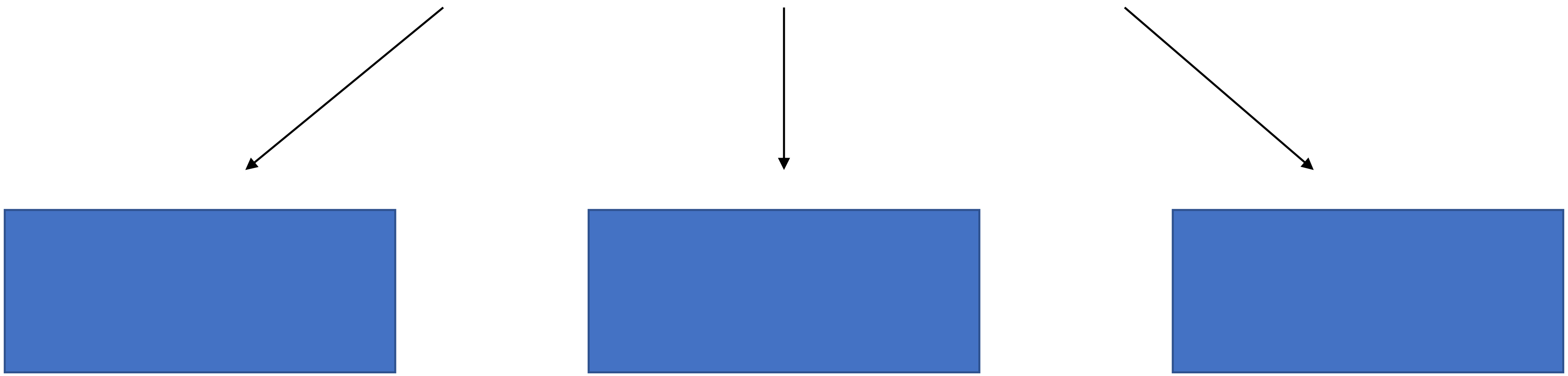
НИУ ВШЭ, 2024





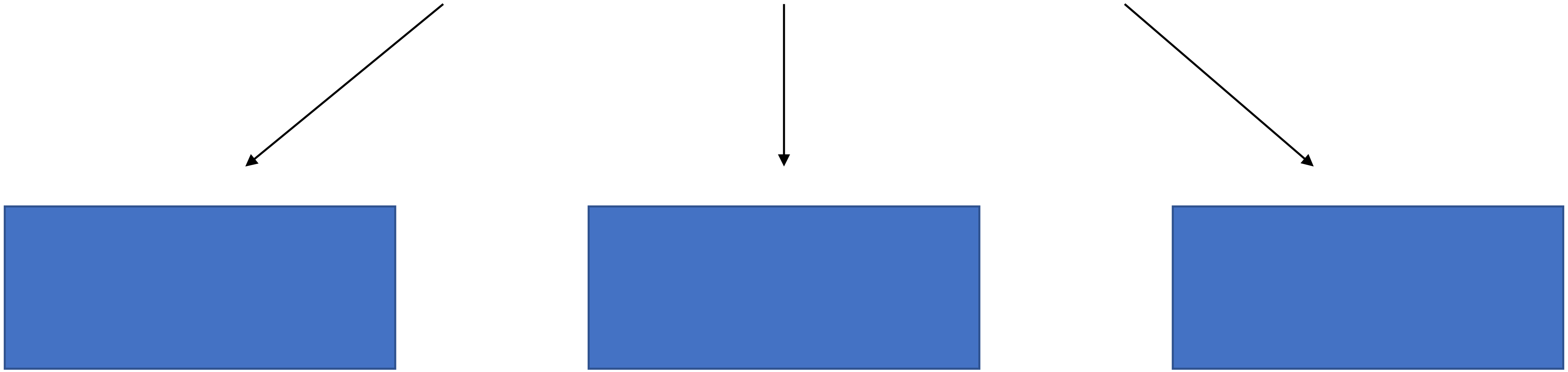
Знакомая картинка? :)
Думаете, знаете, что появится тут на следующем слайде? :)

Математика для анализа данных



Думаю, вы не угадали!

Математика для анализа данных



Какие разделы математики важны в анализе данных и для чего?

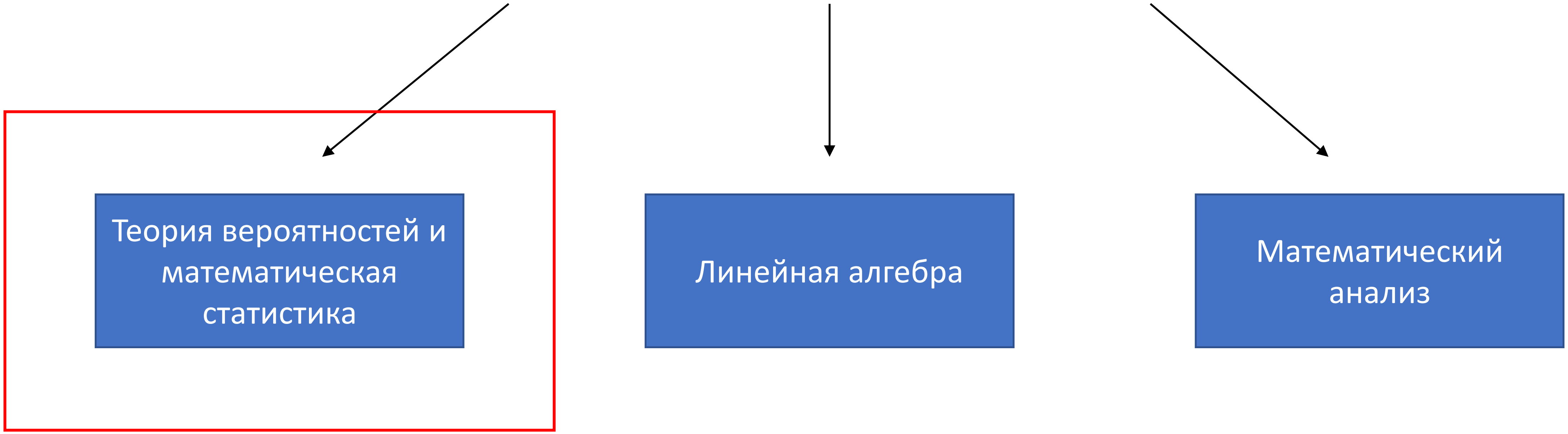
Математика для анализа данных

Математика для анализа данных

- Теория вероятностей и математическая статистика
- Линейная алгебра
- Математический анализ

Это три кита из разделов математики, на которых базируется весь анализ данных

Математика для анализа данных



Это мы уже вспомнили и обсудили достаточно подробно в рамках предыдущего модуля!

Математика для анализа данных

Теория вероятностей и математическая статистика

Линейная алгебра

Математический анализ

Сегодня наш разговор про это!

Математика для анализа данных

Теория вероятностей и математическая статистика

Линейная алгебра

Математический анализ

Это нас ждет впереди!

Линейная алгебра. Мотивация

Линейная алгебра. Мотивация

- Почему именно линейная алгебра?
- Про что этот раздел и почему он так важен для анализа данных и конкретно для машинного обучения?

Линейная алгебра. Мотивация

- За последние 30 лет мы научили машины
  - Смотреть и видеть;
  - Слушать, слышать и говорить в ответ;
  - Сочинять стихи и песни,
  - Прогнозировать котировки бирж;
  - Находить раковые опухоли
- и многое другое…

Линейная алгебра. Мотивация

- За последние 30 лет мы научили машины
  - Смотреть и видеть;
  - Слушать, слышать и говорить в ответ;
  - Сочинять стихи и песни,
  - Прогнозировать котировки бирж;
  - Находить раковые опухоли
- и многое другое…

- Диапазон приложений машинного обучения поражает воображение человека

Линейная алгебра. Мотивация

- Диапазон приложений машинного обучения поражает воображение человека

- Тем более удивительным кажется тот факт, что на всё это способна обычная вычислительная машина
- Фактически – просто продвинутый калькулятор

Линейная алгебра. Мотивация

- Предметно: компьютер это очень глупая коробка!
- Он понимает только нули и единицы

Линейная алгебра. Мотивация

- Предметно: компьютер это очень глупая коробка!
- Он понимает только нули и единицы
- …но делает это очень хорошо!
- Насколько хорошо?

Линейная алгебра. Мотивация

- Предметно: компьютер это очень глупая коробка!
- Он понимает только нули и единицы
- …но делает это очень хорошо!
- Насколько хорошо?
- Настолько хорошо, что если из нулей и единиц составить числа, то он сможет за секунду обработать больше чисел, чем вы за всю свою жизнь!

Линейная алгебра. Мотивация

- Какой отсюда следует вывод?
- Фото, видео, звуки вашего голоса, стихи Маяковского — всё это нужно выразить на языке чисел.

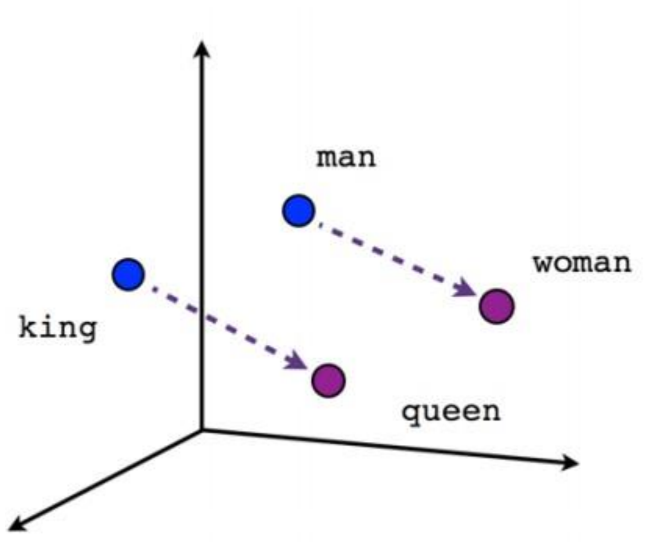
Линейная алгебра. Мотивация

- Какой отсюда следует вывод?
- Фото, видео, звуки вашего голоса, стихи Маяковского — всё это нужно выразить на языке чисел.
- И возможным это делает раздел математики под названием “линейная алгебра”.

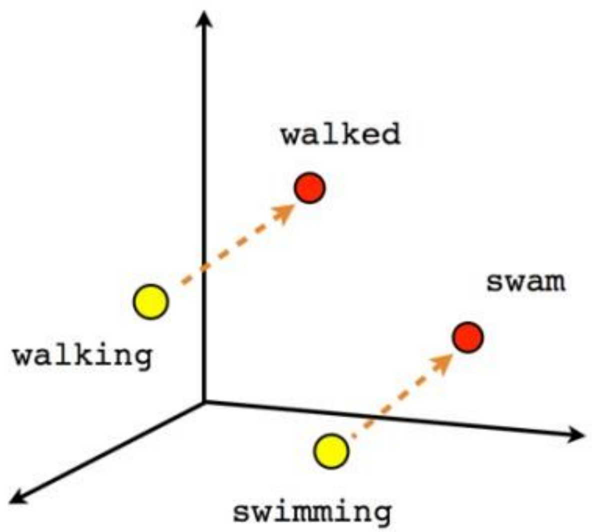
Линейная алгебра. Мотивация

- Линейная алгебра открыла для нас невообразимые вещи:

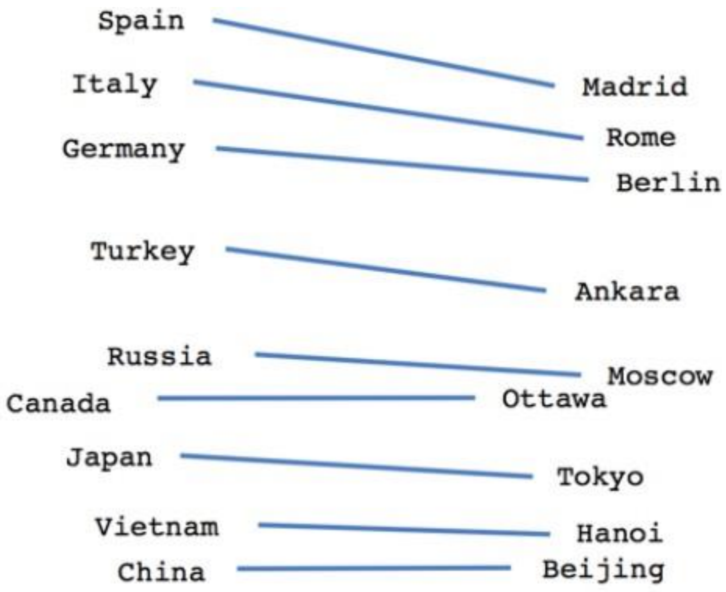
Арифметика слов



Male-Female

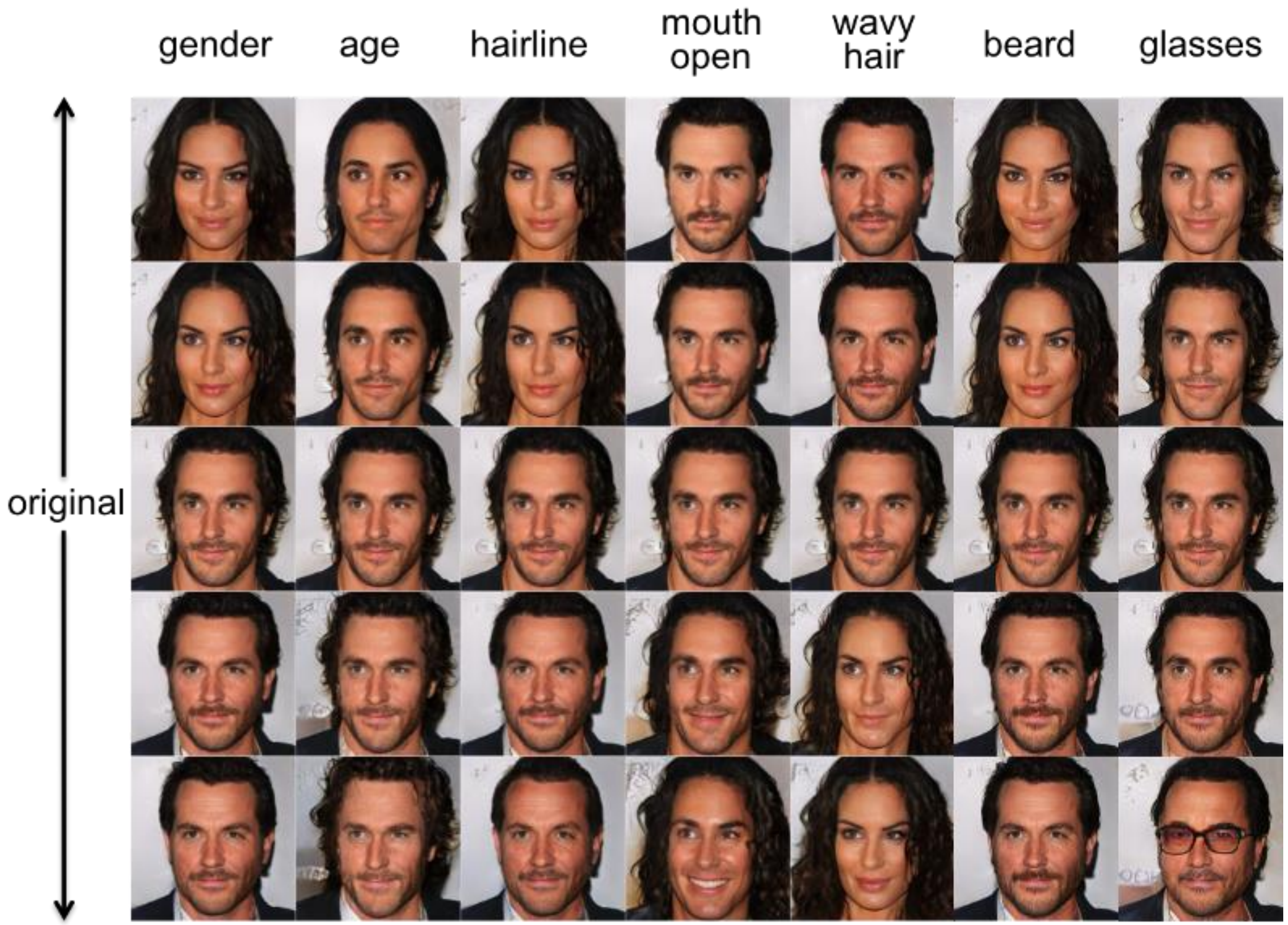


Verb tense



Country-Capital

Генерация лиц с заданными свойствами



Линейная алгебра. Мотивация

- …и многое другое

Линейная алгебра. Мотивация

- …и многое другое
- Линейная алгебра это язык, на котором задачу понимает машина.
- Наша задача — стать переводчиками с одного языка на другой.

