



Aggregations of Fuzzy Equivalences in k-means Algorithm

28TH INTERNATIONAL CONFERENCE ON KNOWLEDGE-BASED AND INTELLIGENT
INFORMATION & ENGINEERING SYSTEMS (KES 2024)

Авторы: Piotr Lasek, Wojciech Rząsa, Anna Król

Презентовано: Горельский Роман

GitHub репозиторий: <https://github.com/RomanGorelsky/Aggregations-of-Fuzzy-Equivalences-in-k-means-Algorithm>



Содержание

01

Введение

02

Цель и
идея

03

Нечеткие
эквивалентност
и

04

Агрегационные
функции

05

Модификации
 k -средних

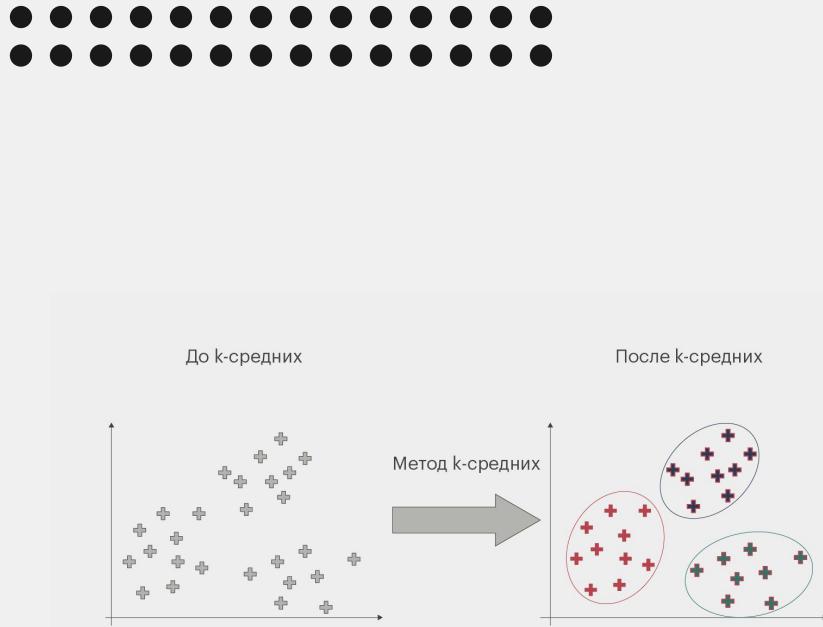
06

Эксперимент
ы и
результаты

Введение

Кластеризация (особенно k-means) широко используется для **группировки данных** по схожести признаков в задачах анализа данных, машинного обучения и распознавания образов.

Классический k-means опирается на **евклидово расстояние**. Однако данный подход **плохо работает** для сложных структур данных, высоких размерностей и несферических кластеров, а также сильно **зависит от начальных центров**.





Идея и цели

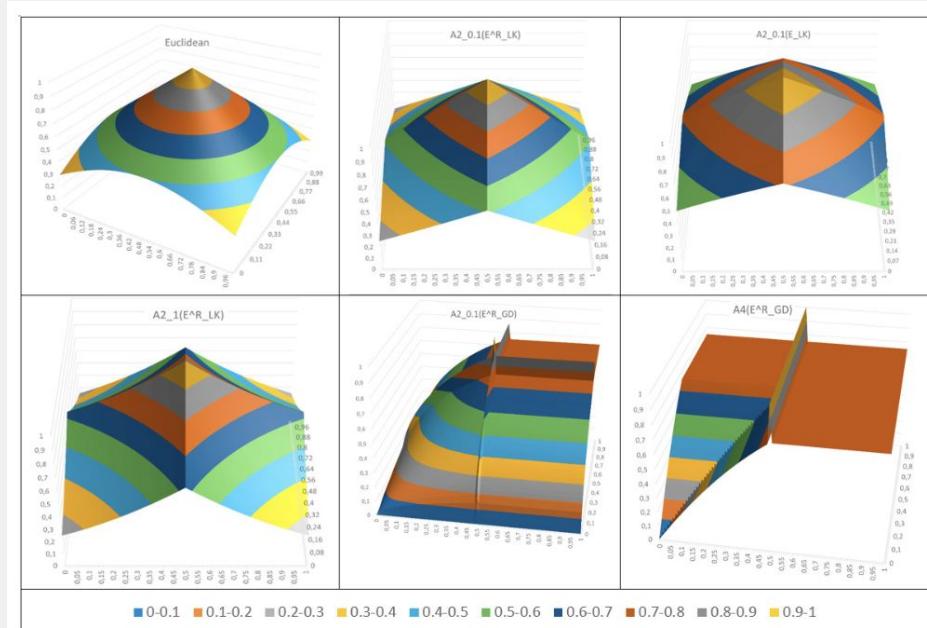
- Предложить модификацию алгоритма k-means, в которой стандартная метрика расстояния заменяется различными **нечеткими эквивалентностями** и их **агрегациями** для более точной оценки **похожести объектов**.
- Исследовать, как **композиции** агрегаций и нечетких эквивалентностей влияют на **качество** кластеризации, включая теоретический анализ их свойств.
- Оценить предлагаемый подход на **реальных и искусственных наборах данных** с помощью **адаптированного коэффициента силуэта** и **сравнить** его с классическим k-means.



