

Линейная алгебра для машинного обучения

Познакомимся?

Маргарита Бурова

- преподаватель департамента больших данных и информационного поиска НИУ ВШЭ
- автор онлайн-курсов НИУ ВШЭ
- магистр прикладной математики и информатики НИУ ВШЭ
- руководитель онлайн-курса по подготовке к ЕГЭ по математике



План вебинара:

ОСНОВНЫЕ МОМЕНТЫ



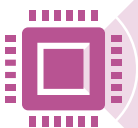
Зачем нужна линейная алгебра?



Линейная алгебра в NLP



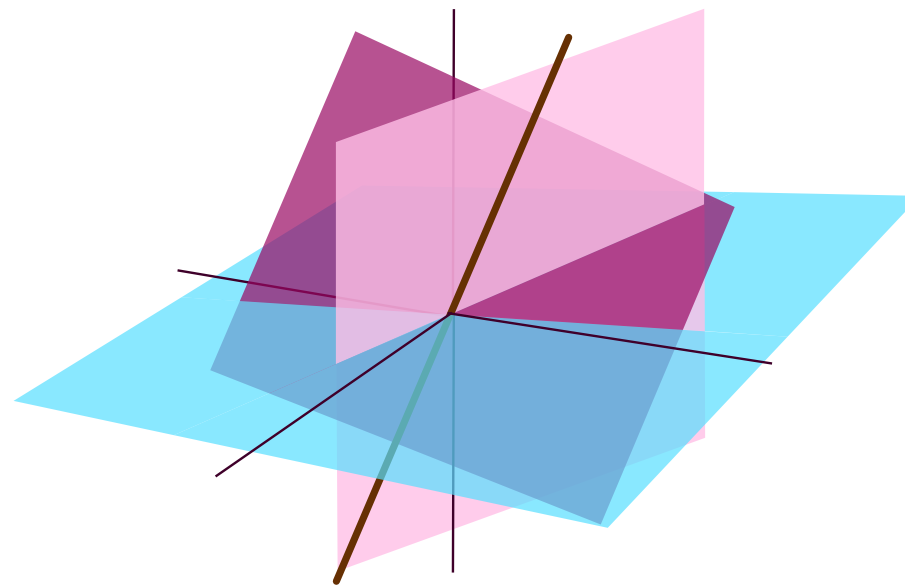
Линейная алгебра в Машинном обучении



Линейная алгебра в снижении размерности



Линейная алгебра в компьютерном зрении



Линейная алгебра- **ЗАЧЕМ**

Линейная алгебра – хлеб насущный для всей науки

Линейная алгебра- ЗАЧЕМ

ML - интуиция

Создание алгоритмов

Структурный элемент

Настройка алгоритмов

Реализация алгоритмов

Линейная алгебра- ЗАЧЕМ

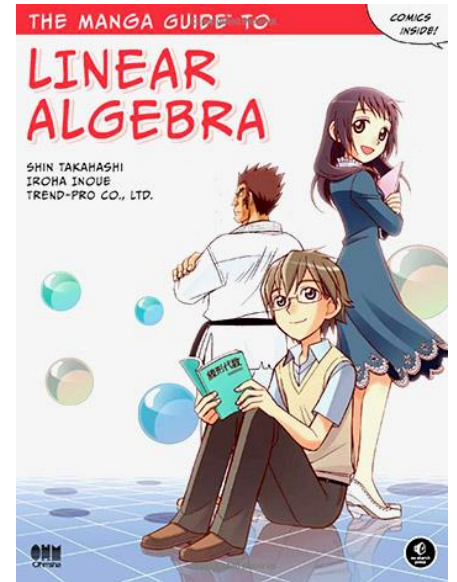


Уровни ПОНИМАНИЯ

- **Поверхностный**— основная рабочая лошадка; уровень «must have»

курс OCW-MIT Гилберта Стрэнга по линейной алгебре + его книжка.

ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/index.htm



- **Для любителей челленджей**— уровень повыше, позволяет оглядеть с высоты птичьего полёта уровень 1, систематизирует знания, объединяет различные области знаний

Axler «*Linear algebra done right*»; Гельфанд «Лекции по линейной алгебре»; курс МФТИ

lectoriy.mipt.ru/course/LinearAlgebra; Кострикин «Введение в алгебру, часть 2», Тиртышников «Матричный анализ и линейная алгебра».

- **Ночной ужас** — для сильных духом, уровень мехмата

Кострикин-Манин «Линейная алгебра и геометрия», Шафаревич-Ремизов «Линейная алгебра и геометрия».

MUST-HAVE ТЕМЫ

Векторы и векторное пространство

Матричные разложения

Операторы

Собственные значения и собственные вектора

Связь линейных операторов и
матриц

Квадратичные формы

4 основных объекта

СКАЛЯР

1

ВЕКТОР

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

МАТРИЦА

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

ТЕНЗОР

$$\begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} & \begin{bmatrix} 3 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & 7 \end{bmatrix} & \begin{bmatrix} 5 & 4 \end{bmatrix} \end{bmatrix}$$