Основные понятия

Линейное пространство

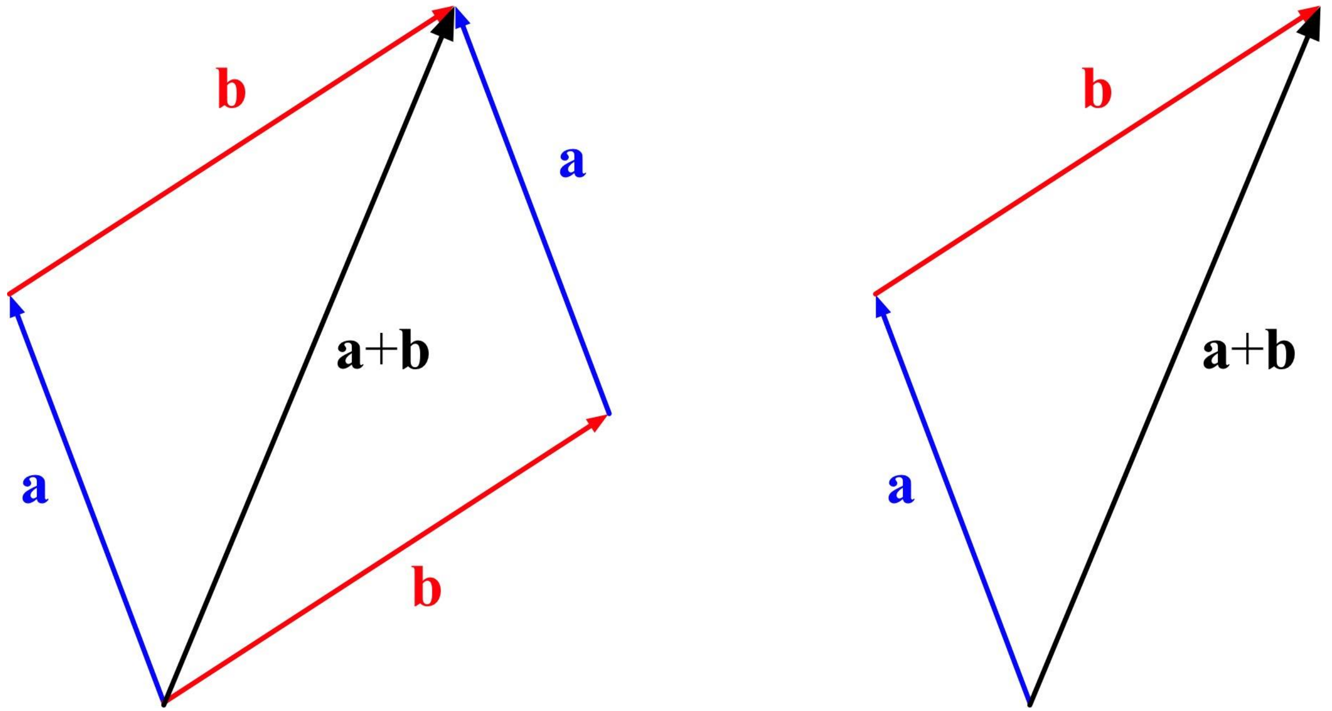
- Основополагающим понятием линейной алгебры является понятие линейного пространства.
- Давайте вспомним, что это!

Линейное пространство

- Основополагающим понятием линейной алгебры является понятие линейного пространства.
- Давайте вспомним, что это!
- Математическая структура, представляющая собой набор элементов, называемых векторами, для которых определены операции сложения друг с другом и умножения на число — скаляр

Линейное пространство

- Пример: множество вещественных векторов из n компонент (например (0, 0, 0, …, 0))



Линейное пространство

- На практике вам придётся иметь дело преимущественно с вещественными векторами, матрицами и тензорами.
- Тем не менее, линейные пространства на этом не исчерпываются.

Линейное пространство

- На практике вам придётся иметь дело преимущественно с вещественными векторами, матрицами и тензорами.
- Тем не менее, линейные пространства на этом не исчерпываются.

В общем случае, линейное пространство состоит из абелевой группы векторов (с операцией сложения), поля скаляров и операции умножения вектора на скаляр, согласованной со сложением векторов при помощи аксиом дистрибутивности.

Эти условия легко проверить, таким образом определив, является ли пространство линейным (векторным).

Линейное пространство

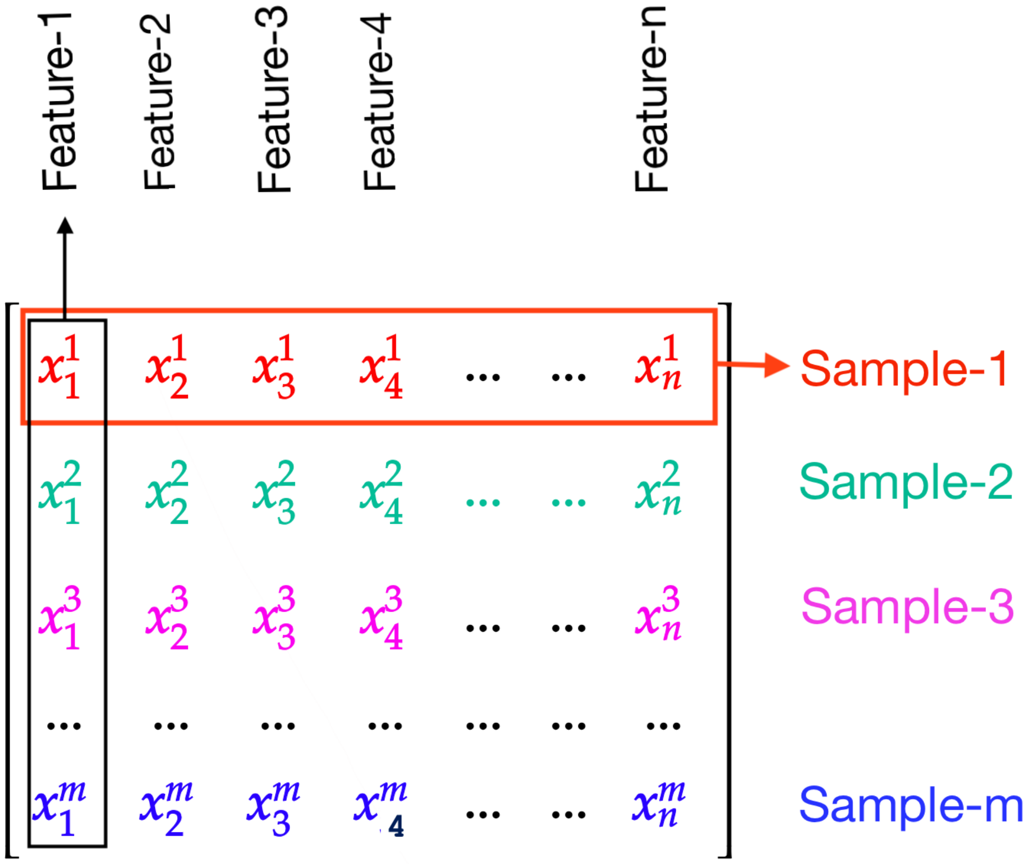
- Приведите примеры других линейных пространств!

Линейное пространство

- Приведите примеры других линейных пространств!
  - Пространство матриц одинаковой размерности.
  - Пространство тензоров (“многомерных” матриц) одинаковой размерности.
  - Пространство многочленов (обычных или тригонометрических).
  - Пространство решений однородной системы линейных уравнений, алгебраических или дифференциальных.
  - Пространство двоичных векторов с операцией сложения по модулю 2 и операцией умножения на 0 и 1.

Линейное пространство

- Отлично, мы вспомнили линейные пространства!
- Тем не менее, на этом занятии мы сосредоточимся на вещественных векторах, матрицах и связи между ними, поскольку именно с ними в большинстве своем работает Data Scientist



Матрицы

- А напомните, пожалуйста, что такое вообще матрица?

Матрицы

- А напомните, пожалуйста, что такое вообще матрица?
- Матрица – представление чисел в виде двумерной таблицы

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & -1 \\ 3 & 4 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

Матрицы

- Транспонированная матрица?

Матрицы

- Транспонированная матрица?
- Транспонированная матрица – зеркально отображенная относительно главной диагонали матрица.

$$\begin{pmatrix} 1 & 2 & -1 \\ 3 & 4 & 0 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ -1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

Матрицы

- Единичная матрица?

Матрицы

- Единичная матрица?
- Единичная матрица – квадратная матрица, у которой на главной диагонали стоят единицы, а на всех остальных местах нули. Обычно обозначают I или E.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Матрицы

- Диагональная матрица?

Матрицы

- Диагональная матрица?
- Диагональная матрица – матрица, у которой на главной диагонали стоят любые числа, а на всех остальных местах нули.

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 10 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 70 \end{pmatrix}$$

Матрицы

- Принцип умножения матрицы на вектор?

Матрицы

- Принцип умножения матрицы на вектор?

$$\begin{pmatrix} 1 & 5 \\ 3 & 8 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix} \qquad \begin{pmatrix} 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 5$$

Матрицы

- Принцип умножения матрицы на вектор?

$$\begin{pmatrix}1 & 5\\ 3 & 8\end{pmatrix} \cdot \begin{pmatrix}-1\\ 1\end{pmatrix} = \begin{pmatrix}4\\ 5\end{pmatrix} \qquad \begin{pmatrix}2 & 3\end{pmatrix} \cdot \begin{pmatrix}1\\ 1\end{pmatrix} = 5$$

$$\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} \cdot \begin{pmatrix}1\\ 1\\ 1\end{pmatrix} =$$

Матрицы

- Принцип умножения матрицы на матрицу?

Матрицы

- Принцип умножения матрицы на матрицу?

$$\begin{pmatrix} -1 & 3 \\ 4 & 0 \end{pmatrix} \cdot \begin{pmatrix} -1 & 4 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 10 & -4 \\ -4 & 16 \end{pmatrix}$$

Матрицы

- Принцип умножения матрицы на матрицу?

$$\begin{pmatrix} -1 & 3 \\ 4 & 0 \end{pmatrix} \cdot \begin{pmatrix} -1 & 4 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 10 & -4 \\ -4 & 16 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 3 & 1 \\ 4 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} =$$

Матрицы

- Обратная матрица?

Матрицы

- Обратная матрица?
- Обратная матрица – такая матрица, при умножение на которую исходная матрица станет единичной. Обозначают верхним индексом -1

$$A \cdot A^{-1} = I$$

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}$$

Матрицы

- Обратная матрица?
- Обратная матрица – такая матрица, при умножение на которую исходная матрица станет единичной. Обозначают верхним индексом -1

$$A \cdot A^{-1} = I$$

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}$$

В общем случае, находить обратную матрицу вычислительно очень сложно, и это на практике часто делают приближенно.

Базис линейного пространства

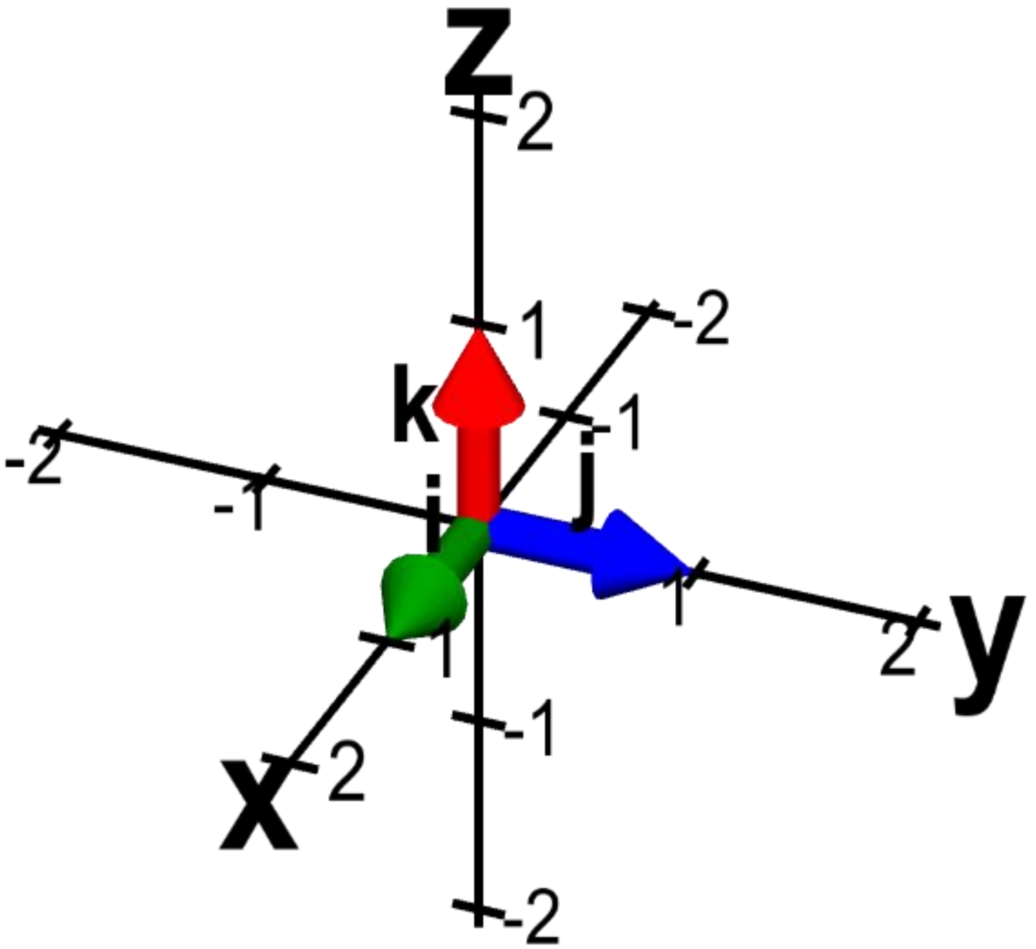
- Что такое базис линейного пространства?

Базис линейного пространства

- Что такое базис линейного пространства?
- Во всяком линейном пространстве есть несколько главных векторов — так называемый базис. Остальные вектора выражаются через базисные.
- Количество базисных векторов в базисе определяется размерностью пространства.

Базис линейного пространства

- Пример:
- В случае $\mathbb{R}^n$ базисных векторов ровно n.
- Они задают координатные оси.



Базис линейного пространства

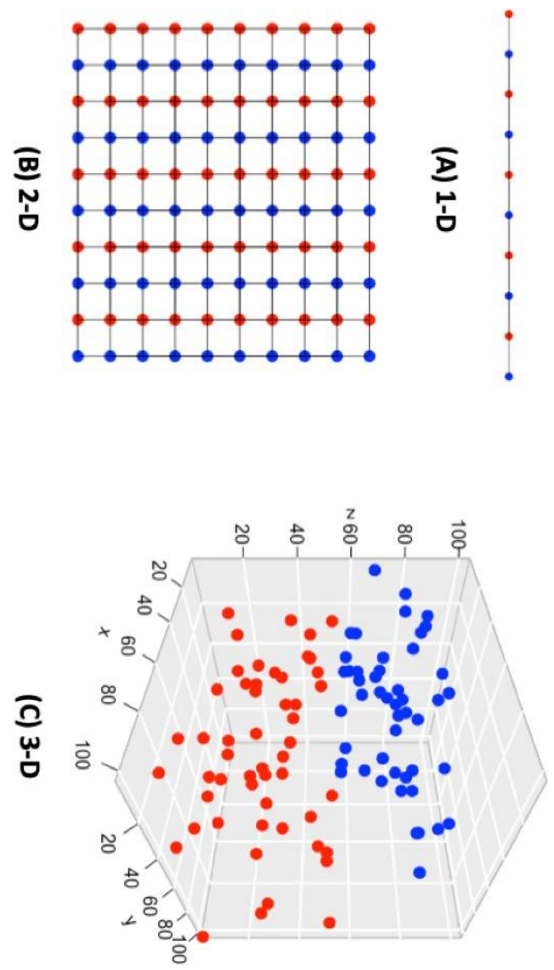
- Базис — такое минимальное по включению множество векторов, что:
  - Ни один из них не выражается линейной комбинацией остальных;
  - Каждый элемент пространства можно единственным образом представить линейной комбинацией конечного набора векторов из этого множества.