Нечеткие эквивалентности

Нечеткая эквивалентность $E(x,y)$ – бинарная функция на $[0, 1]^2$, измеряющая «степень эквивалентности» двух значений и удовлетворяющая ряду аксиом:

- $E(0, 1) = 0$
- $E(x, x) = 1$ (для всех x из $[0, 1]$)
- Симметричность: $E(x, y) = E(y, x)$
- Монотонность: $E(x, y) \leq E(u, v)$ (при $x \leq u \leq v \leq y$)



Нечеткие эквивалентности

- $E_{LK}(x, y) = 1 - |x - y|,$
- $E_{GG}(x, y) = \begin{cases} 1, & \text{if } x = y \\ \frac{\min(x, y)}{\max(x, y)}, & \text{if } x \neq y \end{cases},$
- $E_{GD}(x, y) = \begin{cases} 1, & \text{if } x = y \\ \min(x, y), & \text{if } x \neq y \end{cases},$
- $E_{FD}(x, y) = \begin{cases} 1, & \text{if } x = y \\ \max(1 - y, x), & \text{if } x < y \\ \max(1 - x, y), & \text{if } x > y \end{cases}.$

+ возрастающая
биекция

- $E_{LK}^R(x, y) = (E_{LK}(x, y))^2,$
- $E_{GG}^R(x, y) = E_{GG}(x, y),$
- $E_{GD}^R(x, y) = E_{GD}(x, y) \cdot (2 - E_{GD}(x, y)),$
- $E_{FD}^R(x, y) = \begin{cases} 1 - 2 \cdot (1 - E_{FD}(x, y))^2, & \text{if } E_{FD}(x, y) \geq 0.5 \\ 2 \cdot (E_{FD}(x, y))^2, & \text{otherwise} \end{cases},$
- $E_3(x, y) = \frac{2\min(x, y)}{x+y}$ for $x + y > 0$, otherwise $E_3(x, y) = 1,$
- $E_4(x, y) = \frac{2xy}{x^2+y^2}$ for $x + y > 0$, otherwise $E_4(x, y) = 1,$
- $E_5(x, y) = \frac{2\min(x^2, y^2)}{x^2+y^2}$ for $x + y > 0$, otherwise $E_5(x, y) = 1,$
- $E_6(x, y) = \frac{1-|x-y|}{1+|x-y|},$

Агрегационные функции



Агрегационная функция $A: [0, 1]^n \rightarrow [0, 1]$ объединяет несколько значений близости (по координатам) в одно итоговое.

- *Arithmetic Mean:* $A1(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i,$
- *Power Mean:* $A2_p(x_1, \dots, x_n) = \left(\frac{1}{n} \sum_{i=1}^n (x_i)^p \right)^{\frac{1}{p}}, \text{ where } p > 0,$
- *Minimum:* $A3(x_1, \dots, x_n) = \min\{x_1, \dots, x_n\},$
- *Maximum:* $A4(x_1, \dots, x_n) = \max\{x_1, \dots, x_n\},$

+ нечеткие
эквивалентности

$$A(E)(\mathbf{x}, \mathbf{y}) = A(E(x_1, y_1), \dots, E(x_n, y_n)).$$

Модификации k-means



```
1: Input:  $D$  the input not clustered data set of  $m$  data points  $x_i$ 
2:       $k$  the number of clusters to find ( $k \leq m$ )
3:       $A, E$  the aggregation and the fuzzy equivalence used as  $A(E)$  to compute closeness of points  $x_i$  to centroids
4: Output: The clustered data points from set  $D$  into  $k$  clusters, with their centers  $\mu_1, \mu_2, \dots, \mu_k$ 
5: Initialize cluster centroids  $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$  randomly
6: repeat
7:     for each  $i$  (data point  $x_i \in \mathbb{R}^n$ ) do
8:          $c_i \leftarrow \arg \min_j \{1 - A(E)(x_i, \mu_j)\}$ 
9:     end for
10:    for each  $j$  (centroid  $\mu_j$ ) do
11:         $\mu_j \leftarrow \text{mean of } \{x_i : c_i = \mu_j\}$ 
12:    end for
13: until convergence or max number of iterations was reached
```



Эксперименты и результаты

- Эксперименты авторов (данные)

Использованные реальные (из ucimlrepo) и синтетические (make_blob из sklearn с заданным количеством объектов, кластеров и размерностью) данные для экспериментов.

ID	Name	Cnt.	Dim.	Cat.
1	Fertility	100	9	2
2	Iris	150	4	3
3	Zoo	101	16	7
4	Ecoli	336	7	8
5	Glass Identification	214	9	6
6	Leaf	340	14	30
7	LED Display Domain	500	7	10
8	Parkinson Dataset	195	22	2
9	Speaker accent	329	12	6
10	Statlog (German Credit Data)	1000	20	2
11	Banknote Authentication	1372	4	2
12	Wine Quality	6497	11	7

Table 1. The UCI Machine Learning Repository data sets used for experiments. Cnt. - Number of objects, Dim. - dimensionality, Cat. - Number of categories (k).

ID	Name	Dim.	Cat.
1	gen_1000_2_4	2	4
2	gen_1000_4_4	4	4
3	gen_1000_6_4	6	4
4	gen_1000_8_4	8	4
5	gen_1000_2_6	2	6
6	gen_1000_4_6	4	6
7	gen_1000_6_6	6	6
8	gen_1000_8_6	8	6
9	gen_1000_2_8	2	8
10	gen_1000_4_8	4	8
11	gen_1000_6_8	6	8
12	gen_1000_8_8	8	8

Table 2. Artificial data sets. Dim. - dimensionality, Cat. - Number of categories (k), Cnt. = 1000.