Классическая история: датасет как матрица

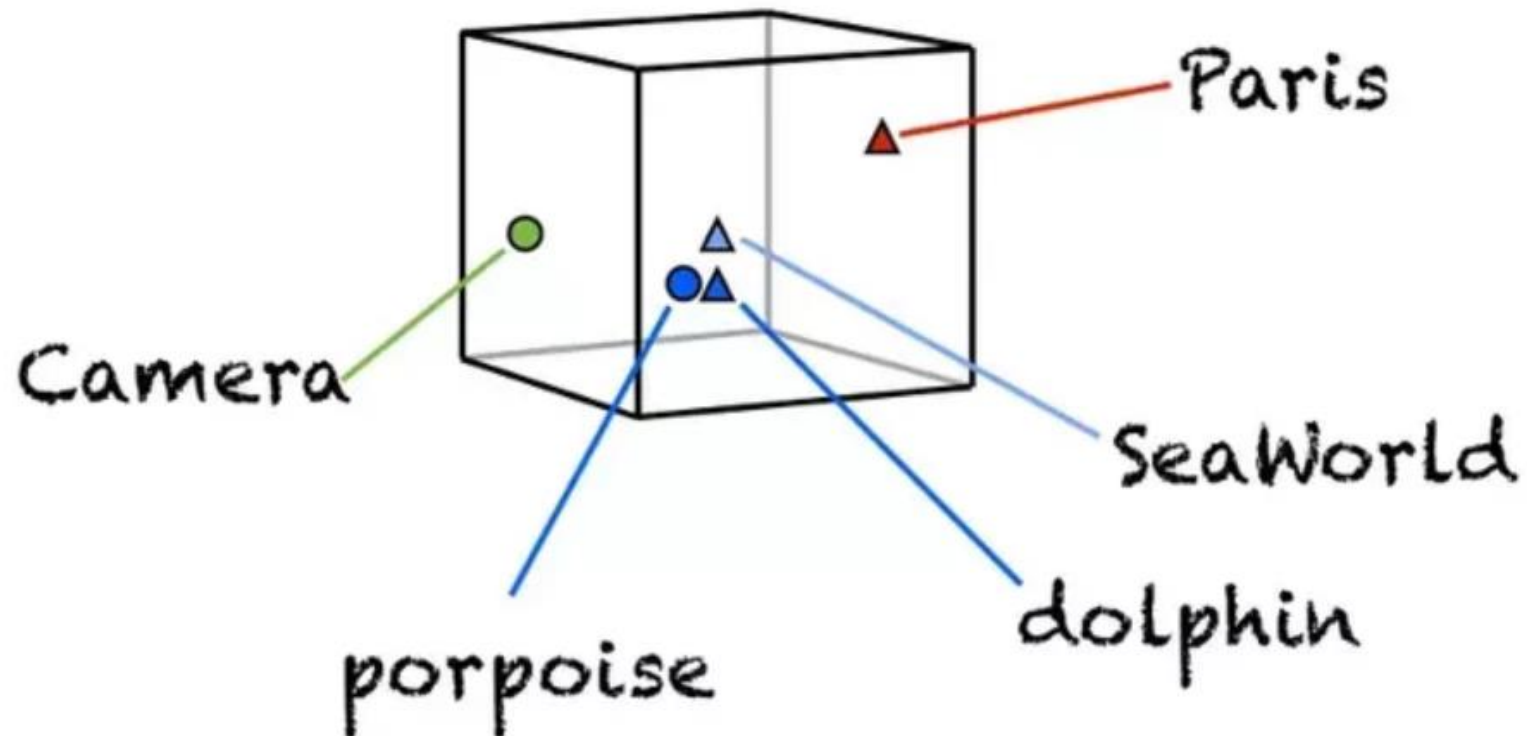
объекты

признаки

	Sex	Race	Height	Income	Marital Status	Years of Educ.
R1001	M	1	70	50	1	12
R1002	M	2	72	100	2	20
R1003	F	1	55	250	1	16
R1004	M	2	65	20	2	16
R1005	F	1	60	10	3	12
R1006	M	1	68	30	1	9
R1007	F	5	66	25	2	21
R1008	F	4	61	43	1	18
R1009	M	1	69	67	1	12

Векторы: word embedding

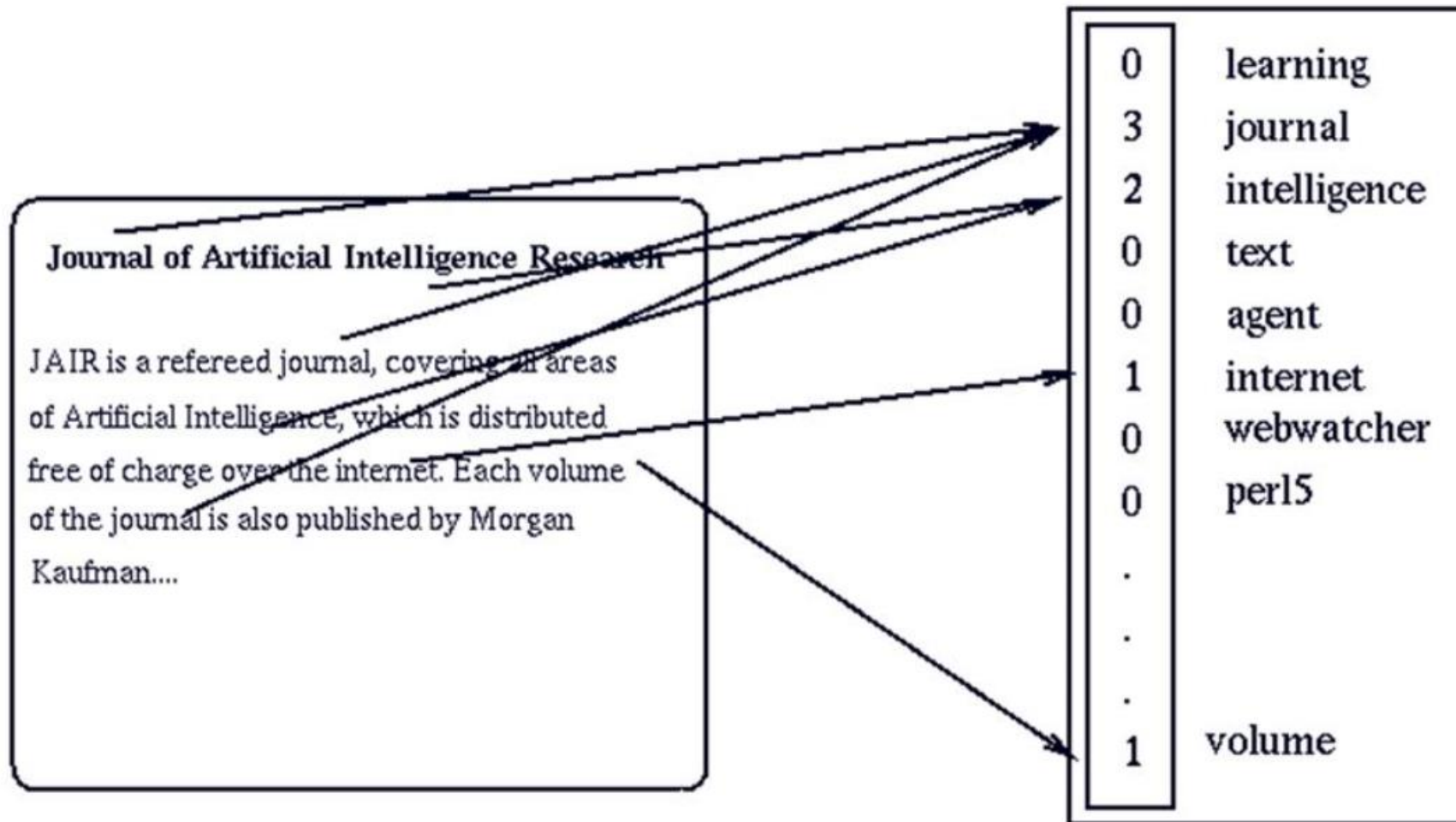
→ сопоставление
произвольной
сущности
какому-то
вектору



Как сделать из слова вектор?

motel [0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0] AND
hotel [0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0] = 0

Как сделать из слова вектор?