Базис линейного пространства

- Базис — такое минимальное по включению множество векторов, что:
  - Ни один из них не выражается линейной комбинацией остальных;
  - Каждый элемент пространства можно единственным образом представить линейной комбинацией конечного набора векторов из этого множества.
- Доказать, что всякое линейное пространство имеет базис, можно при помощи леммы Цорна.

Размерность линейного пространства

- Размерность — одна из главных характеристик линейного пространства.
- Размерность в точности совпадает с числом базисных векторов.
- В машинном обучении есть даже одноимённое проклятие :)



Матрицы

- Определитель?

Матрицы

- Определитель?
- А вот тут уже интереснее – да, это такая, некоторая мера для матриц. Но что же она показывает?

Матрицы

- Определитель?
- А вот тут уже интереснее – да, это такая, некоторая мера для матриц. Но что же она показывает?
- Определитель (детерминант) – численная характеристика матрицы, которая в некотором смысле описывает сжатие или растяжение пространства при преобразовании этой матрицей.

$$det \begin{pmatrix} -1 & -1 \\ 1 & 3 \end{pmatrix} = (-1) \cdot 3 - (-1) \cdot 1 = -2$$

Матрицы

- Если связать получившееся значение с геометрическим смыслом, то если рассмотреть то, что мы матрицей преобразуем единичный квадрат, то:
  - Знак минус будет говорить, что наш квадрат будет иметь противоположную ориентацию
  - Значение 2 будет говорить о том, что после преобразования наш квадрат будет иметь площадь, в два раза превосходящую площадь оригинального квадрата.

$$det \begin{pmatrix} -1 & -1 \\ 1 & 3 \end{pmatrix} = (-1) \cdot 3 - (-1) \cdot 1 = -2$$

Линейные преобразования

- А вот теперь серьезный вопрос: как матрицы связаны с линейными преобразованиями?

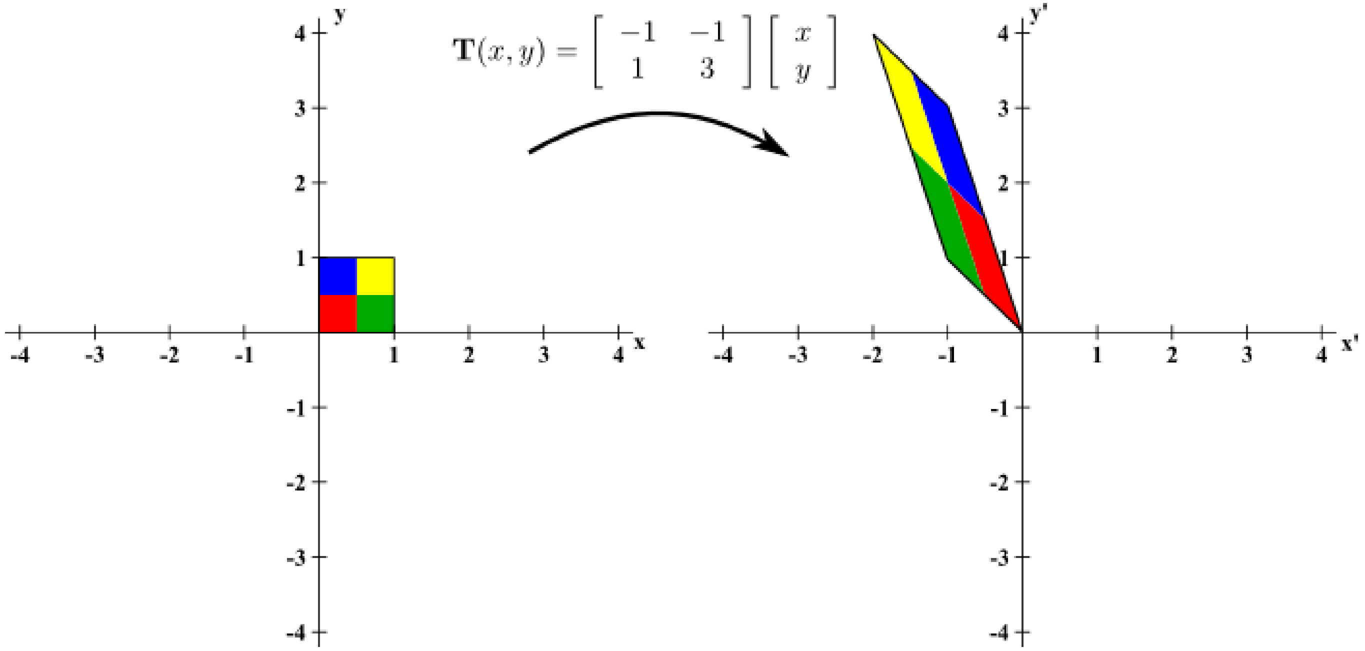
Линейные преобразования

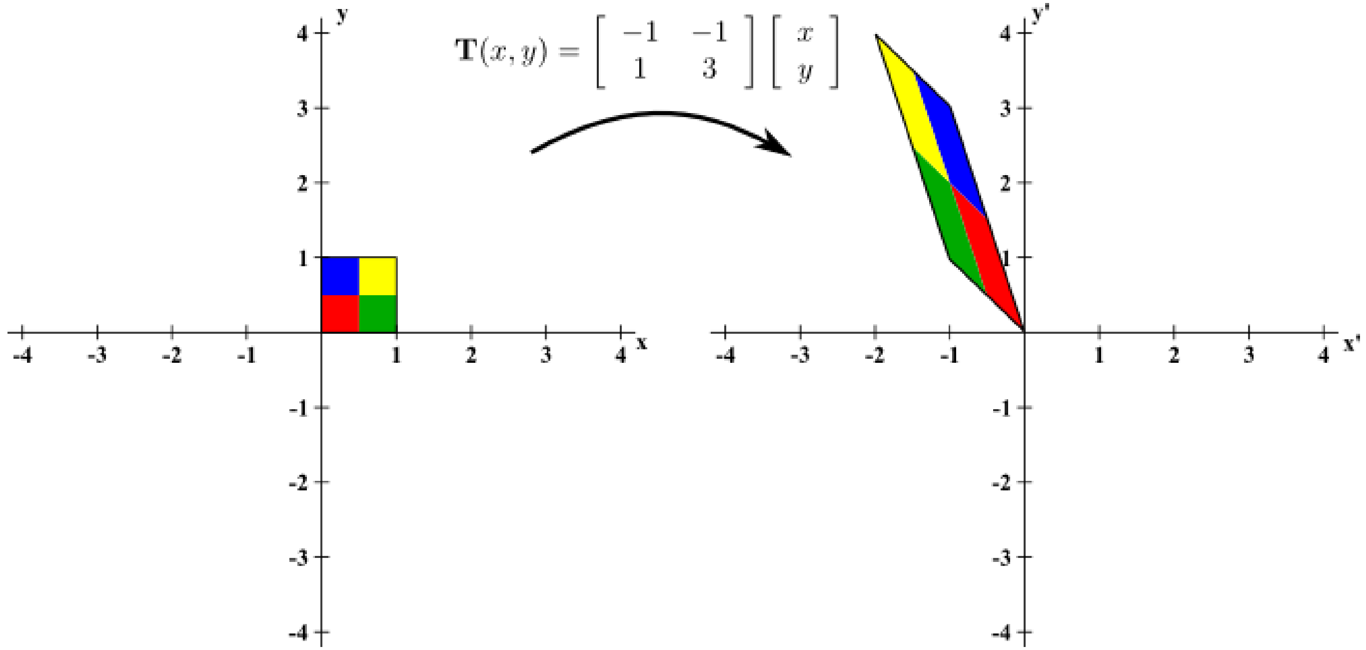
- А вот теперь серьезный вопрос: как матрицы связаны с линейными преобразованиями?
- Каждая матрица кодирует линейное преобразование линейных пространств в некоторой системе координат: одно пространство отображается на подпространство другого.
- Все эти преобразования можно разложить на комбинацию проекций, вращений, отражений и масштабирования вдоль координатных осей. Других преобразований нет.

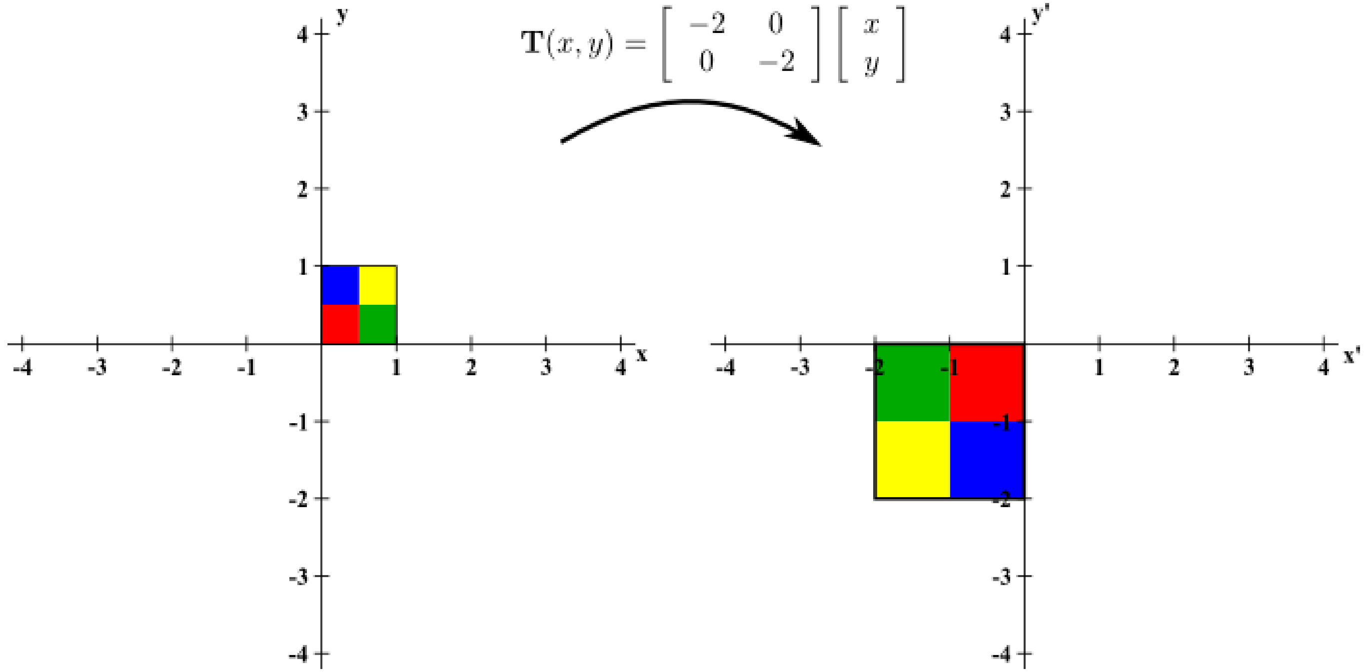
Линейные преобразования

- Одному преобразованию можно сопоставить бесконечное количество матриц! Столько же, сколько есть систем координат. Сама по себе матрица это просто таблица чисел!

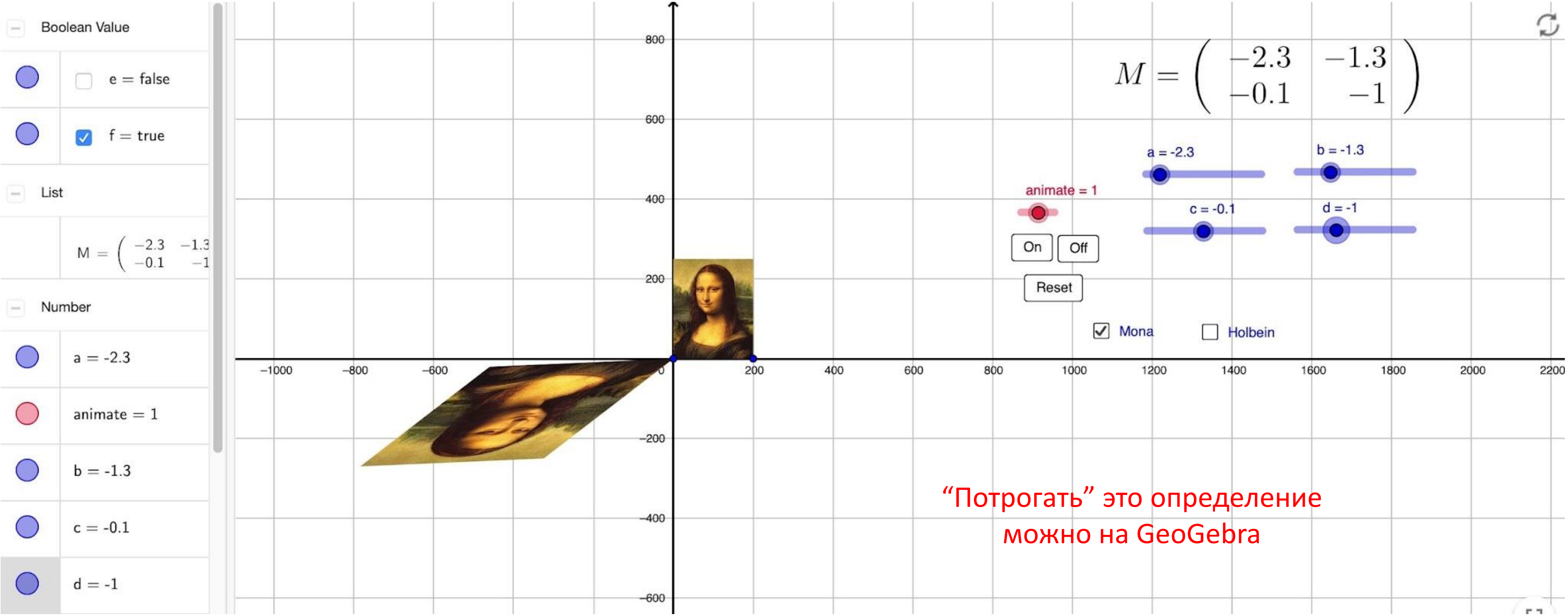
$$\begin{pmatrix}-1 & -1\\ 1 & 3\end{pmatrix}\cdot\begin{pmatrix}\quad\end{pmatrix} =$$




$\mathbf{T}(x,y) = \begin{bmatrix} -1 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

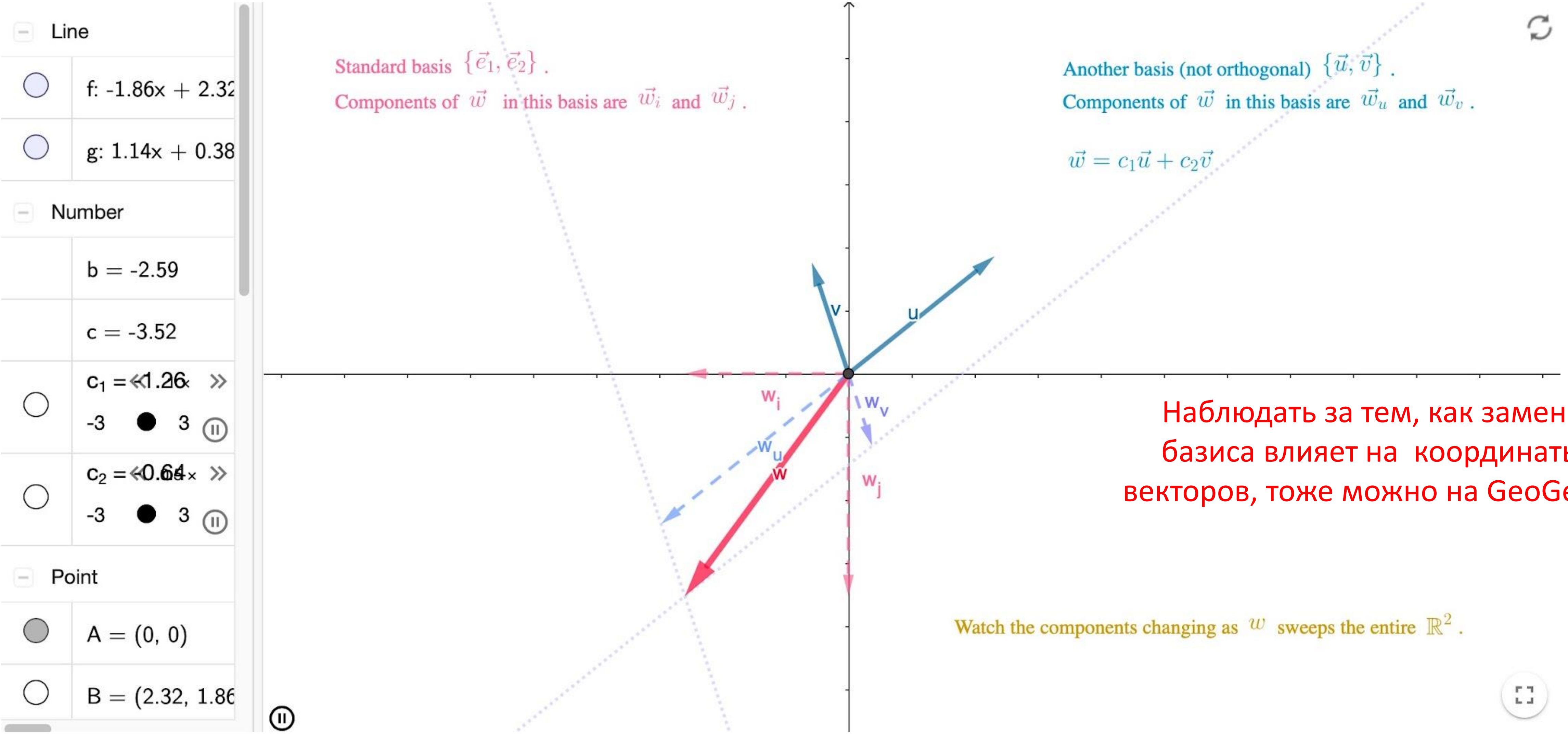

$\mathbf{T}(x,y) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

Линейные преобразования



“Потрогать” это определение можно на GeoGebra

Линейные преобразования



Наблюдать за тем, как замена базиса влияет на координаты векторов, тоже можно на GeoGebra

Собственные направления

- А как вот эти все чудеса связаны с собственными направлениями и собственными значениями?
- Помните, да, вроде было что-то такое? :)

Собственные направления

- У линейного преобразования $A : V \to V$ есть некоторое количество собственных направлений.
- Вдоль них оно не вращает, не отражает, а масштабирует!
- Это пример т.н. инвариантных подпространств: они не изменяются под действием A .