Эксперименты и результаты

- Эксперименты авторов
(метрика)

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Коэффициент силуэта, который измеряет насколько каждая точка похожа на свой собственный кластер по сравнению с другими кластерами (чем больше, тем лучше). Данная метрика была

адаптирована под различные конфигурации, а именно:

a(i) - среднее расстояние до точек в своём кластере

b(i) - среднее расстояние до точек в ближайшем кластере

Эксперименты и результаты

- Результаты авторов (метрика)

#	Agg (A..I)	Dist. (E)	Real datasets												Artificial datasets											
			Fer.	Iris	Zoo	Eco.	Gla	Lea	LED	Par.	Spe.	Cre.	Ban.	Win.	1	2	3	4	5	6	7	8	9	10	11	12
1	Euclidean		0.13	0.51	0.28	0.16	0.35	0.26	0.32	0.24	0.14	0.08	0.33	0.13	0.64	0.43	0.65	0.73	0.36	0.43	0.52	0.25	0.43	0.32	0.35	0.43
2	2..0..1	E_LK	0.24	0.54	0.24	0.07	0.34	0.25	0.31	0.33	0.15	0.12	0.36	0.15	0.65	0.45	0.66	0.45	0.42	0.49	0.51	0.50	0.41	0.40	0.41	0.51
3	2..0..1	E_LK^R	0.19	0.50	0.24	0.13	0.32	0.24	0.30	0.27	0.14	0.09	0.33	0.15	0.43	0.66	0.52	0.72	0.36	0.33	0.29	0.24	0.36	0.52	0.39	0.38
4	2..0..1	E_GG	0.10	0.31	0.23	0.07	0.43	0.14	0.32	0.24	0.02	0.06	0.28	0.13	0.43	0.32	0.32	0.57	0.34	0.41	0.20	0.50	0.39	0.26	0.33	0.46
5	2..0..1	E_GD	0.06	0.10	0.26	0.00	0.14	-0.02	0.30	0.00	-0.01	0.05	0.06	0.01	0.01	0.17	0.11	0.04	0.01	-0.02	0.00	0.09	0.06	-0.05	0.02	0.02
6	2..0..1	E_GD^R	0.11	0.45	0.23	0.08	0.50	0.20	0.31	0.35	0.08	0.07	0.40	0.12	0.38	0.71	0.70	0.73	0.53	0.43	0.57	0.66	0.51	0.53	0.55	0.52
7	2..0..1	E_FD	0.22	0.21	0.24	0.04	0.10	0.04	0.32	0.22	0.06	0.10	0.17	0.01	0.31	0.35	0.32	0.22	0.04	0.16	0.20	0.21	0.03	0.18	0.21	0.18
8	2..0..1	E_FD^R	0.22	0.31	0.24	0.10	0.21	0.08	0.30	0.26	0.09	0.10	0.26	0.07	0.42	0.53	0.49	0.34	0.27	0.33	0.30	0.19	0.08	0.26	0.36	0.38
9	2..0..1	E3	0.01	0.35	0.00	0.00	0.00	0.16	0.03	0.29	0.00	0.00	0.31	0.10	0.59	0.48	0.59	0.40	0.42	0.50	0.45	0.38	0.32	0.36	0.33	0.35
10	2..0..1	E4	0.02	0.46	0.00	0.00	0.01	0.21	0.03	0.37	0.08	0.00	0.43	0.15	0.41	0.73	0.73	0.75	0.54	0.62	0.36	0.68	0.53	0.57	0.51	0.58
11	2..0..1	E5	0.01	0.32	0.00	0.00	0.00	0.17	0.03	0.23	0.03	0.00	0.28	0.11	0.37	0.57	0.56	0.59	0.41	0.36	0.24	0.39	0.39	0.44	0.35	0.32
12	2..0..1	E6	0.11	0.49	0.24	0.11	0.30	0.23	0.32	0.27	0.03	0.10	0.32	0.11	0.60	0.64	0.51	0.69	0.41	0.41	0.34	0.21	0.41	0.42	0.26	0.37
13	2..1	E_LK	0.17	0.52	0.37	0.15	0.32	0.24	0.34	0.33	0.14	0.08	0.36	0.12	0.64	0.68	0.55	0.43	0.43	0.43	0.34	0.39	0.43	0.46	0.41	0.54
14	2..1	E_LK^R	0.15	0.49	0.37	0.16	0.28	0.22	0.37	0.32	0.03	0.06	0.32	0.09	0.60	0.39	0.50	0.68	0.42	0.56	0.33	0.44	0.37	0.50	0.48	0.44
15	2..1	E_GG	0.16	0.37	0.36	0.14	0.27	0.15	0.37	0.25	0.05	0.07	0.23	0.11	0.29	0.40	0.52	0.57	0.38	0.25	0.29	0.30	0.33	0.30	0.24	0.02
16	2..1	E_GD	0.16	0.14	0.36	-0.01	0.08	-0.03	0.32	0.01	-0.01	0.05	0.06	0.00	0.04	-0.03	0.11	0.11	0.05	0.04	0.01	0.11	-0.02	0.00	-0.02	0.04
17	2..1	E_GD^R	0.20	0.51	0.36	0.20	0.45	0.24	0.36	0.38	0.05	0.09	0.37	0.16	0.71	0.59	0.46	0.34	0.58	0.54	0.58	0.49	0.50	0.43	0.58	0.39
18	2..1	E_FD	0.16	0.23	0.37	0.05	0.13	0.05	0.37	0.18	0.05	0.07	0.16	0.01	0.25	0.33	0.28	0.18	0.14	0.20	0.12	0.21	0.04	0.08	0.10	0.19
19	2..1	E_FD^R	0.16	0.31	0.37	0.10	0.14	0.05	0.38	0.33	0.07	0.08	0.25	0.04	0.19	0.51	0.47	0.26	0.24	0.33	0.33	0.36	-0.03	0.19	0.23	0.31
20	2..1	E3	0.06	0.40	0.02	0.05	0.03	0.17	0.05	0.32	0.04	0.03	0.29	0.10	0.59	0.38	0.39	0.63	0.36	0.45	0.36	0.36	0.35	0.36	0.37	0.38
21	2..1	E4	0.06	0.54	0.03	0.04	0.06	0.25	-0.02	0.40	0.05	0.05	0.40	0.21	0.73	0.57	0.73	0.50	0.45	0.46	0.60	0.33	0.50	0.57	0.60	0.40
22	2..1	E5	0.08	0.37	0.02	0.07	0.06	0.14	-0.12	0.24	0.04	0.03	0.23	0.08	0.56	0.57	0.53	0.59	0.40	0.39	0.24	0.26	0.39	0.30	0.18	0.33
23	2..1	E6	0.15	0.48	0.37	0.16	0.28	0.21	0.42	0.30	0.03	0.06	0.31	0.15	0.58	0.39	0.49	0.66	0.42	0.37	0.43	0.52	0.36	0.40	0.47	0.36
24	2..4	E_LK	0.16	0.51	0.38	0.15	0.23	0.21	0.39	0.35	0.04	0.07	0.29	0.11	0.42	0.64	0.59	0.40	0.43	0.38	0.18	0.25	0.41	0.33	0.46	0.33
25	2..4	E_LK^R	0.18	0.47	0.37	0.12	0.17	0.15	0.36	0.34	0.03	0.08	0.25	0.10	0.39	0.31	0.51	0.57	0.33	0.43	0.29	0.32	0.38	0.23	0.25	0.38
26	2..4	E_GG	0.16	0.34	0.37	0.11	0.18	0.07	0.33	0.20	0.02	0.06	0.17	0.07	0.46	0.28	0.45	0.51	0.29	0.37	0.34	0.29	0.26	0.18	0.19	0.26
27	2..4	E_GD	0.18	0.24	0.37	0.01	0.06	-0.03	0.34	0.09	0.03	0.08	0.13	0.02	0.04	0.09	0.10	0.16	0.12	-0.08	0.10	0.08	-0.06	-0.07	0.03	-0.02
28	2..4	E_GD^R	0.19	0.51	0.37	0.24	0.16	0.33	0.34	0.05	0.08	0.33	0.13	0.54	0.52	0.66	0.72	0.52	0.50	0.39	0.44	0.38	0.41	0.47	0.53	