Как оценить схожесть текстов?

Рассмотрим пример. Пусть

S1= 'купить пластиковые окна со скидкой',

S2='купить недорого пластиковые окна с бесплатной доставкой по Москве'

$$\cos = \frac{1 \cdot 1 + 1 \cdot 1 + 0 \cdot 1 + 1 \cdot 1 + 1 \cdot 0 + 0 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 + 0 \cdot 1}{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2} \cdot \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2}}$$

Слов	купить	пластиковые	недорого	окна	скидко	с	бесплатной	Москве	по	со	доставкой
В	1	1	0	1	1	0	0	0	0	1	0
В	1	1	1	1	0	1	1	1	1	0	1

$$\frac{3}{\sqrt{5} \cdot \sqrt{5}} = \frac{1}{\sqrt{5}} \approx 0,45$$

$$\cos D = 1 - 0,45 = 0,55$$

расстояние

Скалярное произведение aka косинусное расстояние

$$\vec{a}(x_1; y_1)$$

$$x_1 x_2 + y_1 y_2$$

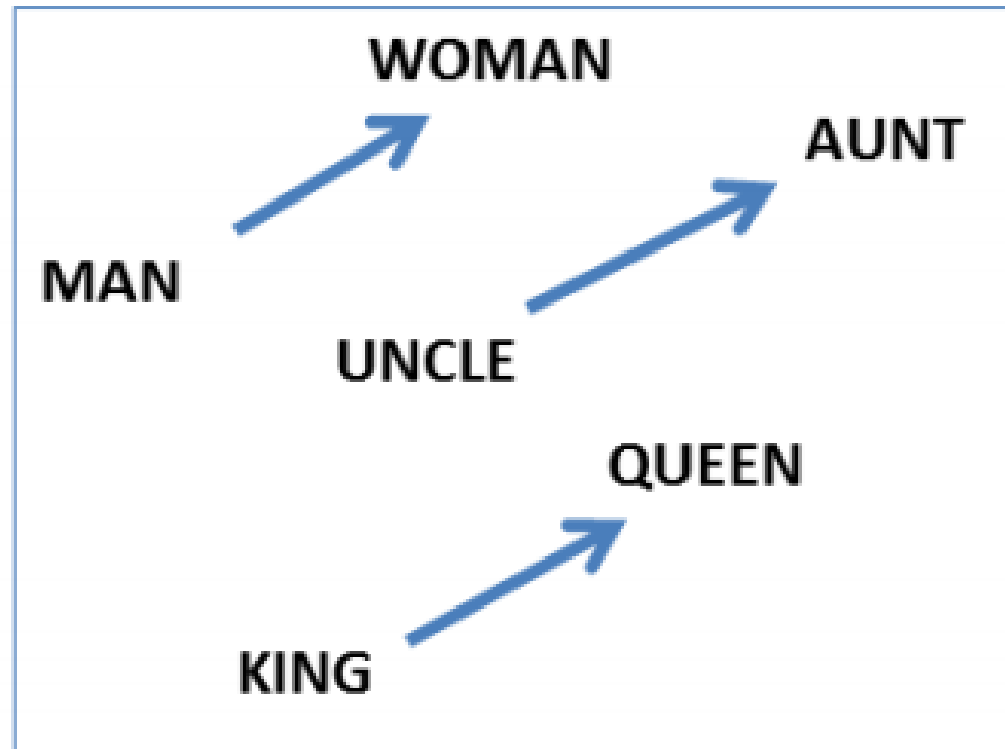
$$\vec{b}(x_2; y_2)$$

$$|\vec{a}| \cdot |\vec{b}| \cdot \cos(\hat{a}, \vec{b})$$

$$\cos(a; b) = \frac{x_1 x_2 + y_1 y_2}{|\vec{a}| \cdot |\vec{b}|} =$$

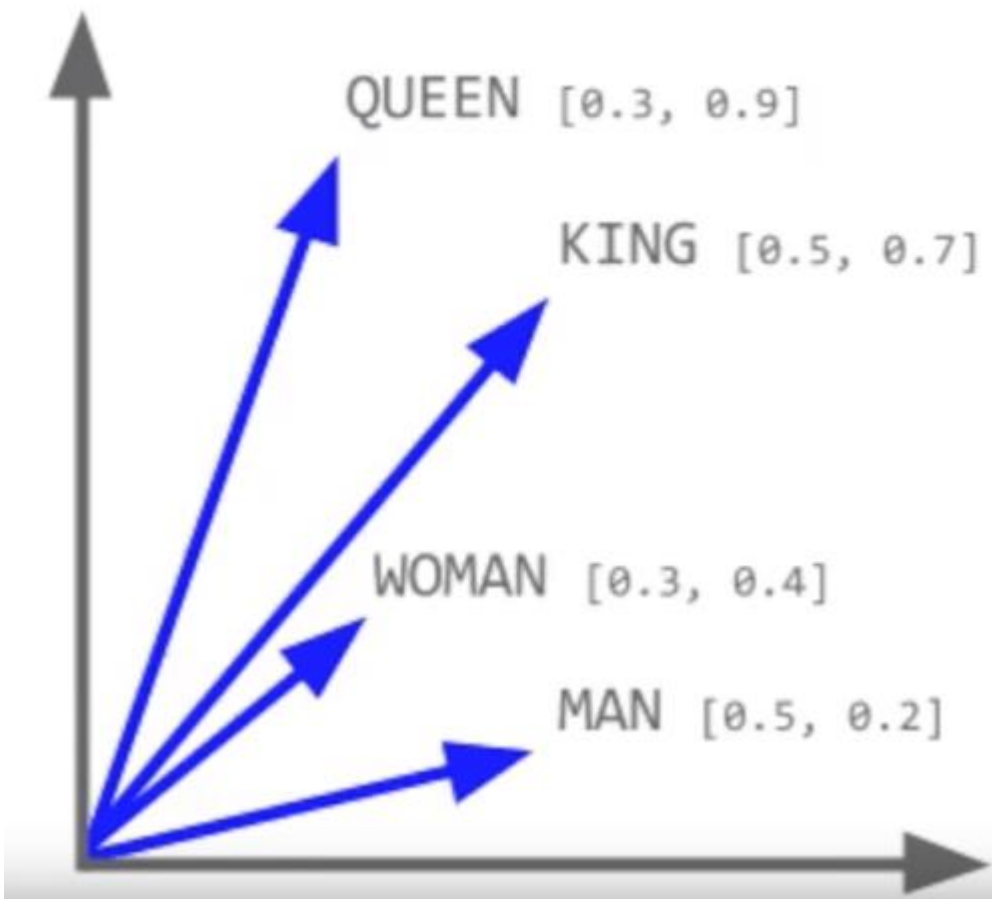
$$= \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2} \cdot \sqrt{x_2^2 + y_2^2}}$$