- Формулами это задается так:
- $Av = \lambda v$
  - λ – собственное значение
  - v – собственный вектор

- Например:

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$$

$$v = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Собственные направления

- Понятие собственных направлений очень полезно, поскольку позволяет находить “хорошие” направления нашего преобразования и, кроме того, если собственных векторов окажется столько, сколько базисных, то мы сможем “перейти” в базис собственных векторов, в котором наша матрица A будет иметь диагональный вид, что упрощает многие выкладки.

Собственные направления

- Понятие собственных направлений очень полезно, поскольку позволяет находить “хорошие” направления нашего преобразования и, кроме того, если собственных векторов окажется столько, сколько базисных, то мы сможем “перейти” в базис собственных векторов, в котором наша матрица А будет иметь диагональный вид, что упрощает многие выкладки.

У собственных направлений, например, может быть следующий физический смысл: в теории колебаний собственные числа – это частоты колебаний системы при свободном движении, а собственные вектора – это их «траектории».

Матричные разложения

Матричные разложения

- Теперь, когда мы с вами вспомнили и еще раз обсудили основные понятия линейной алгебры (которые понадобятся нам уже сегодня), мы можем, наконец, переходить к основному предмету нашего рассмотрения – матричным разложениям!

Матричные разложения

- Теперь, когда мы с вами вспомнили и еще раз обсудили основные понятия линейной алгебры (которые понадобятся нам уже сегодня), мы можем, наконец, переходить к основному предмету нашего рассмотрения – матричным разложениям!
- Давайте еще раз взглянем на общий вид матрицы признаков

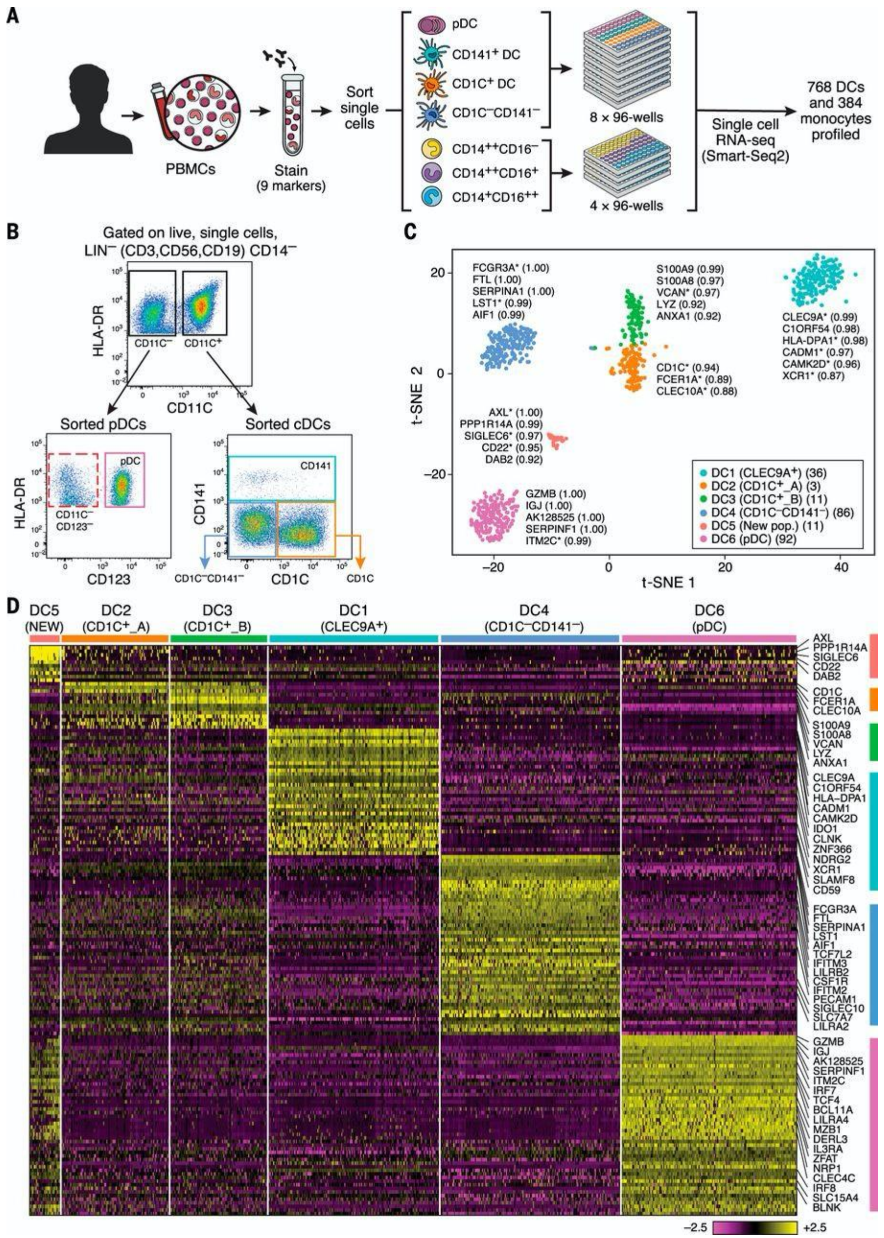
Матрица признаков

- Матрицы признаков – прямоугольные вещественные матрицы размера $m \times n$, где:
  - m – размер выборки
  - n – количество признаков (и размер признакового пространства)

$$
\begin{array}{c|c}
\begin{array}{ccccccc} \text{Feature-1} & \text{Feature-2} & \text{Feature-3} & \text{Feature-4} & & & \text{Feature-n} \end{array} & \\
\begin{bmatrix}
x_1^1 & x_2^1 & x_3^1 & x_4^1 & \dots & \dots & x_n^1 \\
x_1^2 & x_2^2 & x_3^2 & x_4^2 & \dots & \dots & x_n^2 \\
x_1^3 & x_2^3 & x_3^3 & x_4^3 & \dots & \dots & x_n^3 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots \\
x_1^m & x_2^m & x_3^m & x_4^m & \dots & \dots & x_n^m
\end{bmatrix} &
\begin{array}{l} \text{Sample-1} \\ \text{Sample-2} \\ \text{Sample-3} \\ \\ \text{Sample-m} \end{array}
\end{array}
$$

Матрица признаков

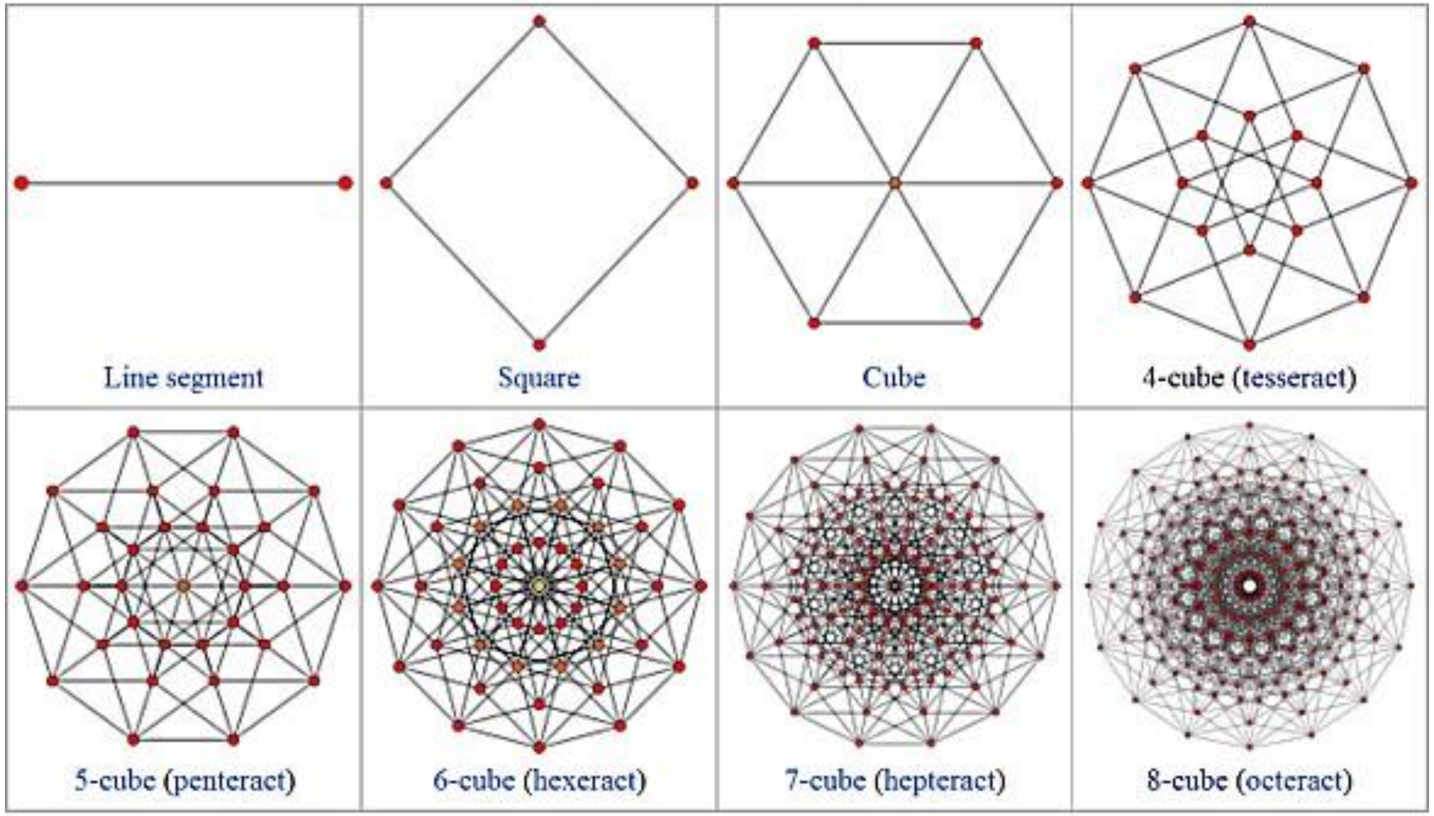
- Линейная алгебра — инструмент, который позволяет вам работать с данными огромной размерности.
- Пример:
- Типичная клетка опухоли в данных секвенирования — это 14000 чисел, а образец обычно содержит от ста до миллиона клеток.



Матрица признаков

- Тем не менее, наш мозг не в состоянии воспринимать данные высокой размерности.
- Всё, что выше размерности 7 (длина, ширина, высота, три цвета, время) — практически за гранью нашего восприятия.
- Даже простейшие геометрические объекты в духе куба становится практически невозможно изобразить без потери информации.

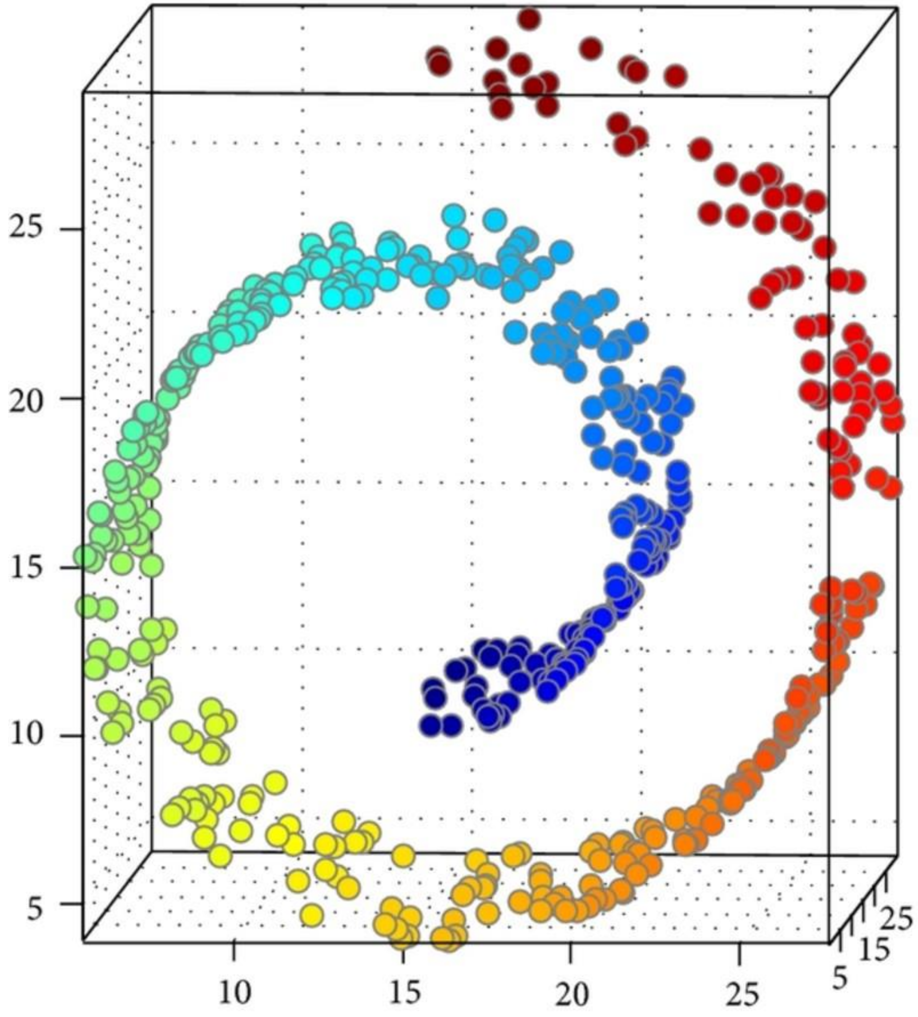
Матрица признаков



Укладка многомерных кубов (гиперкубов) на плоскости

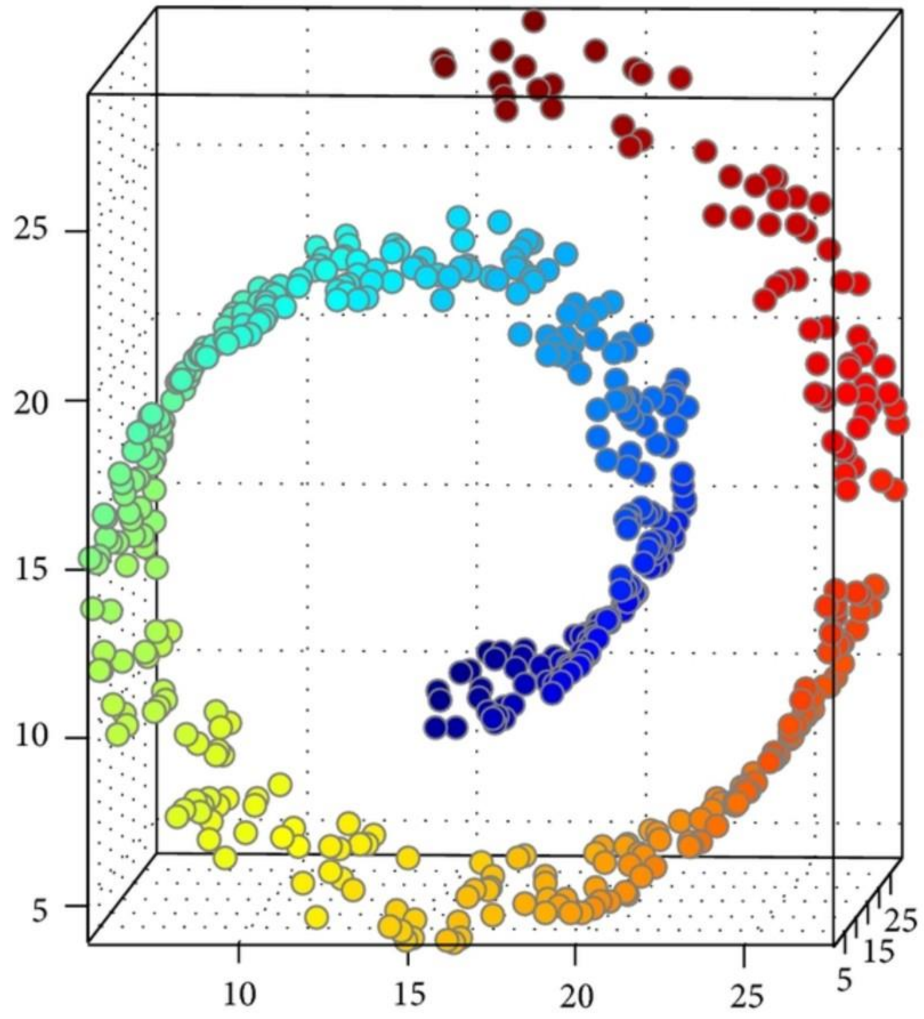
Матрица признаков

- К счастью, реальные данные обычно лежат на т.н. маломерных многообразиях.
- Они не распределены по многомерному пространству равномерно, а лежат на поверхностях существенно меньшей размерности.