Эксперименты и результаты

- Результаты авторов (метрика)

#	Agg (A_i)	Dist. (E)	Real datasets										Artificial datasets													
			Fer.	Iris	Zoo	Eco	Gla	Lea	LED	Par.	Spe.	Cre.	Ban.	Win.	1	2	3	4	5	6	7	8	9	10	11	12
29	2_4	E_FD	0.17	0.26	0.37	0.02	0.06	0.02	0.32	-0.47	0.05	0.08	0.22	0.01	0.26	0.26	0.23	0.20	0.03	0.16	0.13	0.16	0.07	0.10	0.07	0.11
30	2_4	E_FD^R	0.17	0.34	0.37	0.05	0.14	0.03	0.36	0.42	0.08	0.08	0.32	0.01	0.42	0.41	0.29	0.45	0.25	0.27	0.28	0.23	0.14	0.16	0.11	0.15
31	2_4	E3	0.08	0.41	0.03	0.11	0.05	0.11	-0.05	0.23	0.03	0.05	0.24	0.10	0.54	0.56	0.43	0.59	0.41	0.27	0.30	0.30	0.19	0.25	0.40	0.29
32	2_4	E4	0.14	0.53	0.03	0.09	0.04	0.20	0.01	0.37	0.06	0.05	0.35	0.14	0.70	0.74	0.70	0.75	0.58	0.61	0.49	0.47	0.49	0.46	0.51	0.44
33	2_4	E5	0.09	0.35	0.03	0.10	0.06	0.07	-0.06	0.07	0.02	0.07	0.18	0.09	0.19	0.47	0.34	0.11	0.28	0.24	0.27	0.28	0.20	0.23	0.16	0.35
34	2_4	E6	0.18	0.47	0.37	0.12	0.14	0.15	0.33	0.34	0.03	0.08	0.24	0.10	0.39	0.54	0.42	0.30	0.30	0.34	0.13	0.32	0.29	0.28	0.32	0.33
35	3	E_LK	0.10	0.46	0.19	0.14	0.31	0.20	0.22	0.16	0.07	-0.01	0.27	0.12	0.64	0.33	0.49	0.71	0.39	0.58	0.48	0.48	0.35	0.30	0.41	0.46
36	3	E_LK^R	0.05	0.40	0.18	0.14	0.30	0.19	0.21	0.12	0.07	0.00	0.23	0.10	0.36	0.59	0.58	0.62	0.34	0.43	0.45	0.32	0.43	0.41	0.36	0.28
37	3	E_GG	0.04	0.23	0.18	0.13	0.39	0.10	0.25	0.07	0.03	0.00	0.26	0.11	0.35	0.43	0.40	0.38	0.25	0.24	0.26	0.26	0.37	0.27	0.26	0.21
38	3	E_GD	0.02	0.01	0.17	0.00	0.06	0.00	0.21	0.00	-0.03	0.00	0.12	-0.01	0.07	0.01	0.00	-0.04	-0.05	0.00	0.01	-0.03	-0.02	-0.02	0.02	-0.02
39	3	E_GD^R	0.07	0.25	0.19	0.20	0.52	0.15	0.24	0.13	0.06	0.00	0.40	0.13	0.44	0.56	0.45	0.32	0.57	0.17	0.44	0.40	0.51	0.45	0.43	0.37
40	3	E_FD	0.09	0.17	0.17	0.05	0.10	0.04	0.22	0.08	0.02	0.06	0.13	0.04	0.35	0.17	0.25	0.22	0.18	0.14	0.17	0.16	0.12	0.18	0.17	0.19
41	3	E_FD^R	0.08	0.25	0.18	0.09	0.17	0.07	0.21	0.09	0.03	0.04	0.19	0.10	0.46	0.48	0.44	0.25	0.28	0.28	0.22	0.22	0.13	0.30	0.28	0.19
42	3	E3	0.00	0.28	0.00	0.00	0.02	0.14	0.07	0.11	0.03	0.00	0.33	0.16	0.56	0.50	0.49	0.39	0.40	0.41	0.31	0.36	0.32	0.19	0.34	0.32
43	3	E4	0.00	0.26	0.00	0.00	0.02	0.16	0.07	0.14	0.06	0.00	0.44	0.12	0.65	0.57	0.28	0.60	0.58	0.53	0.30	0.54	0.49	0.42	0.38	0.37
44	3	E5	0.00	0.23	0.00	0.00	0.01	0.10	0.07	0.05	0.04	0.00	0.24	0.05	0.53	0.38	0.39	0.22	0.27	0.24	0.27	0.23	0.38	0.28	0.23	0.18
45	3	E6	0.07	0.39	0.18	0.13	0.29	0.18	0.21	0.12	0.06	0.00	0.22	0.12	0.33	0.36	0.56	0.60	0.40	0.36	0.43	0.41	0.34	0.23	0.36	0.45