Спасение всех дата саентистов by Томаш Миколов



Спасение всех дата саентистов by Томаш Миколов

Load up the word vectors



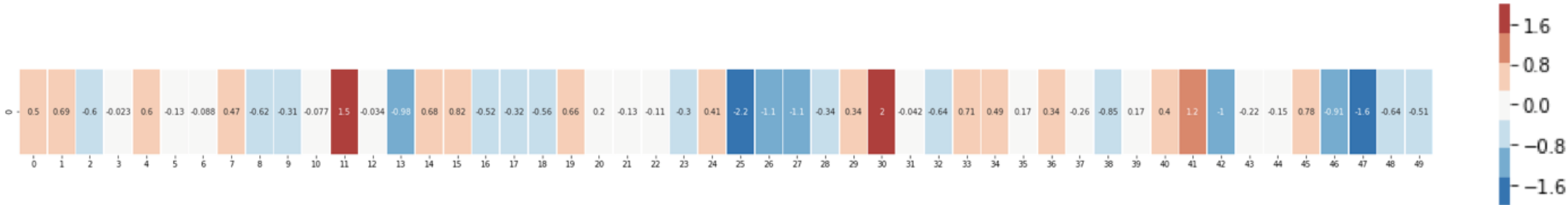
$$\begin{aligned} & \text{KING} - \text{MAN} \\ & (0.5, 0.7) - (0.5, 0.2) = \\ & = (0, 0.5) \end{aligned}$$

$$\begin{aligned} & (0, 0.5) + (0.3, 0.4) = \\ & = (0.3, 0.9) \end{aligned}$$

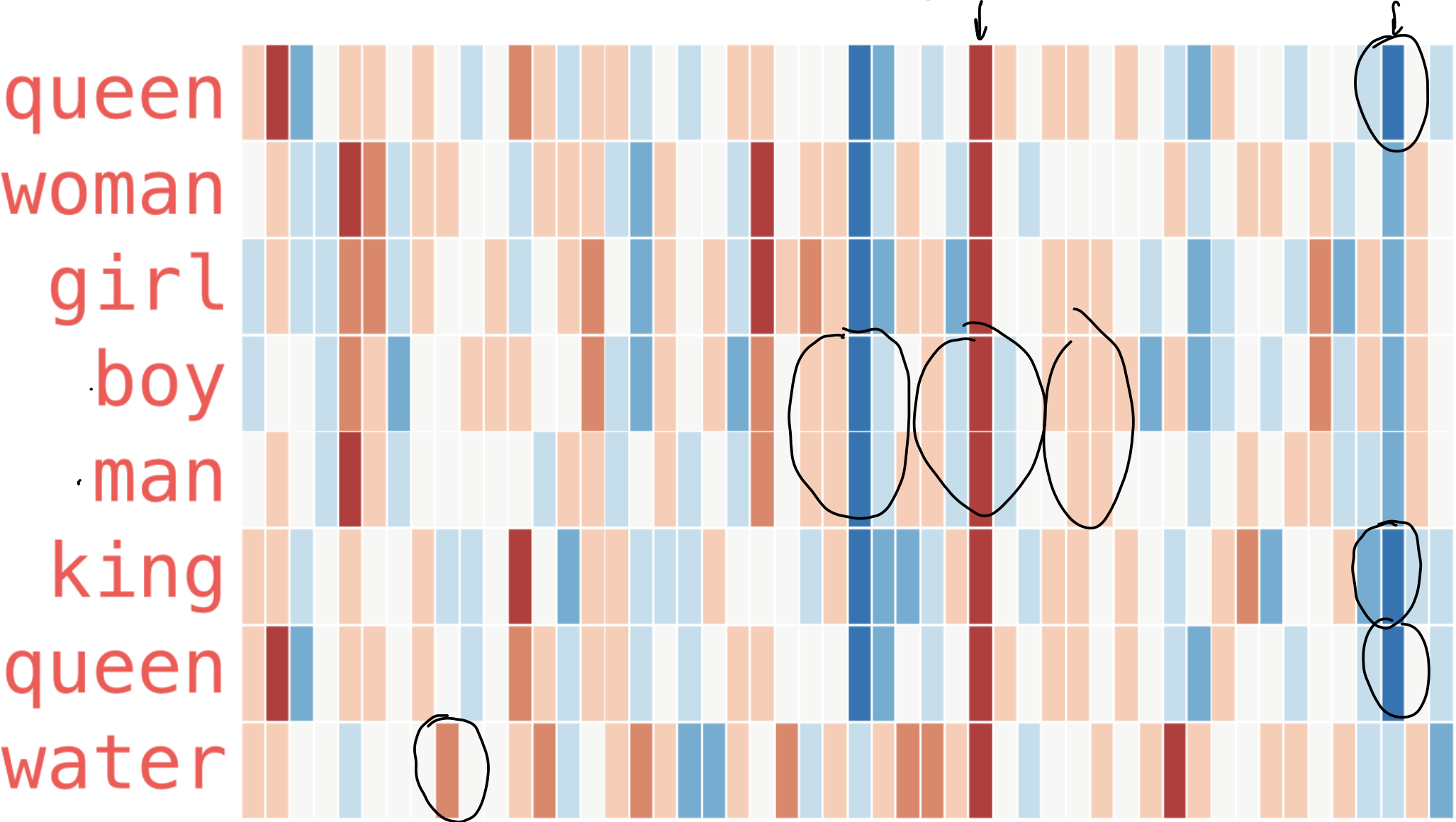
Спасение всех дата саентистов by Томаш Миколов

Король:

[0.50451 , 0.68607 , -0.59517 , -0.022801, 0.60046 , -0.13498 ,
-0.08813 , 0.47377 , -0.61798 , -0.31012 , -0.076666, 1.493 , -
0.034189, -0.98173 , 0.68229 , 0.81722 , -0.51874 , -0.31503 , -
0.55809 , 0.66421 , 0.1961 , -0.13495 , -0.11476 , -0.30344 ,
0.41177 , -2.223 , -1.0756 , -1.0783 , -0.34354 , 0.33505 ,
1.9927 , -0.04234 , -0.64319 , 0.71125 , 0.49159 , 0.16754 ,
0.34344 , -0.25663 , -0.8523 , 0.1661 , 0.40102 , 1.1685 , -
1.0137 , -0.21585 , -0.15155 , 0.78321 , -0.91241 , -1.6106 , -]



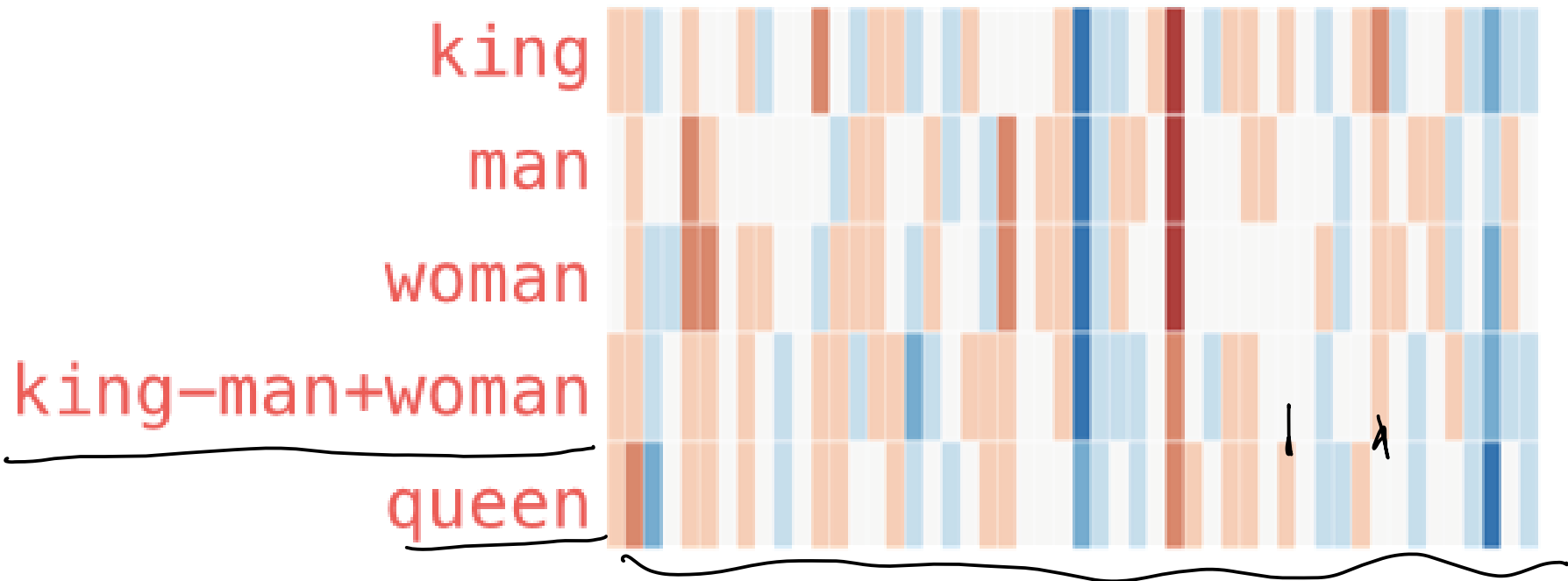
Спасение всех дата саентистов by Томаш Миколов



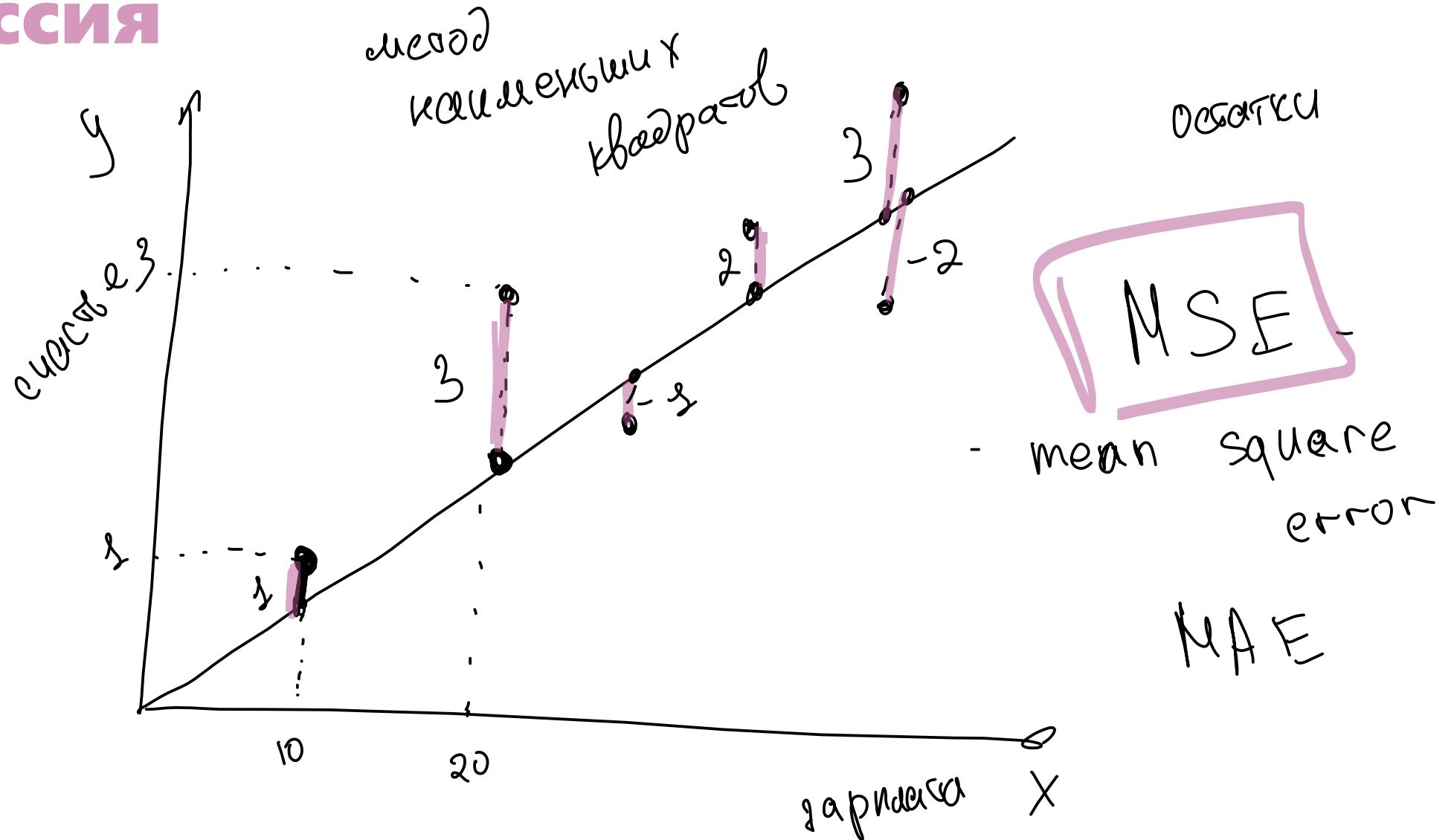
Спасение всех дата саентистов by Томаш Миколов