Матрица признаков

- К счастью, реальные данные обычно лежат на т.н. маломерных многообразиях.
- Они не распределены по многомерному пространству равномерно, а лежат на поверхностях существенно меньшей размерности.
- Обратите внимание на пример. Формально это трёхмерные данные, но по факту они почти полностью лежат на двумерной полосе!

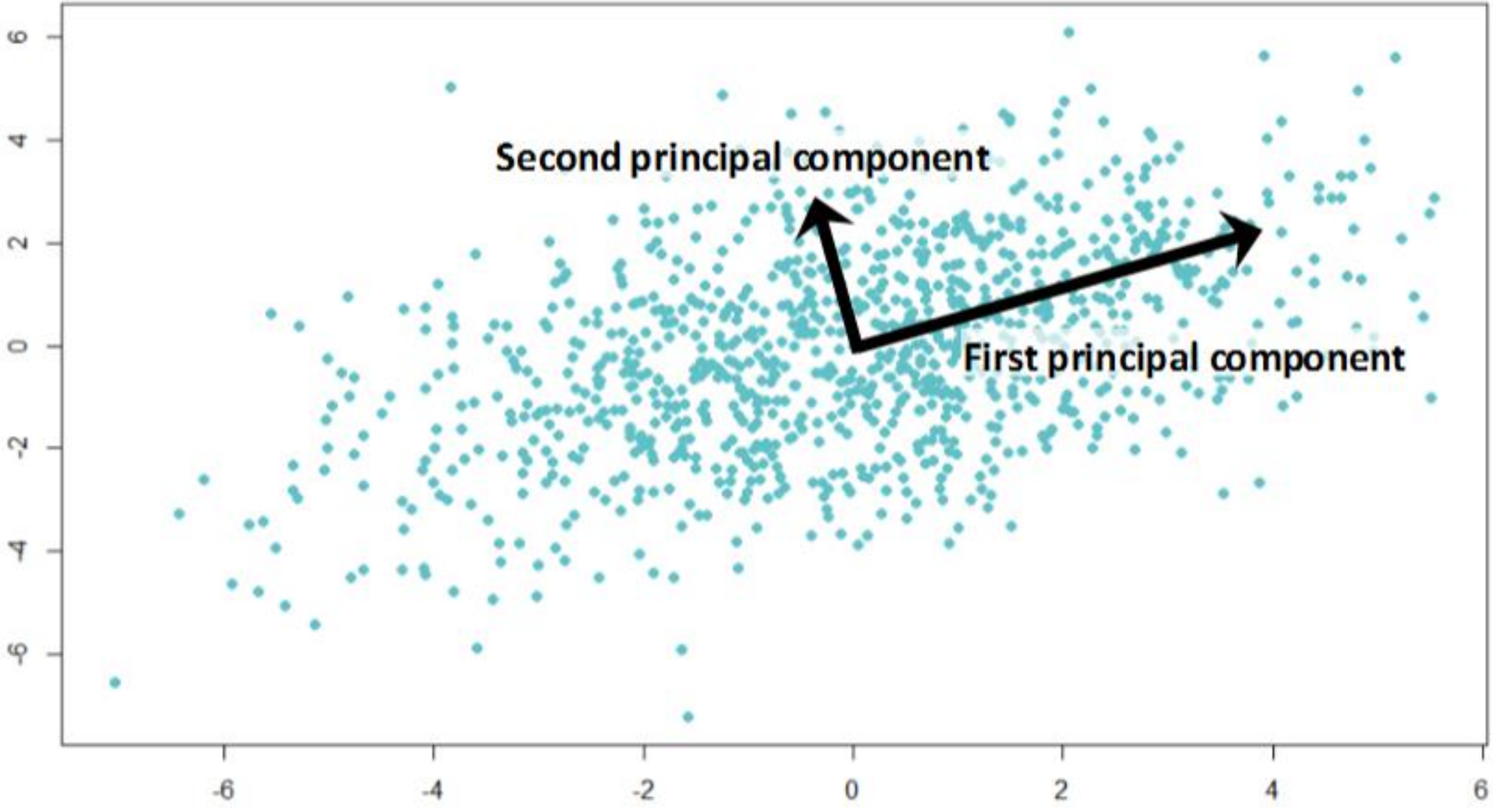


Матрица признаков

- Определением тонкой геометрической структуры данных занимается раздел ML под названием manifold learning.
- Но часто понизить размерность, скажем, с 14000 до 30-100 без существенной потери информации можно и при помощи более простых методов.
- Например, при помощи PCA — метода главных компонент, который ищет оптимальное приближение размерности для вашей матрицы данных.

Метод главных компонент (PCA)

- Алгоритм работы PCA можно описать следующими шагами:
  1. Найти такое направление в пространстве, вдоль которого дисперсия данных максимальна.
  2. Среди оставшихся направлений, ортогональных предыдущим, найти направление, вдоль которого дисперсия максимальна.
  3. Повторять шаг 2 до тех пор, пока это возможно.
- Получившиеся направления и называют главными компонентами.


Second principal component
First principal component
6
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6

Метод главных компонент (PCA)

- Но как искать такие направления?!

Метод главных компонент (PCA)

- Но как искать такие направления?!
- Мы могли бы находить такие направления полным перебором, но сколько бы это заняло времени хотя бы в 10-мерном пространстве?...

Метод главных компонент (PCA)

- Но как искать такие направления?!
- Мы могли бы находить такие направления полным перебором, но сколько бы это заняло времени хотя бы в 10-мерном пространстве?...
- Поэтому мы, как настоящие математики, воспользуемся доступным мат. аппаратом, чтобы решить эту задачу!

Важные понятия линейной алгебры

- Ортогональная матрица – такая матрица, для которой обратная матрица совпадает с транспонированной.

$$A \cdot A^T = A \cdot A^{-1} = I$$

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Важные понятия линейной алгебры

- Сингулярные направления и значения:

$$Mv = \sigma u$$
$$M^T u = \sigma v$$

- где
  - v – правый сингулярный вектор,
  - u – левый сингулярный вектор,
  - σ – сингулярное значение.
- Сингулярное значение это корень из собственного значения матрицы $M^T M$

Важные понятия линейной алгебры

- Сингулярные направления и значения:
- где
  - v – правый сингулярный вектор,
  - u – левый сингулярный вектор,
  - σ – сингулярное значение.
- Сингулярное значение это корень из собственного значения матрицы $M^T M$

$$Mv = \sigma u$$
$$M^T u = \sigma v$$

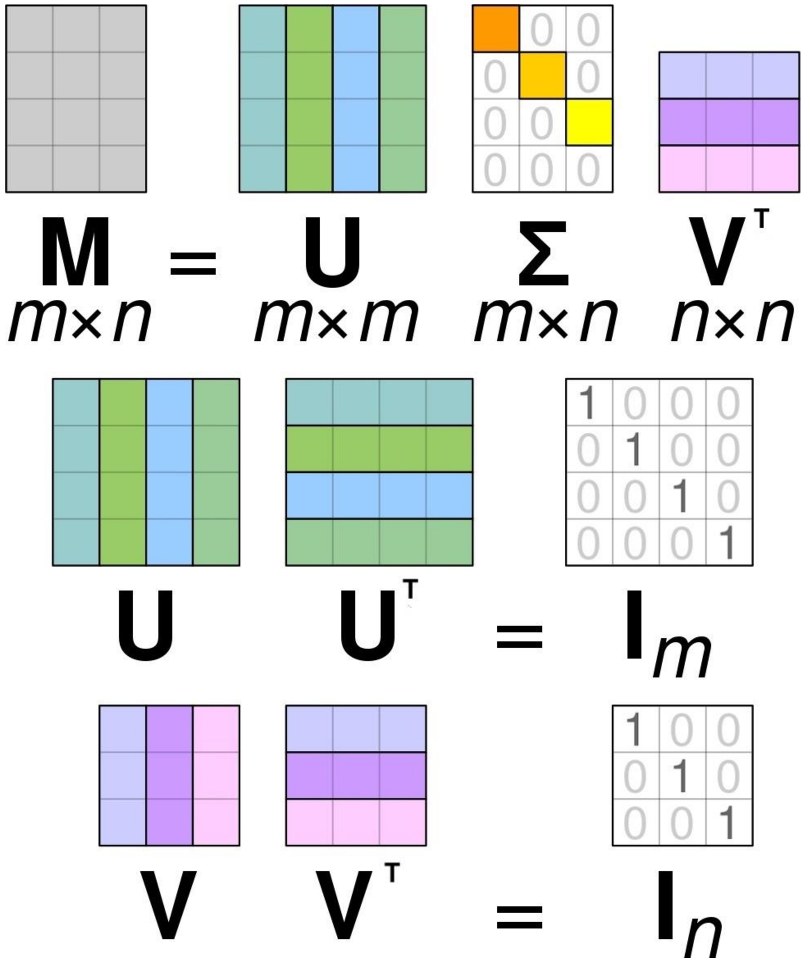
Сингулярные вектора и значения это в некотором смысле аналоги собственных векторов и чисел для неквадратных матриц. Причем, если матрица квадратная, то квадрат сингулярного значения равен модулю соответствующего собственного значения (если такое существует).

Сингулярное матричное разложение (SVD)

- Так вот оказывается, что любую матрицу M можно разложить в произведение трех таких матриц U , Σ и V^T, что:

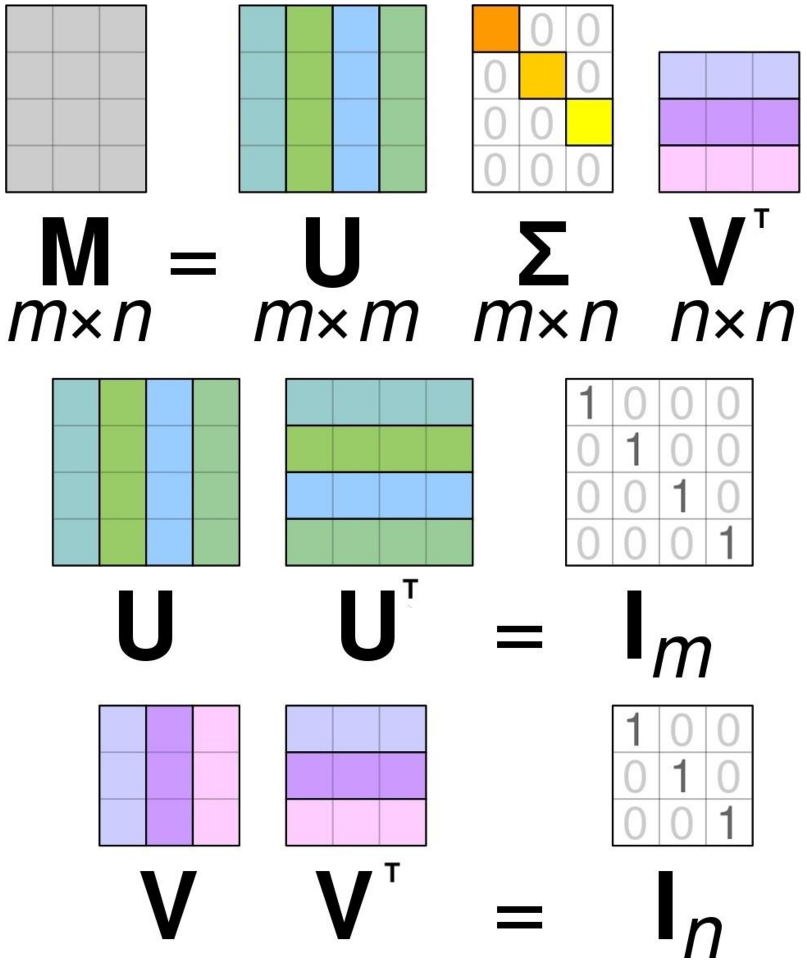
Сингулярное матричное разложение (SVD)

- Так вот оказывается, что любую матрицу M можно разложить в произведение трех таких матриц U, Σ и V^T, что:

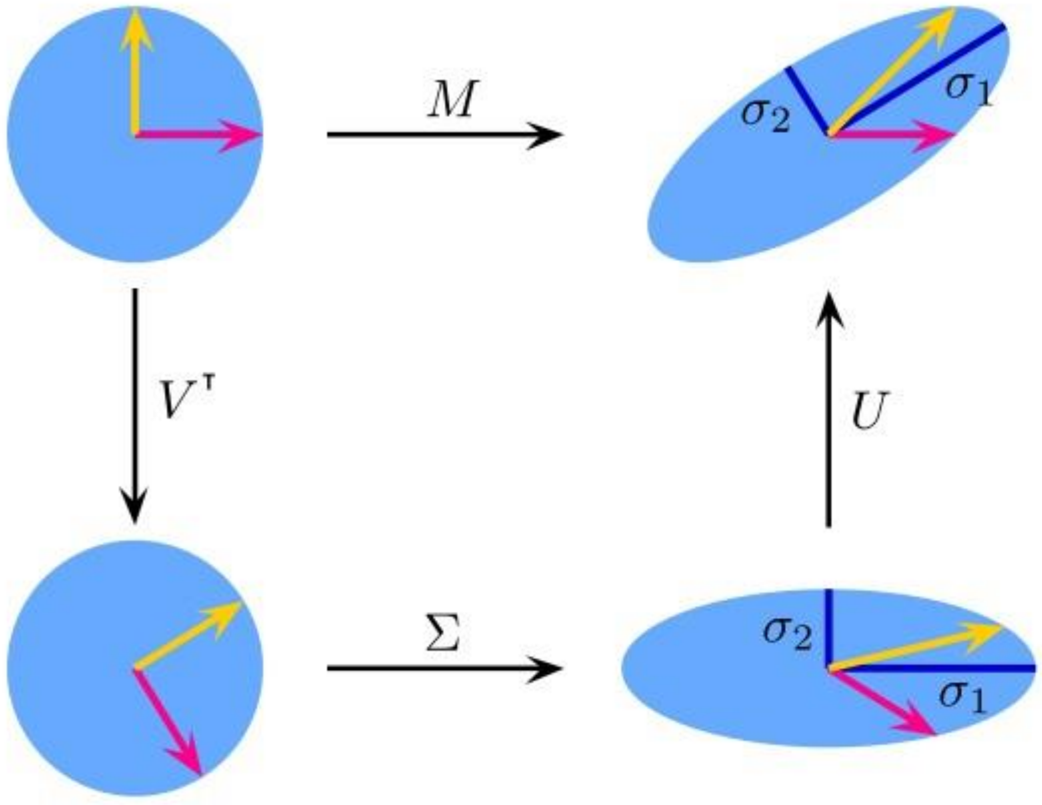


Сингулярное матричное разложение (SVD)