46	4	E_LK	-0.32	0.38	-	0.90	0.46	0.29	-	0.39	-0.11	-	0.05	-0.03	0.19	0.13	0.16	0.05	0.07	0.01	0.04	-0.04	0.17	-0.03	-0.06	-0.10
47	4	E_LK^R	-0.32	0.38	-	0.90	0.47	0.31	-	0.38	-0.10	-	0.05	-0.04	0.08	0.15	0.16	0.01	0.05	0.00	-0.02	-0.06	0.03	-0.01	-0.08	-0.08
48	4	E_GG	-0.32	0.31	-	-	0.47	0.14	-	0.25	-0.10	-	0.15	-0.01	0.36	0.26	0.14	0.08	0.12	0.04	0.01	-0.04	0.04	-0.05	0.00	-0.05
49	4	E_GD	-0.32	0.34	-	0.90	0.42	-0.09	-	0.26	0.03	-	-0.12	0.03	0.01	-0.02	0.12	-0.06	0.21	0.06	0.11	0.07	0.01	-0.10	-0.07	0.10
50	4	E_GD^R	-0.32	0.30	-	0.90	0.28	-0.01	-	0.15	-0.24	-	0.14	-0.07	0.41	0.44	-0.12	0.18	0.20	0.01	-0.07	-0.03	-0.08	-0.07	-0.12	-0.17
51	4	E_FD	-0.32	0.22	-	-	0.53	0.10	-	0.51	0.05	-	0.14	-0.04	0.33	0.21	0.13	0.18	0.05	-0.07	0.17	0.15	-0.08	-0.05	-0.03	0.15
52	4	E_FD^R	-0.32	0.34	-	0.90	0.15	-0.13	-	0.73	0.06	-	0.23	-0.16	0.07	0.05	0.26	0.29	-0.21	-0.05	0.20	0.19	-0.04	0.14	0.12	0.26
53	4	E3	-0.59	0.31	-	-0.06	-0.06	0.18	-0.46	0.25	-0.10	-0.35	0.15	-0.02	0.44	0.07	0.24	-0.10	0.16	0.03	-0.04	-0.05	0.09	-0.05	-0.04	-0.07
54	4	E4	-0.59	0.29	-	-0.22	-0.21	-0.01	-0.46	0.15	-0.26	-0.36	0.13	-0.07	0.11	0.47	0.24	-0.09	0.11	-0.03	0.00	-0.10	0.12	-0.17	-0.14	-0.18
55	4	E5	-0.58	0.31	-	-0.06	-0.05	0.18	-0.46	0.25	-0.10	-0.35	0.15	-0.01	0.14	0.07	0.10	0.01	0.22	0.04	-0.05	-0.07	0.10	-0.11	-0.05	-0.09
56	4	E6	-0.32	0.38	-	0.90	0.47	0.37	-	0.38	-0.10	-	0.06	-0.05	0.24	0.13	0.09	0.00	0.14	0.07	-0.02	-0.02	0.07	-0.02	-0.03	-0.08



Эксперименты и результаты



- **Результаты авторов (выводы)**
- По частоте попадания в топ-3 наиболее «успешными» оказались:
 - Агрегации:** $A_{2(1)}$ – 19 раз; $A_{2(0.1)}$ – 18 раз; $A_{2(4)}$ – 11 раз; A_4 – 10 раз; A_3 – 3 раза.
 - Нечеткие эквивалентности:** ELK – 12 раз; $ELK(R)$ – 7 раз; EFD – 6 раз; $EFD(R)$ – 6 раз; EGD – 3 раза; $EGD(R)$ – 7 раз; EGG – 1 раз; $E3$ – 1 раз; $E4$ – 7 раз; $E6$ – 7 раз.
 - Пары:** $A_{2(0.1)}(ELK)$ - 5 раз; $A_{2(1)}(EGD(R))$ - 4 раза.
- Для многих наборов данных композиции на основе ELK (особенно с $A_{2(0.1)}$) давали силуэты, сопоставимые или лучше, чем классический k-means с евклидовой метрикой.
- Агрегация – не техническая деталь, а полноценный «гиперпараметр» алгоритма, так как её смена приводит к заметным изменениям силуэта

В финальных выводах авторы подводят, что несмотря на более высокую гибкость в комбинировании агрегаций с нечеткими эквивалентностями, их производительность (как и производительность стандартного k-means) далека от идеала. Следовательно, в будущих работах следует исследовать новые классы агрегаций и эквивалентностей.