king - man + woman ≈ queen

word2vec
pip install
word2vec

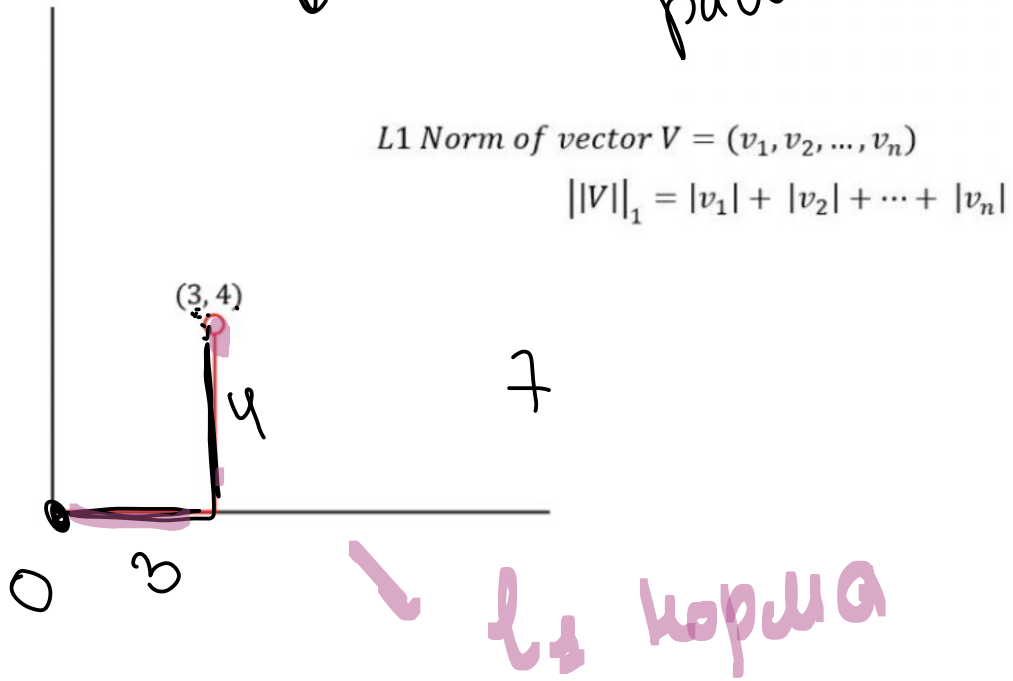


Линейная алгебра и машинное обучение: регрессия

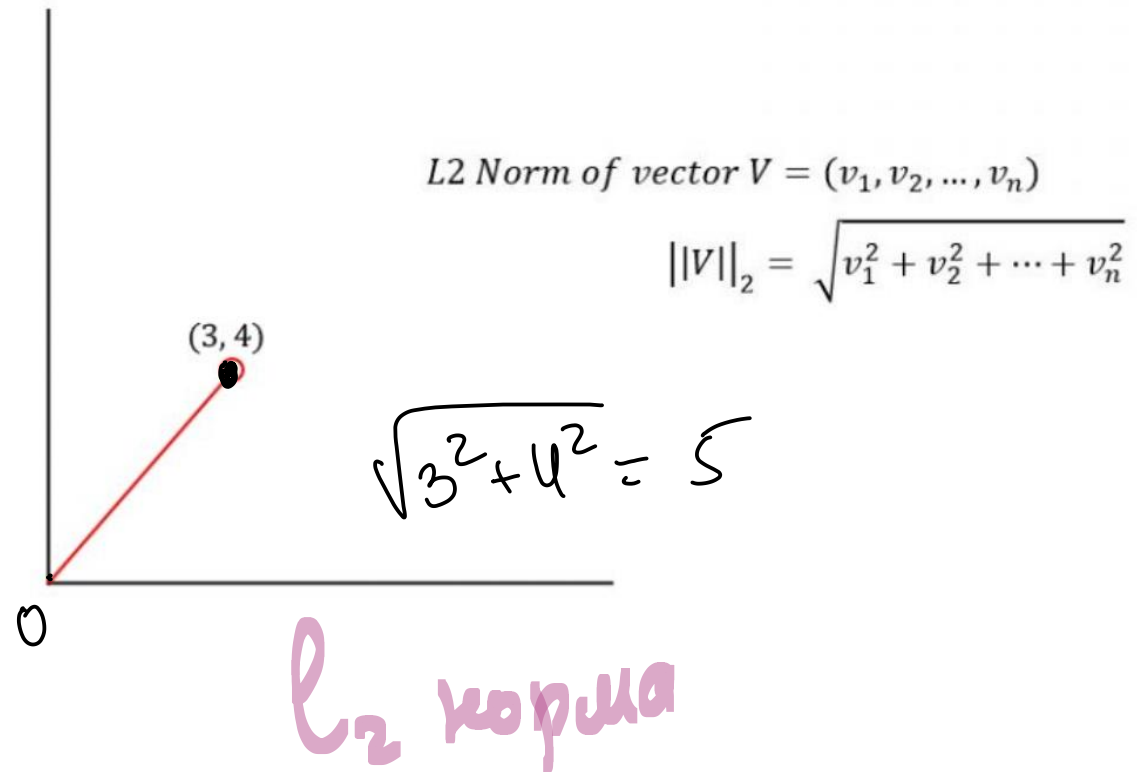


Линейная алгебра и машинное обучение: LOSS FUNCTION

↙ манхеттенское
расстояние



расстояние
Евклида



Матрицы для представления всего на свете

Изображения



```
0 2 15 0 0 11 10 0 0 0 0 9 9 0 0 0
0 0 0 4 60 157 236 255 255 177 95 61 32 0 0 29
0 10 16 119 238 255 244 245 243 250 249 255 222 103 10 0
0 14 170 255 255 244 254 255 253 245 255 249 253 251 124 1
2 98 255 228 255 251 254 211 141 116 122 215 251 238 255 49
13 217 243 255 155 33 226 52 2 0 10 13 232 255 255 36
16 229 252 254 49 12 0 0 7 7 0 70 237 252 235 62
6 141 245 255 212 25 11 9 3 0 115 236 243 255 137 0
0 87 252 250 248 215 60 0 1 121 252 255 248 144 6 0
0 13 113 255 255 245 255 182 181 248 252 242 208 36 0 19
1 0 5 117 251 255 241 255 247 255 241 162 17 0 7 0
0 0 0 4 58 251 255 246 254 253 255 120 11 0 1 0
0 0 4 97 255 255 255 248 252 255 244 255 182 10 0 4
0 22 206 252 246 251 241 100 24 113 255 245 255 194 9 0
0 111 255 242 255 158 24 0 0 6 39 255 232 230 56 0
0 218 251 250 137 7 11 0 0 0 2 62 255 250 125 3
0 173 255 255 101 9 20 0 13 3 13 182 251 245 61 0
0 107 251 241 255 230 98 55 19 118 217 248 253 255 52 4
0 18 146 250 255 247 255 255 255 249 255 240 255 129 0 5
0 0 23 113 215 255 250 248 255 255 248 248 118 14 12 0
0 0 6 1 0 52 153 233 255 252 147 37 0 0 4 1
0 0 5 5 0 0 0 0 0 14 1 0 6 6 0 0
```



0	2	15	0	0	11	10	0	0	0	0	9	9	0	0	0
0	0	0	4	60	157	236	255	255	177	95	61	32	0	0	29
0	10	16	119	238	255	244	245	243	250	249	255	222	103	10	0
0	14	170	255	255	244	254	255	253	245	255	249	253	251	124	1
2	98	255	228	255	251	254	211	141	116	122	215	251	238	255	49
13	217	243	255	155	33	226	52	2	0	10	13	232	255	255	36
16	229	252	254	49	12	0	0	7	7	0	70	237	252	235	62
6	141	245	255	212	25	11	9	3	0	115	236	243	255	137	0