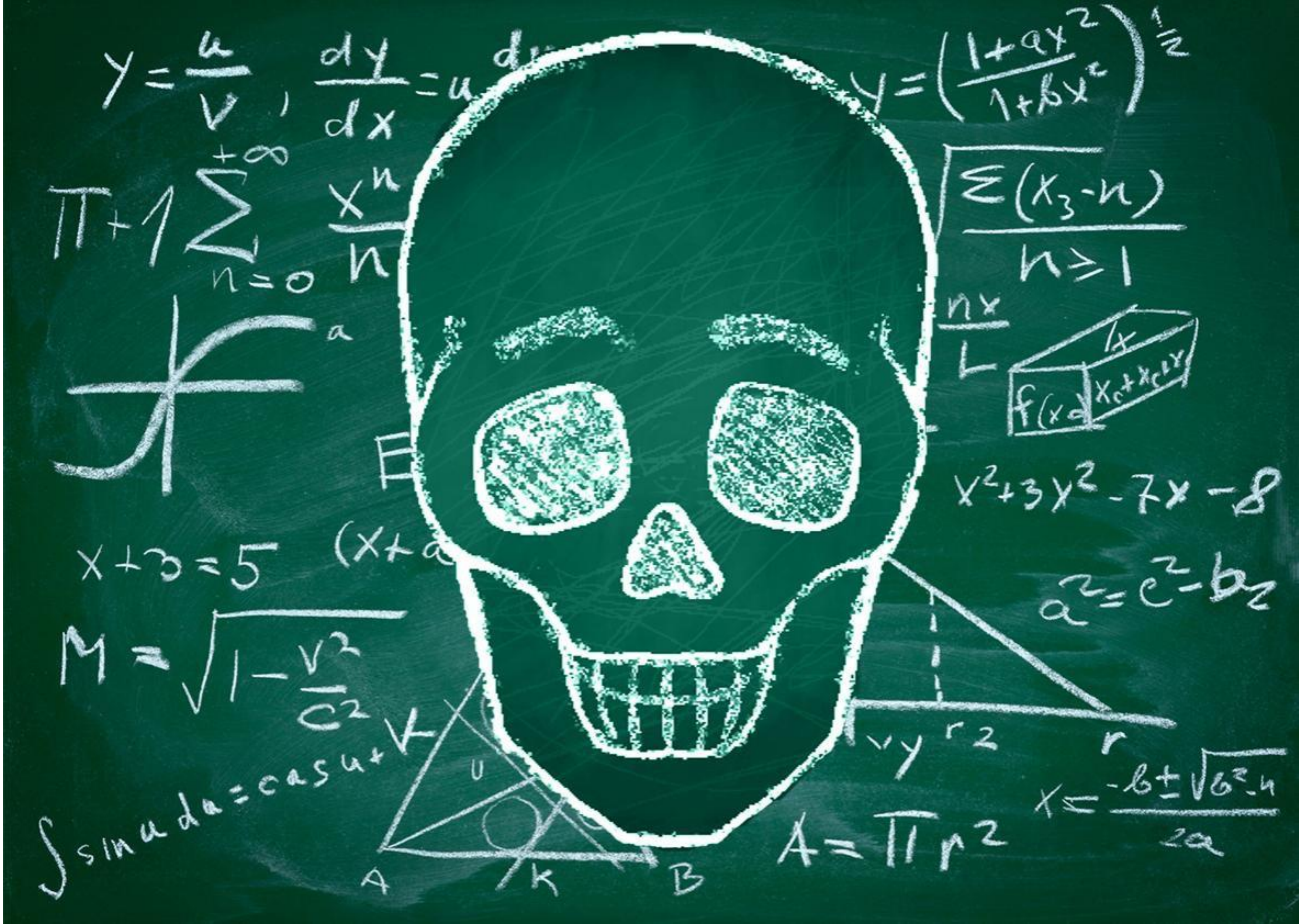
- Так вот оказывается, что любую матрицу M можно разложить в произведение трех таких матриц U, Σ и V^T, что:
  - $M \in \mathbb{R}^{m \times n}$ — Исходная матрица;
  - $U \in \mathbb{R}^{m \times m}$— Матрица левых сингулярных векторов;
  - $\Sigma \in \mathbb{R}^{m \times n}$— Матрица, на главной диагонали которой находятся т.н. сингулярные числа (в порядке невозрастания);
  - $V \in \mathbb{R}^{n \times n}$— Матрица правых сингулярных векторов.



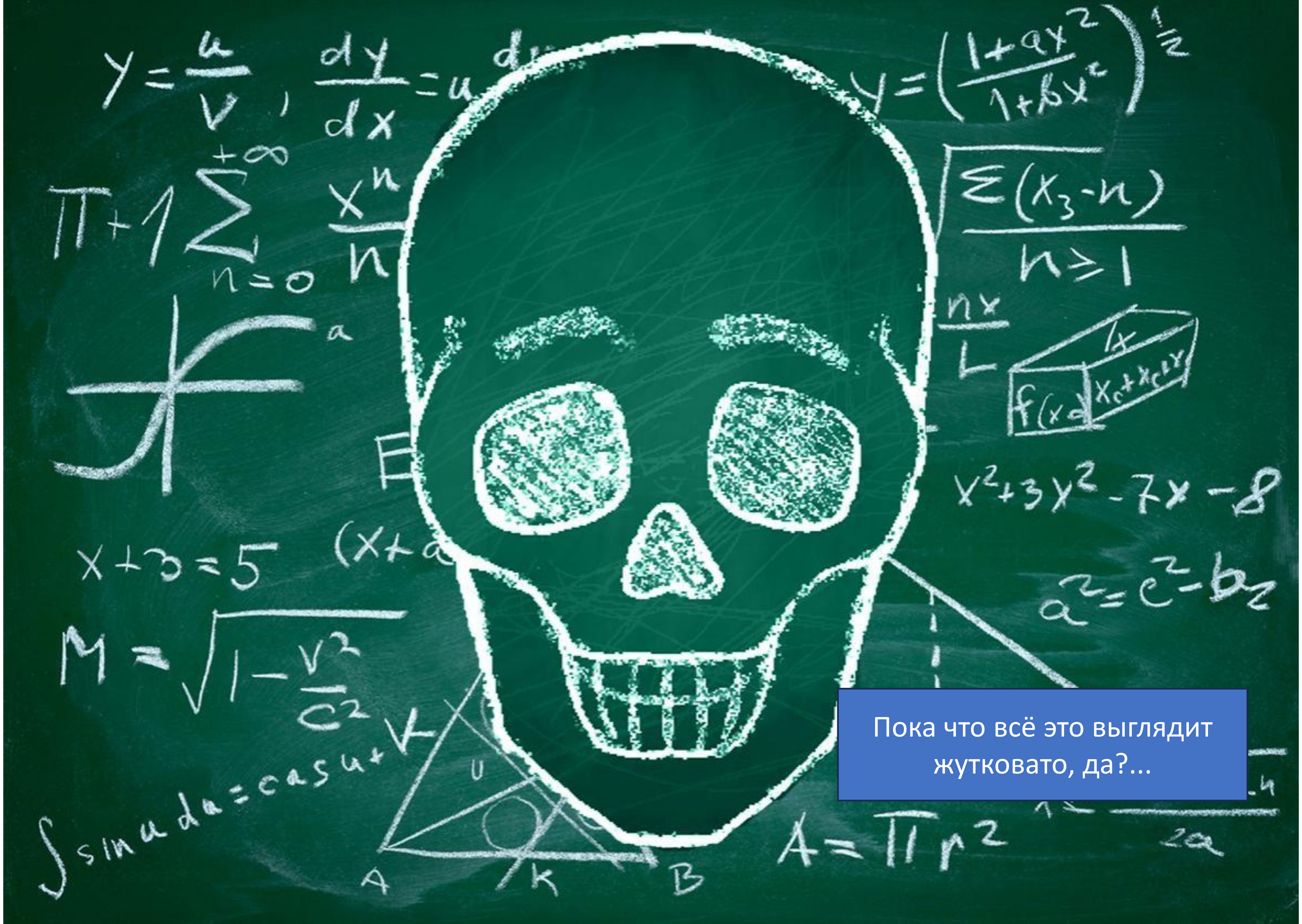
Сингулярное матричное разложение (SVD)



$$M = U \cdot \Sigma \cdot V^{\intercal}$$



x+3=5
M = √(1−v²/c²)
∫sin u du = cos u + C
x²+3x²−7x−8
a²=c²−b²
A=πr²
x = (−b ± √(b²−4ac))/2a
n=0
n≥1
A
B
K

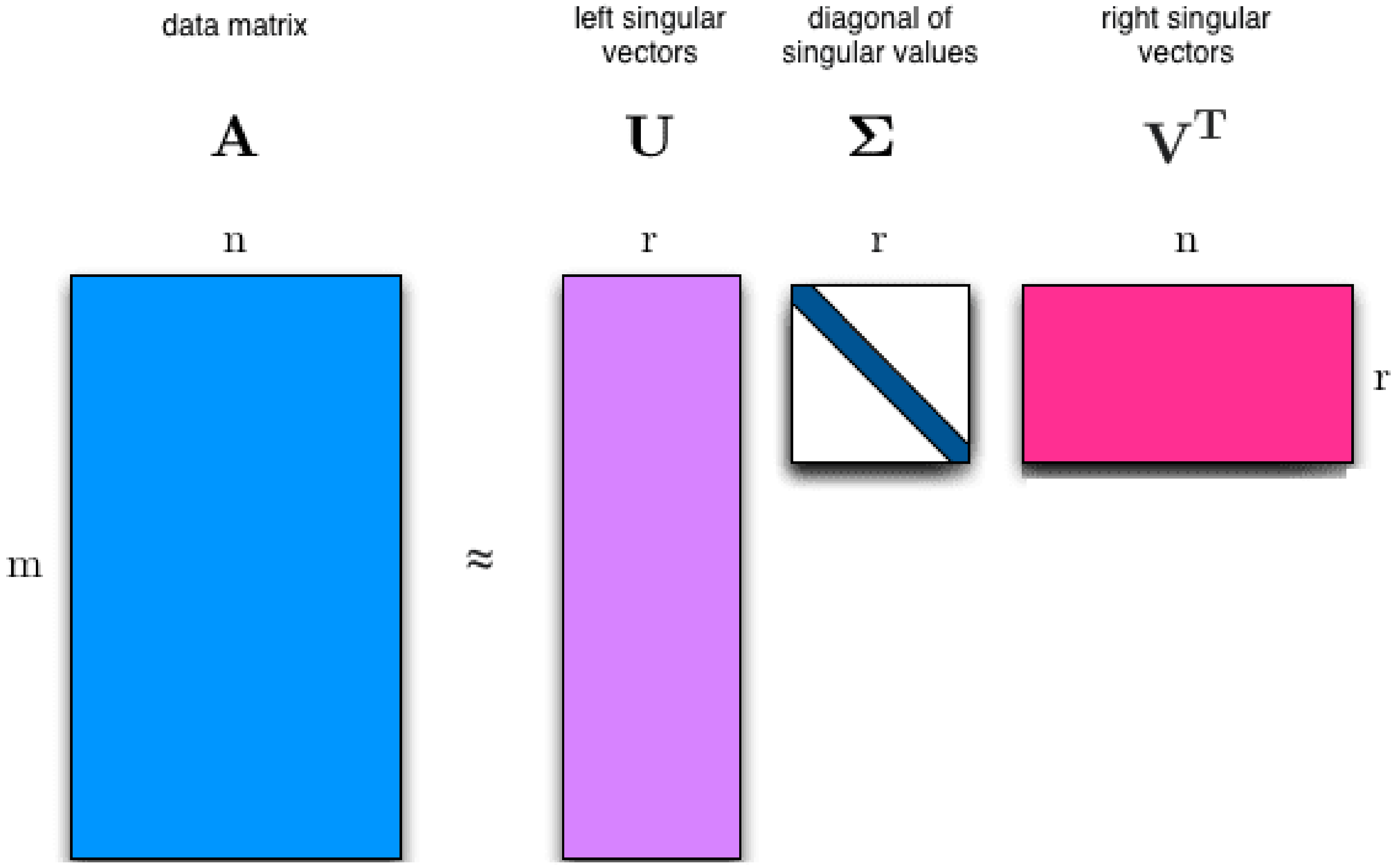


Пока что всё это выглядит жутковато, да?...

Усечённое SVD-разложение ранга r

- Усечённое SVD-разложение ранга r матрицы $M \in \mathbb{R}^{m \times n}$ — это обычное SVD-разложение, где оставили только самые большие сингулярные значения вместе с соответствующими сингулярными векторами.

Усечённое SVD-разложение ранга r



Усечённое SVD-разложение ранга r

- Усечённое SVD-разложение ранга r матрицы $M \in \mathbb{R}^{m \times n}$ — это обычное SVD-разложение, где оставили только самые большие сингулярные значения вместе с соответствующими сингулярными векторами.

Усечённое SVD-разложение ранга r

- Усечённое SVD-разложение ранга r матрицы $M \in \mathbb{R}^{m \times n}$ — это обычное SVD-разложение, где оставили только самые большие сингулярные значения вместе с соответствующими сингулярными векторами.
- Можно доказать, что усечённое SVD-разложение ранга r это оптимальная по Фробениусовой норме аппроксимация ранга r.

Усечённое SVD-разложение ранга r

- Усечённое SVD-разложение ранга r матрицы $M \in \mathbb{R}^{m \times n}$ — это обычное SVD-разложение, где оставили только самые большие сингулярные значения вместе с соответствующими сингулярными векторами.
- Можно доказать, что усечённое SVD-разложение ранга r это оптимальная по Фробениусовой норме аппроксимация ранга r.

Усечённое SVD-разложение ранга r

- Усечённое SVD-разложение ранга r матрицы $M \in \mathbb{R}^{m \times n}$ — это обычное SVD-разложение, где оставили только самые большие сингулярные значения вместе с соответствующими сингулярными векторами.
- Можно доказать, что усечённое SVD-разложение ранга r это оптимальная по Фробениусовой норме аппроксимация ранга r.