Эксперименты и результаты

- **Эксперименты презентующего (данные)**

Для экспериментов также были взяты как реальные, так и синтетические данные. Список данных:

- **Iris (UCIML)** - использован для приблизительной сверки результатов авторов и презентующего
- **Wine (UCIML)** - использован для приблизительной сверки результатов авторов и презентующего
- **Absenteeism at work (UCIML)**
- **Heart Failure Clinical Records (UCIML)**
- **Wholesale customers (UCIML)**
- **Breast Cancer (sklearn)**
- **make_blob (sklearn)** - использован для приблизительной сверки результатов авторов и презентующего
- **make_circles (sklearn)**
- **make_moons (sklearn)**

Эксперименты и результаты



- Эксперименты презентующего (метрики)

В дополнение к коэффициенту силуэта, также были добавлены:

Индекс Дэвиса-Болдина

$$DB = \frac{1}{N} \sum_{i=1}^N D_i$$

$$D_i = \max_{j: i \neq j} R_{i,j}$$

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}}$$

Среднее сходство между каждым кластером и наиболее похожим на него кластером (чем меньше, тем лучше).

Индекс Калински-Харабаша

k = Number of clusters

n_q = Number of points in cluster q

c_q = Cluster center of cluster q

n_E = Number of data points

c_E = Cluster center of all points

Between-cluster dispersion, $B = \sum_{q \in k} n_q (c_q - c_E)(c_q - c_E)^T$

Within-cluster dispersion, $W = \sum_{q \in k} \sum_{x \in \text{cluster } q} (x - c_q)(x - c_q)^T$

Calinski-Harabasz score (s) = $\frac{B}{W} \times \frac{n_E - k}{k - 1}$

Отношение дисперсии между кластерами к дисперсии внутри кластера.

Эксперименты и результаты

- Результаты презентующего (метрики)

		Iris				Wine				make_blobs						
combination	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	
0	Standard_KMeans	nan	0.459948	0.833595	241.904402	3	nan	0.284859	1.389188	70.940008	3	nan	0.797474	0.281107	3290.737331	4
1	E_LK_A2_p0.1	0.100000	0.373546	0.653371	142.309905	3	0.100000	0.214728	0.726906	105.286977	3	0.100000	0.574044	0.497811	361.688337	4
2	E_LK_A2_p1.0	0.100000	0.384307	0.674844	145.510686	3	1.000000	0.214630	1.066157	105.373902	3	1.000000	0.559484	0.538308	341.019579	4
3	E_LK_A2_p4.0	0.400000	0.377721	0.661518	158.450555	3	4.000000	0.000000	nan	0.000000	0	4.000000	0.462007	0.714803	286.969372	4
4	E_LK_A3	nan	0.210948	0.988354	104.138667	3	nan	0.136121	1.259468	87.046574	3	nan	0.545042	0.500736	373.312483	4
5	E_LK_A4	nan	0.478973	4.899230	278.355721	3	nan	0.053912	0.710012	237.575934	3	nan	0.041398	4.300635	135.500985	4
6	E_GG_A2_p0.1	0.100000	0.175215	1.498438	69.082637	3	0.100000	0.000000	nan	0.000000	1	0.100000	0.117283	0.935915	81.190512	4
7	E_GG_A2_p1.0	0.100000	-0.131620	1.377588	77.033123	3	1.000000	0.166671	1.681113	80.345416	3	1.000000	0.146022	1.161028	112.688811	4
8	E_GG_A2_p4.0	0.400000	-0.072385	2.248107	86.147323	3	4.000000	0.094006	1.777326	82.408241	3	4.000000	0.071419	2.099753	158.648776	4
9	E_GG_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.121217	1.226949	99.708338	4
10	E_GG_A4	nan	-0.080237	1.664873	106.699681	3	nan	0.436293	0.225192	167.058096	3	nan	0.090420	1.288073	174.751540	4
11	E_GD_A2_p0.1	0.100000	0.248460	1.543068	68.989685	3	0.100000	0.000000	nan	0.000000	1	0.100000	0.057068	1.504782	95.137015	4
12	E_GD_A2_p1.0	0.100000	0.124218	1.735794	76.908055	3	1.000000	0.108835	1.962911	89.120952	3	1.000000	-0.166681	1.508702	105.127308	4
13	E_GD_A2_p4.0	0.400000	-0.002365	1.562838	96.593873	3	4.000000	0.094877	2.144873	111.258490	3	4.000000	-0.258656	1.442202	130.879671	4
14	E_GD_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	-0.035978	1.668137	92.633651	4
15	E_GD_A4	nan	0.169421	1.663060	128.240242	3	nan	0.228507	1.550683	194.378262	3	nan	-0.002404	1.962648	137.918840	4
16	E_FD_A2_p0.1	0.100000	0.221265	1.306645	114.882862	3	0.100000	0.149149	0.848762	109.160684	3	0.100000	0.294672	1.467564	186.693041	4
17	E_FD_A2_p1.0	0.100000	0.236725	1.365479	120.207855	3	1.000000	0.137704	1.384085	115.160590	3	1.000000	0.307145	1.508572	194.496889	4
18	E_FD_A2_p4.0	0.400000	0.245711	1.441635	135.775888	3	4.000000	0.131011	1.911785	131.041382	3	4.000000	0.224894	2.361763	224.482342	4
19	E_FD_A3	nan	0.174683	1.382418	87.551806	3	nan	0.105044	1.391873	80.492945	3	nan	0.211569	1.632435	145.531673	4
20	E_FD_A4	nan	0.396244	1.094753	216.809095	3	nan	0.342259	22.728931	328.940505	3	nan	0.475852	0.768449	581.501799	3
21	E_LK_R_A2_p0.1	0.100000	0.313480	0.839560	126.951822	3	0.100000	0.166173	1.138205	100.337610	3	0.100000	0.521004	0.593911	310.772594	4
22	E_LK_R_A2_p1.0	0.100000	0.337299	0.809013	133.457336	3	1.000000	0.180213	1.340048	103.670664	3	1.000000	0.489632	0.684193	281.474055	4
23	E_LK_R_A2_p4.0	0.400000	0.315607	0.697831	145.840462	3	4.000000	0.180164	1.436306	106.233454	3	4.000000	0.059369	1.521383	126.316534	4
24	E_LK_R_A3	nan	0.146792	1.268308	95.858191	3	nan	0.064807	1.704558	87.989393	3	nan	0.474685	0.642836	300.075943	4
25	E_LK_R_A4	nan	0.452685	4.684038	244.769984	3	nan	0.055392	0.728353	231.757290	3	nan	0.031379	3.939551	129.227702	4
26	E_GG_R_A2_p0.1	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0
27	E_GG_R_A2_p1.0	0.100000	0.000000	nan	0.000000	0	1.000000	0.000000	nan	0.000000	0	1.000000	0.000000	nan	0.000000	0
28	E_GG_R_A2_p4.0	0.400000	0.000000	nan	0.000000	0	4.000000	0.000000	nan	0.000000	0	4.000000	0.000000	nan	0.000000	0
29	E_GG_R_A3	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0
30	E_GG_R_A4	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0