0	87	252	250	248	215	60	0	1	121	252	255	248	144	6	0
0	13	113	255	255	245	255	182	181	248	252	242	208	36	0	19
1	0	5	117	251	255	241	255	247	255	241	162	17	0	7	0
0	0	0	4	58	251	255	246	254	253	255	120	11	0	1	0
0	0	4	97	255	255	255	248	252	255	244	255	182	10	0	4
0	22	206	252	246	251	241	100	24	113	255	245	255	194	9	0
0	111	255	242	255	158	24	0	0	6	39	255	232	230	56	0
0	218	251	250	137	7	11	0	0	0	2	62	255	250	125	3
0	173	255	255	101	9	20	0	13	3	13	182	251	245	61	0
0	107	251	241	255	230	98	55	19	118	217	248	253	255	52	4
0	18	146	250	255	247	255	255	255	249	255	240	255	129	0	5
0	0	23	113	215	255	250	248	255	255	248	248	118	14	12	0
0	0	6	1	0	52	153	233	255	252	147	37	0	0	4	1
0	0	5	5	0	0	0	0	0	14	1	0	6	6	0	0

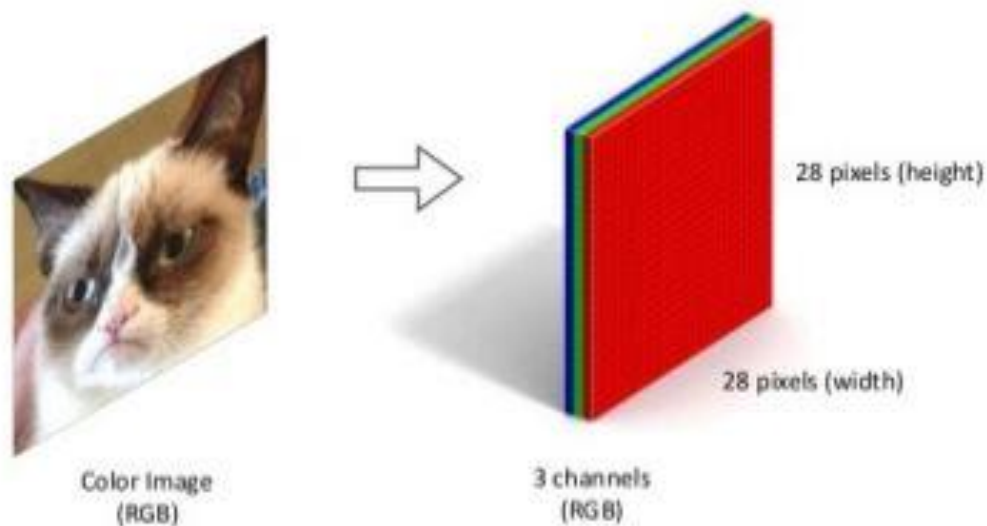
```
0 2 15 0 0 11 10 0 0 0 0 9 9 0 0 0
0 0 0 4 60 157 236 255 255 177 95 61 32 0 0 29
0 10 16 119 238 255 244 245 243 250 249 255 222 103 10 0
0 14 170 255 255 244 254 255 253 245 255 249 253 251 124 1
2 98 255 228 255 251 254 211 141 116 122 215 251 238 255 49
13 217 243 255 155 33 226 52 2 0 10 13 232 255 255 36
16 229 252 254 49 12 0 0 7 7 0 70 237 252 235 62
6 141 245 255 212 25 11 9 3 0 115 236 243 255 137 0
0 87 252 250 248 215 60 0 1 121 252 255 248 144 6 0
0 13 113 255 255 245 255 182 181 248 252 242 208 36 0 19
1 0 5 117 251 255 241 255 247 255 241 162 17 0 7 0
0 0 0 4 58 251 255 246 254 253 255 120 11 0 1 0
0 0 4 97 255 255 255 248 252 255 244 255 182 10 0 4
0 22 206 252 246 251 241 100 24 113 255 245 255 194 9 0
0 111 255 242 255 158 24 0 0 6 39 255 232 230 56 0
0 218 251 250 137 7 11 0 0 0 2 62 255 250 125 3
0 173 255 255 101 9 20 0 13 3 13 182 251 245 61 0
0 107 251 241 255 230 98 55 19 118 217 248 253 255 52 4
0 18 146 250 255 247 255 255 255 249 255 240 255 129 0 5
0 0 23 113 215 255 250 248 255 255 248 248 118 14 12 0
0 0 6 1 0 52 153 233 255 252 147 37 0 0 4 1
0 0 5 5 0 0 0 0 0 14 1 0 6 6 0 0
```