Норма Фробениуса, или евклидова норма (для евклидова пространства), представляет собой частный случай p-нормы для $p = 2$:

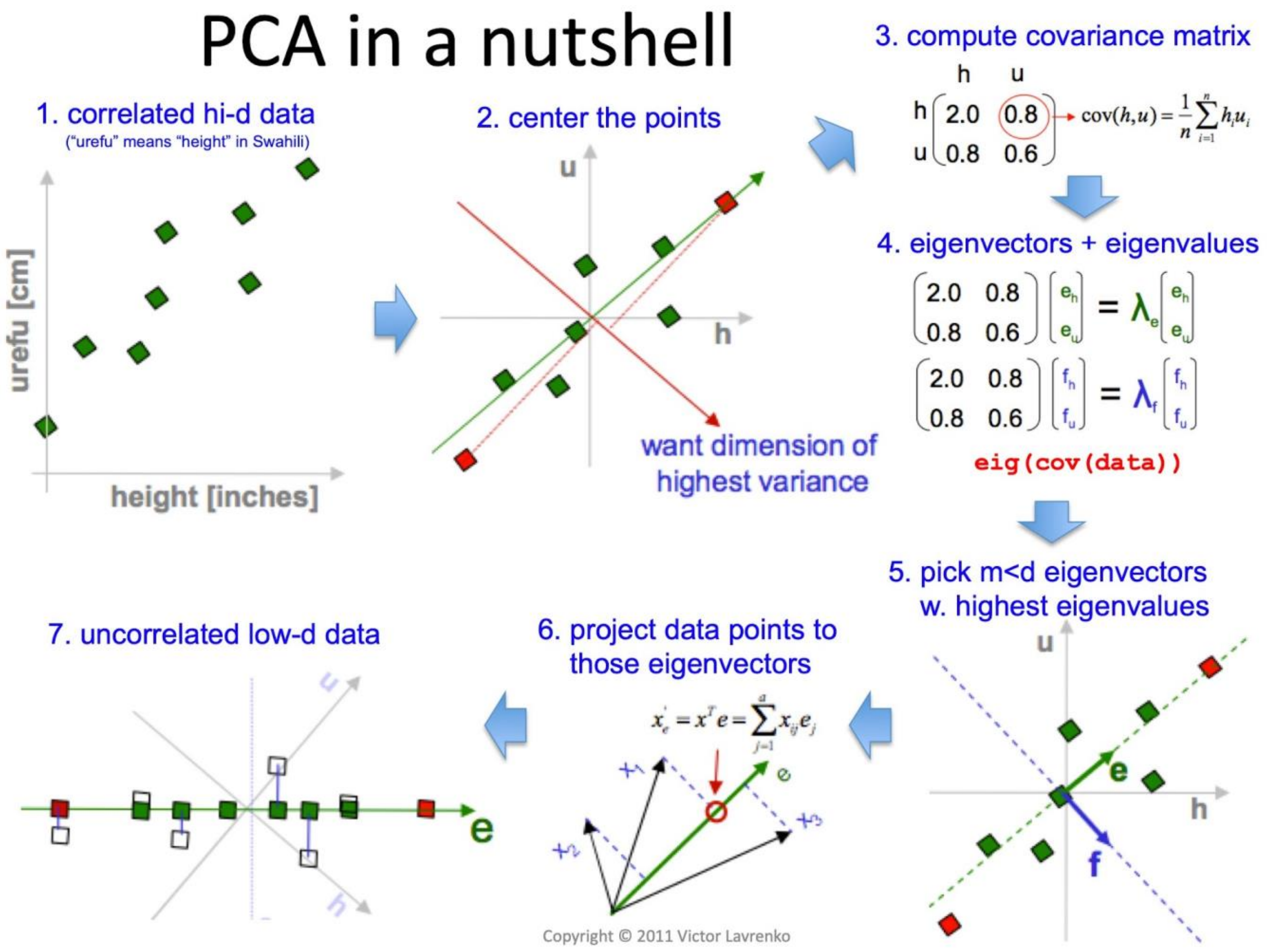
$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

Усечённое SVD-разложение ранга r

- Усечённое SVD-разложение ранга r матрицы $M \in \mathbb{R}^{m \times n}$ — это обычное SVD-разложение, где оставили только самые большие сингулярные значения вместе с соответствующими сингулярными векторами.
- Можно доказать, что усечённое SVD-разложение ранга r это оптимальная по Фробениусовой норме аппроксимация ранга r.
- Поздравляю, именно благодаря этому факту мы можем финализировать алгоритм метода главных компонент! :)

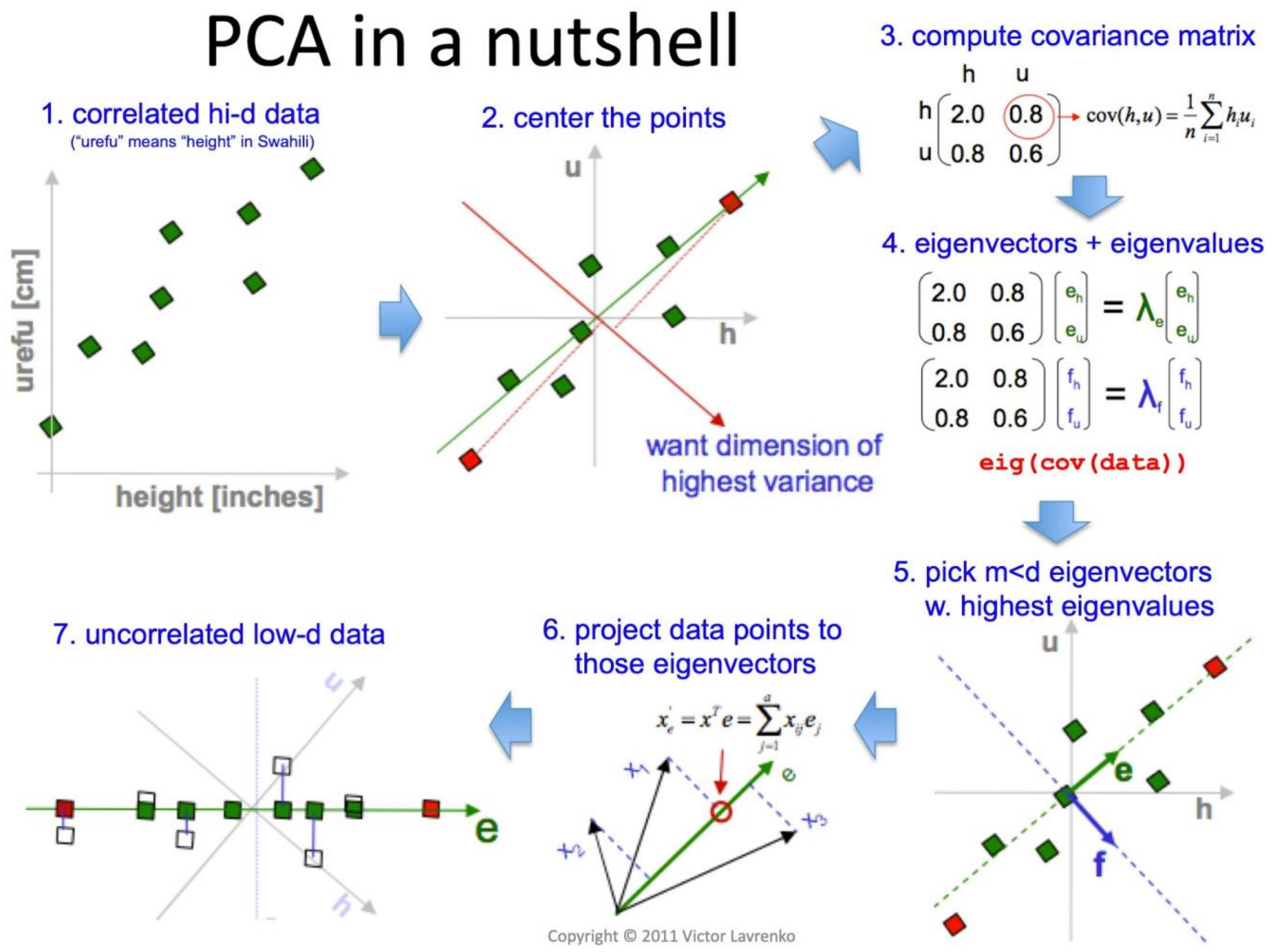
Связь SVD и PCA

- PCA-преобразование можно было бы делать при помощи нахождения собственных векторов (как на алгоритме справа), но к сожалению почти всегда наши датасеты имеют разное число строк и столбцов, и поэтому мы вынуждены использовать SVD.



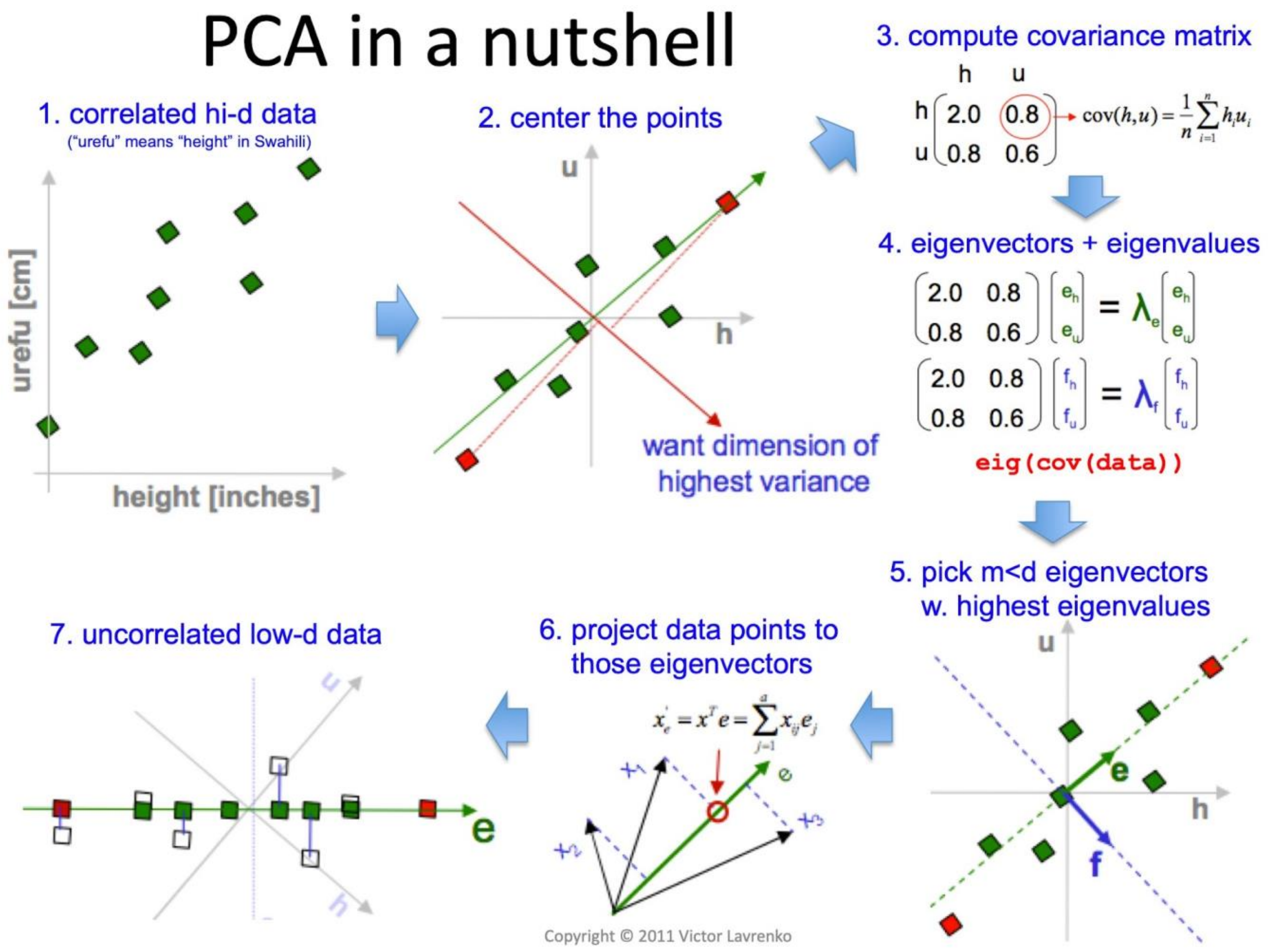
Связь SVD и PCA

- Но есть и хорошая новость! Для того, чтобы получить PCA-преобразование, достаточно всего лишь посчитать $U\Sigma$, то есть произведение первых двух матриц в разложении SVD.



Связь SVD и PCA

- Но есть и хорошая новость! Для того, чтобы получить PCA-преобразование, достаточно всего лишь посчитать $U\Sigma$, то есть произведение первых двух матриц в разложении SVD.
- Подробнее – обязательно разберем на семинарах!

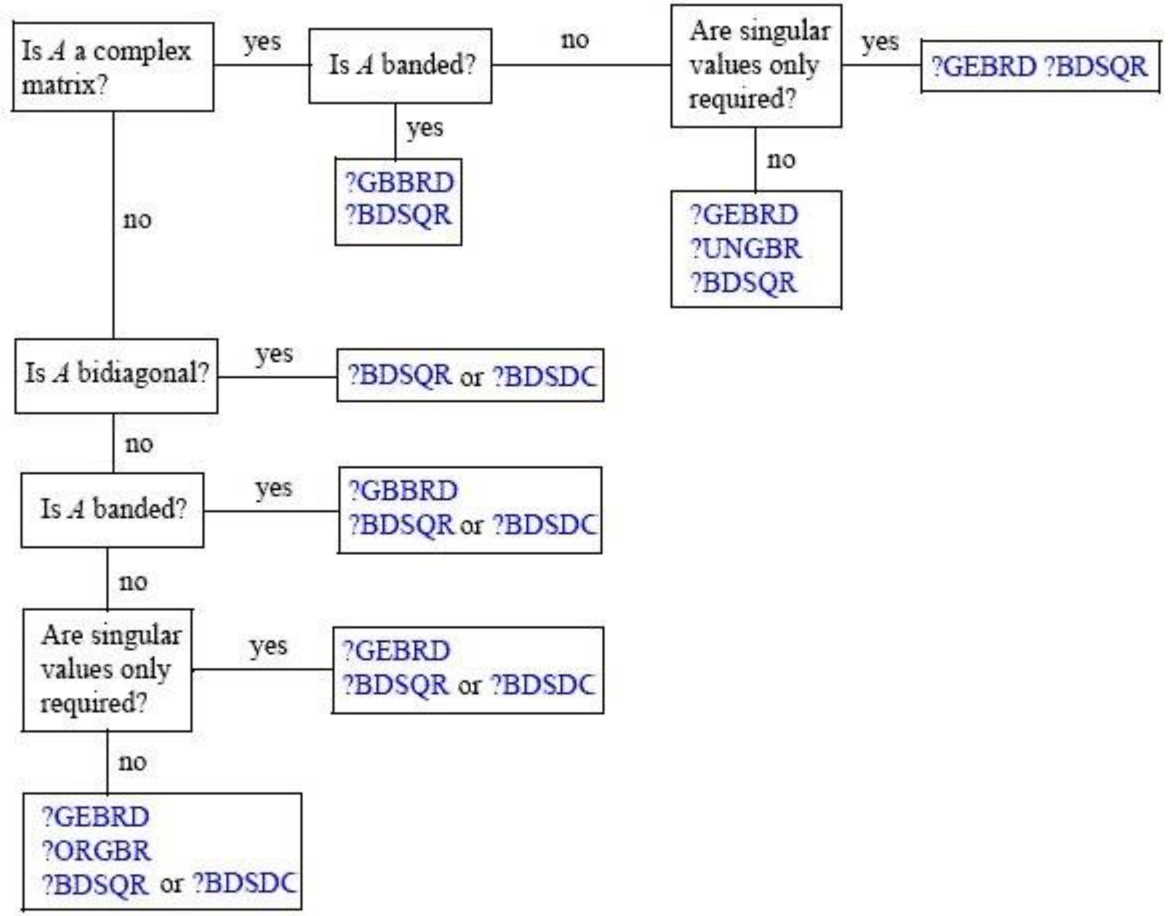


Связь SVD и PCA

- Конкурентное преимущество SVD-разложения в том, что для его вычисления есть эффективные алгоритмы.
- В том числе для огромных разреженных матриц, которые часто встречаются на практике (в рекомендательных системах, биоинформатике etc).

Связь SVD и PCA

Упрощённая схема вычисления SVD-разложения.
На каждый случай — свой алгоритм!
И это ещё — без учёта разреженности матрицы.



- Конкурентное преимущество SVD-разложения в том, что для его вычисления есть эффективные алгоритмы.
- В том числе для огромных разреженных матриц, которые часто встречаются на практике (в рекомендательных системах, биоинформатике etc).

Связь SVD и PCA

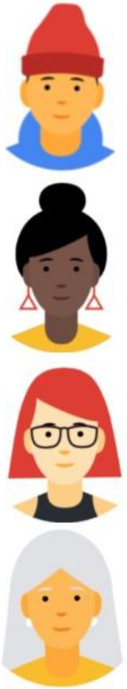
- Конкурентное преимущество SVD-разложения в том, что для его вычисления есть эффективные алгоритмы.
- В том числе для огромных разреженных матриц, которые часто встречаются на практике (в рекомендательных системах, биоинформатике etc).
- Но не PCA единым хорош SVD!

Приложения SVD

Приложения SVD

- В задачах построения рекомендательных систем важную роль играет матрица взаимодействий пользователя с контентом.

	Harry Potter	The Triplets of Belleville	Shrek	The Dark Knight Rises	Memento
■	.9	-.8	1	1	-.9
▲	-.2	-.8	-1	.9	1



●	◆	Harry Potter	The Triplets of Belleville	Shrek	The Dark Knight Rises	Memento
1	.1	✓		✓	✓	
-1	0		✓			✓
.2	-1	✓	✓	✓		
.1	1			?	✓	✓

■ arthouse <-> blockbuster

▲ children's <-> adult's

● preference for arthouse <-> blockbuster

◆ preference for children's <-> adult's

Приложения SVD

- В задачах построения рекомендательных систем важную роль играет матрица взаимодействий пользователя с контентом.
- Представьте Netflix (хорошо нам знакомый):

Приложения SVD

- В задачах построения рекомендательных систем важную роль играет матрица взаимодействий пользователя с контентом.
- Представьте Netflix (хорошо нам знакомый):
  - Миллионы пользователей;
  - Тысячи фильмов и сериалов;
  - Каждый пользователь, в среднем, смотрит < 100 из них.