Эксперименты и результаты

- Результаты презентующего (метрики)

		Iris				Wine				make_blobs						
combination	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	
31	E_GD_R_A2_p0.1	0.100000	0.394891	1.354484	55.060252	3	0.100000	0.000000	nan	0.000000	1	0.100000	0.282175	1.371481	73.651417	4
32	E_GD_R_A2_p1.0	1.000000	0.275633	1.466379	85.834548	3	1.000000	0.295263	1.536720	78.007532	3	1.000000	0.037046	1.576736	81.505552	4
33	E_GD_R_A2_p4.0	4.000000	0.131347	1.742097	95.430139	3	4.000000	0.132146	1.963156	111.446706	3	4.000000	-0.285388	1.519307	107.857681	4
34	E_GD_R_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.099591	1.575197	81.301760	4
35	E_GD_R_A4	nan	0.248884	1.962559	167.875283	3	nan	0.310820	1.657179	317.428293	3	nan	-0.118053	4.169188	105.833895	4
36	E_FD_R_A2_p0.1	0.100000	0.285587	0.589173	140.03939	3	0.100000	0.208625	0.820720	115.853163	3	0.100000	0.473544	1.302852	267.817142	4
37	E_FD_R_A2_p1.0	1.000000	0.285294	1.171708	152.001146	3	1.000000	0.196261	1.345676	128.779119	3	1.000000	0.441212	1.370622	281.072278	4
38	E_FD_R_A2_p4.0	4.000000	0.3232113	1.427932	184.466986	3	4.000000	0.180443	1.7979630	155.230637	3	4.000000	0.353318	1.727786	352.817706	4
39	E_FD_R_A3	nan	0.259049	1.221122	100.334369	3	nan	0.093983	1.598669	79.679906	3	nan	0.285117	1.854488	190.090121	4
40	E_FD_R_A4	nan	0.520982	0.854473	398.426402	3	nan	0.458652	592.766818	791.683024	3	nan	0.720604	0.351654	195.831600	3
41	E_A_3_p0.1	0.100000	0.271465	1.361299	60.982607	3	0.100000	0.000000	nan	0.000000	1	0.100000	0.132047	0.847677	69.104344	4
42	E_A_3_p1.0	1.000000	0.219282	2.109128	70.575905	3	1.000000	0.298711	1.404910	76.139490	3	1.000000	0.177070	1.054457	84.808222	4
43	E_A_3_p4.0	4.000000	0.018243	1.454699	88.769956	3	4.000000	0.172355	1.3991170	92.413042	3	4.000000	0.175763	2.048011	155.692841	4
44	E_A_3_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.192727	1.164469	87.563170	4
45	E_A_3_A4	nan	-0.043018	1.816295	108.883861	3	nan	0.453032	0.182310	184.644867	3	nan	0.091883	1.188629	188.001731	4
46	E_A_4_A2_p0.1	0.100000	0.340675	1.270142	49.694032	3	0.100000	0.000000	nan	0.000000	1	0.100000	0.145681	0.714442	40.365925	4
47	E_A_4_A2_p1.0	1.000000	0.280531	1.726922	60.318877	3	1.000000	0.000000	nan	0.000000	0	1.000000	0.188200	0.903908	51.187446	4
48	E_A_4_A2_p4.0	4.000000	0.307285	1.829554	53.325657	3	4.000000	0.028422	1.473380	112.433651	3	4.000000	0.328177	1.100606	137.595631	4
49	E_A_4_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.236687	1.153269	67.269798	4
50	E_A_4_A4	nan	0.095875	3.262939	121.076072	3	nan	0.560912	0.045644	495.901921	3	nan	0.122580	1.089752	228.103708	4
51	E_A_5_A2_p0.1	0.100000	0.109129	1.601266	71.437202	3	0.100000	0.000000	nan	0.000000	1	0.100000	0.089296	0.938223	78.041105	4
52	E_A_5_A2_p1.0	1.000000	0.209773	1.568653	69.516876	3	1.000000	0.121500	1.816155	83.297820	3	1.000000	0.205671	1.310777	122.072479	4
53	E_A_5_A2_p4.0	4.000000	0.133355	1.368065	100.855089	3	4.000000	0.020514	2.424387	111.208430	3	4.000000	0.058388	1.110928	165.998302	4
54	E_A_5_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.067281	1.247887	103.806478	4
55	E_A_5_A4	nan	-0.083128	1.676958	109.832330	3	nan	0.448983	0.199951	177.648644	3	nan	0.098265	1.330899	179.149925	4
56	E_A_6_A2_p0.1	0.100000	0.307695	0.826098	124.134585	3	0.100000	0.177773	1.015677	99.138045	3	0.100000	0.498372	0.637688	286.576502	4
57	E_A_6_A2_p1.0	1.000000	0.330222	0.824242	130.478284	3	1.000000	0.178177	1.296189	102.714971	3	1.000000	0.474815	0.701020	265.961866	4
58	E_A_6_A2_p4.0	4.000000	0.306634	0.731357	143.190707	3	4.000000	0.000000	nan	0.000000	0	4.000000	0.292645	1.352098	214.683614	4
59	E_A_6_A3	nan	0.147474	1.233935	94.989164	3	nan	0.083568	1.5355543	87.643018	3	nan	0.456770	0.671244	283.726969	4
60	E_A_6_A4	nan	0.438738	4.509612	232.890994	3	nan	0.056061	0.742115	227.543376	3	nan	0.025058	3.765959	126.960433	4