Матрицы для представления всего на свете

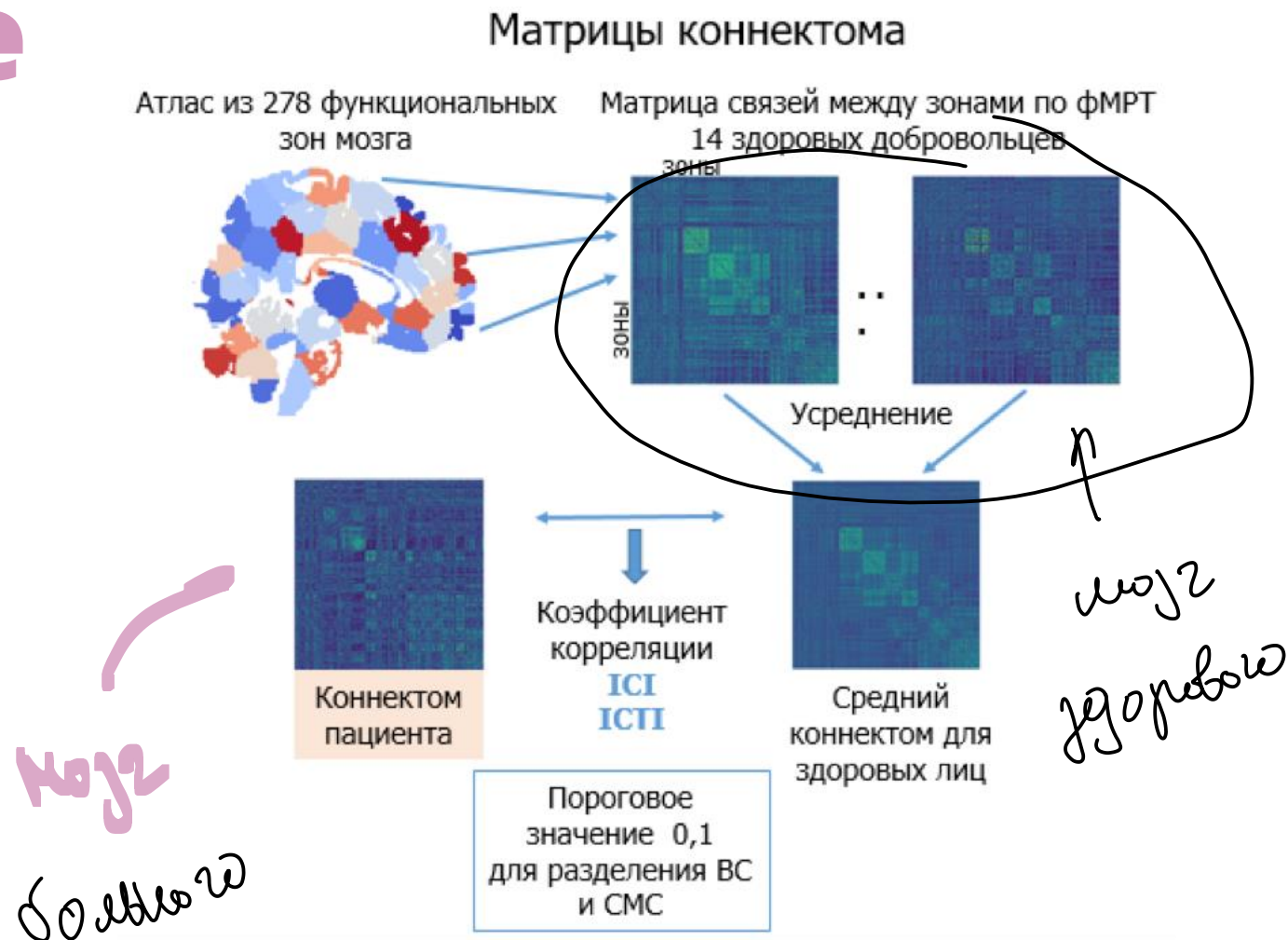
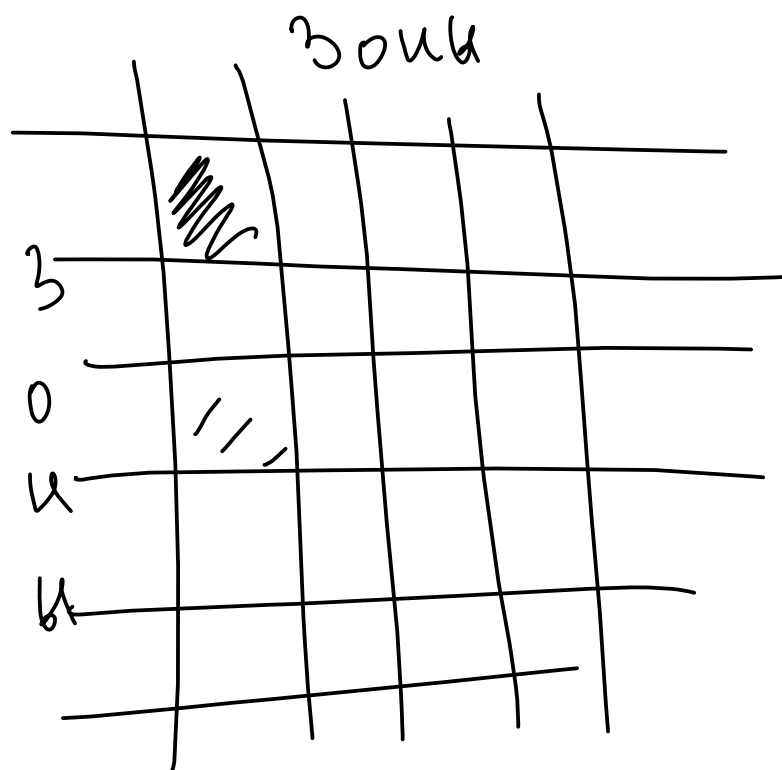


Матрицы для представления всего на свете

color image is 3rd-order tensor



Матрицы для представления всего на свете



ВОПРОСЫ?