Приложения SVD

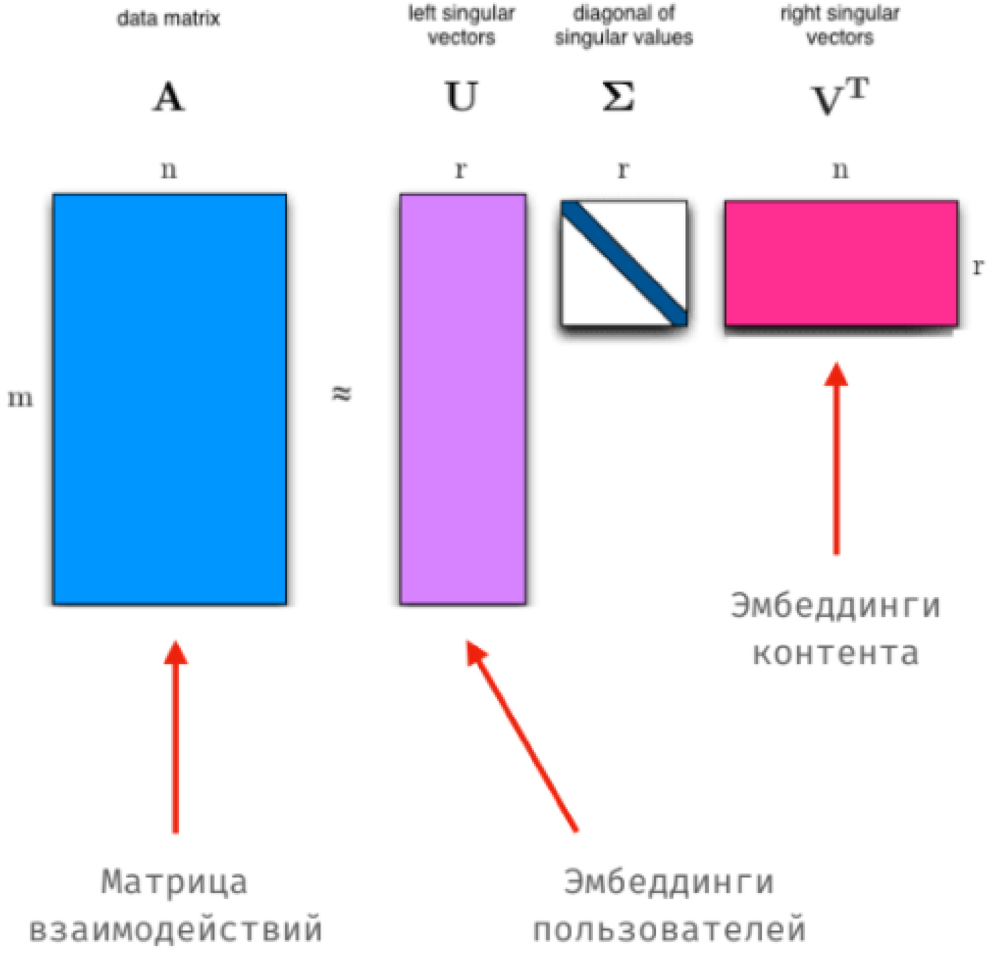
- В задачах построения рекомендательных систем важную роль играет матрица взаимодействий пользователя с контентом.
- Представьте Netflix (хорошо нам знакомый):
  - Миллионы пользователей;
  - Тысячи фильмов и сериалов;
  - Каждый пользователь, в среднем, смотрит < 100 из них.
- Что можно сказать о матрице оценок?

Приложения SVD

- В задачах построения рекомендательных систем важную роль играет матрица взаимодействий пользователя с контентом.
- Представьте Netflix (хорошо нам знакомый):
    - Миллионы пользователей;
    - Тысячи фильмов и сериалов;
    - Каждый пользователь, в среднем, смотрит < 100 из них.
- Что можно сказать о матрице оценок?
- Гигантская разреженная матрица. На пересечении (пользователь, столбец) — оценка или любая другая полезная информация ("посмотрел ли до конца" и другие прокси-метрики)

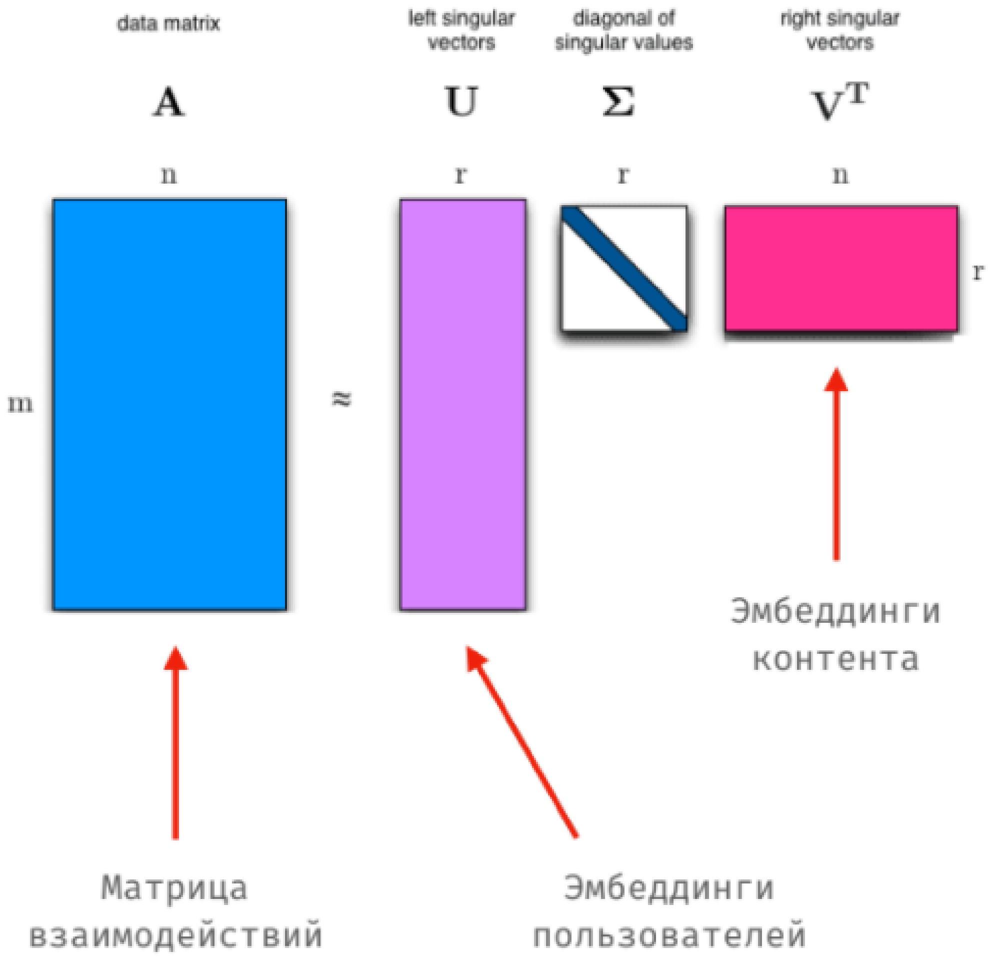
Приложения SVD

- Усеченное SVD-разложение ранга r матрицы взаимодействий позволяет получить эмбеддинги размерности r как для пользователей, так и для контента!



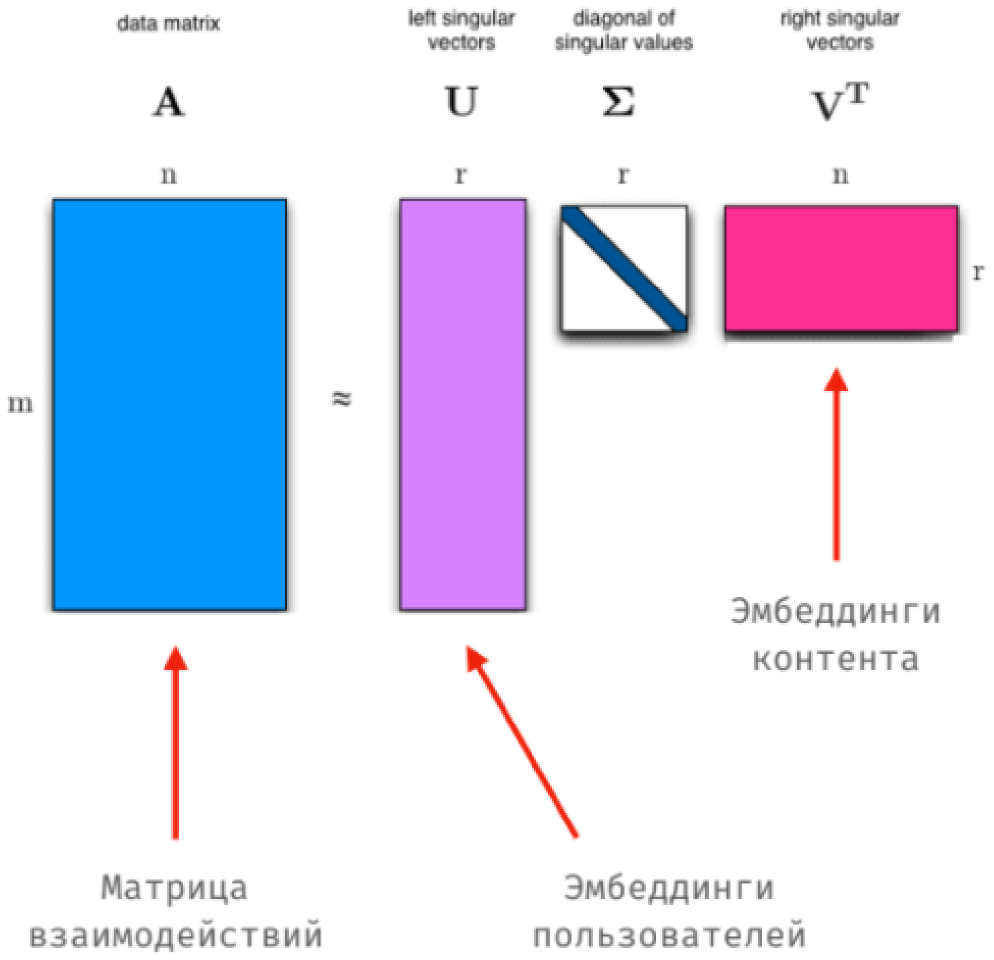
Приложения SVD

- Усеченное SVD-разложение ранга r матрицы взаимодействий позволяет получить эмбеддинги размерности r как для пользователей, так и для контента!
- Более того: чем больше скалярное произведение эмбеддингов пользователя и контента, тем выше шанс, что пользователю понравится контент!



Приложения SVD

- Усеченное SVD-разложение ранга r матрицы взаимодействий позволяет получить эмбеддинги размерности r как для пользователей, так и для контента!
- Более того: чем больше скалярное произведение эмбеддингов пользователя и контента, тем выше шанс, что пользователю понравится контент!
- Это позволяет построить простейшую рекомендательную систему!



Приложения SVD

- Рекомендательные системы на основе коллаборативной фильтрации — основа рекомендательных систем в Яндексе — в рекламе, Дзене и т.д.

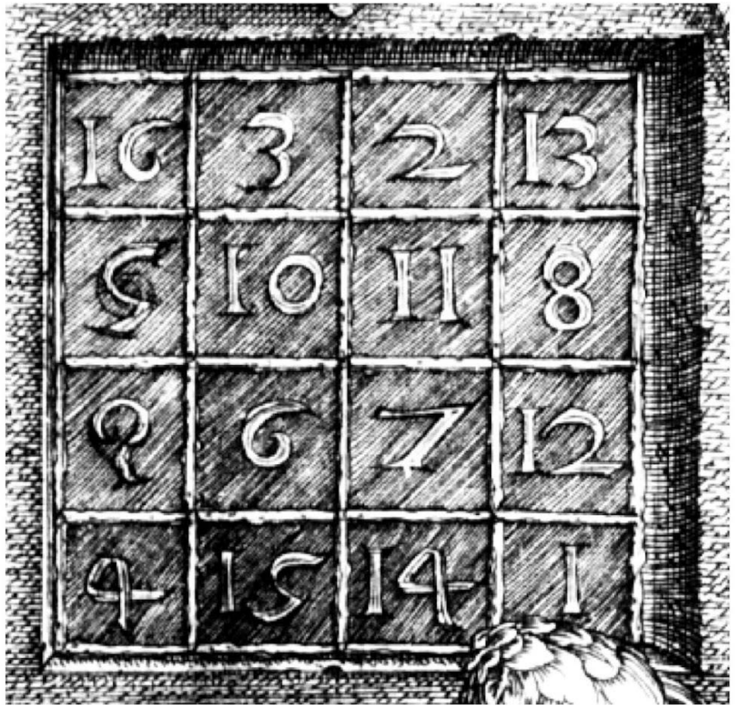
Приложения SVD

- Рекомендательные системы на основе коллаборативной фильтрации — основа рекомендательных систем в Яндексе — в рекламе, Дзене и т.д.
- Разумеется, там используются более продвинутые матричные факторизации.
- Но идея та же!

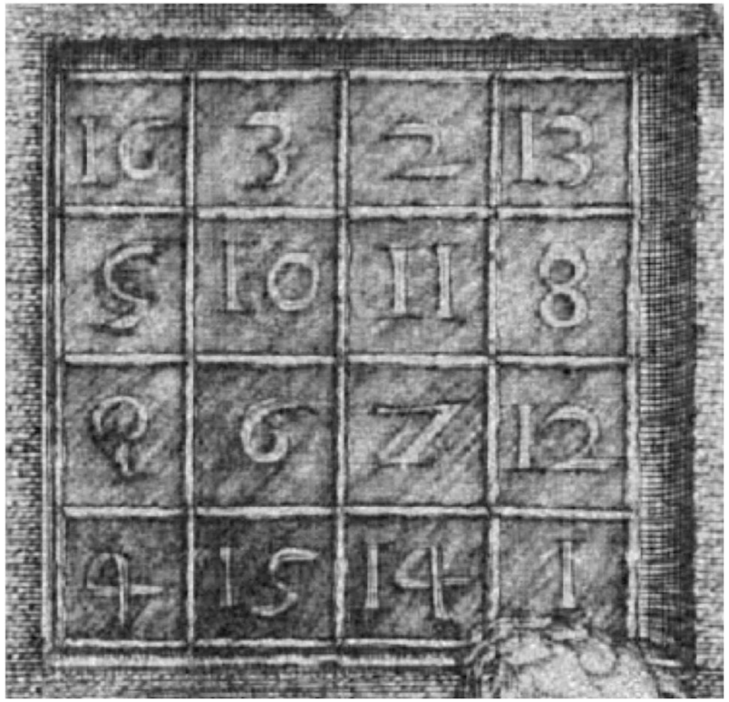
Приложения SVD

- Кроме того, SVD-разложение можно использовать как механизм сжатия изображений и видео…

Detail from Durer's Melancolia, dated 1514., 359x371 image



EOF reconstruction with 50 modes

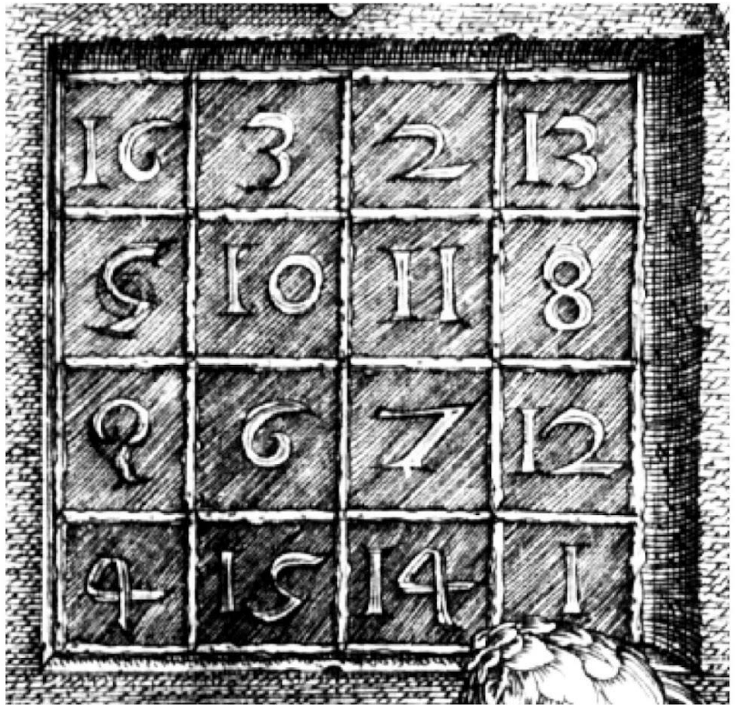


Приложения SVD

Конечно же, мы еще вернемся к этому на семинарах!

- Кроме того, SVD-разложение можно использовать как механизм сжатия изображений и видео…

Detail from Durer's Melancolia, dated 1514., 359x371 image



EOF reconstruction with 50 modes