Эксперименты и результаты

- Результаты презентующего (метрики)

Absenteeism										Heart Failure					Wholesale Customers				
combination	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found				
488	Standard_KMeans	nan	0.167139	2.008950	25.211153	4	nan	0.117851	2.622139	35.894672	2	nan	0.548287	0.927935	140.135183	3			
489	E_LK_A2_p0.1	0.100000	0.248792	0.976703	49.409376	4	0.100000	0.085943	1.656504	198.516396	2	0.100000	0.852194	0.569734	249.978102	3			
490	E_LK_A2_p1.0	0.000000	0.145521	1.222746	78.066799	4	1.000000	0.243780	0.829758	330.699745	2	1.000000	0.818414	0.628292	209.036014	3			
491	E_LK_A2_p4.0	4.000000	0.125581	1.311828	101.211745	4	4.000000	0.247623	0.395615	556.267640	2	4.000000	0.681749	0.626734	220.096487	3			
492	E_LK_A3	nan	0.028659	1.701360	60.159921	4	nan	-0.082940	1.798270	276.595754	2	nan	0.758311	1.164741	155.406410	3			
493	E_LK_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.403862	0.251920	494.237191	3			
494	E_GG_A2_p0.1	0.100000	0.027963	1.932708	52.649041	4	0.100000	0.003024	1.992273	298.245566	2	0.100000	0.000000	nan	0.000000	0			
495	E_GG_A2_p1.0	1.000000	0.084217	1.636474	77.481294	4	1.000000	0.133564	1.207878	397.699783	2	1.000000	0.046453	1.655294	206.006767	3			
496	E_GG_A2_p4.0	4.000000	-0.024804	1.498048	99.733695	4	4.000000	0.216410	0.499127	744.731991	2	4.000000	0.125805	1.509075	186.940866	3			
497	E_GG_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	0			
498	E_GG_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.285407	0.633294	334.499176	3			
499	E_GD_A2_p0.1	0.100000	0.016544	1.947848	58.663661	4	0.100000	0.003024	1.992273	298.245566	2	0.100000	0.000000	nan	0.000000	0			
500	E_GD_A2_p1.0	1.000000	0.049989	1.776249	81.389040	4	1.000000	0.133564	1.207878	397.699783	2	1.000000	0.006246	2.464582	222.486314	3			
501	E_GD_A2_p4.0	4.000000	0.104837	1.645369	125.531171	4	4.000000	0.216410	0.499127	744.731991	2	4.000000	0.003055	6.455757	233.083164	3			
502	E_GD_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	0			
503	E_GD_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	-0.088336	5.368799	244.311602	3			
504	E_FD_A2_p0.1	0.100000	0.141886	1.371610	60.123537	4	0.100000	0.085943	1.656504	198.516396	2	0.100000	0.851298	0.667035	248.285653	3			
505	E_FD_A2_p1.0	1.000000	0.121806	1.368953	82.994408	4	1.000000	0.243780	0.829758	330.699745	2	1.000000	0.819587	0.752518	224.382281	3			
506	E_FD_A2_p4.0	4.000000	0.116781	1.296427	100.658413	4	4.000000	0.247623	0.395615	556.267640	2	4.000000	0.721148	0.951887	263.784627	3			
507	E_FD_A3	nan	-0.039323	1.854473	62.846661	4	nan	-0.082940	1.798270	276.595754	2	nan	0.768129	1.225525	155.872325	3			
508	E_FD_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.468474	0.277673	1030.259889	3			
509	E_LK_R_A2_p0.1	0.100000	0.132705	1.352127	58.786548	4	0.100000	0.035966	1.840834	255.545874	2	0.100000	0.796993	0.912710	242.055236	3			
510	E_LK_R_A2_p1.0	1.000000	0.121774	1.382424	78.671737	4	1.000000	0.210592	1.012225	395.180471	2	1.000000	0.718076	0.779210	252.095290	3			
511	E_LK_R_A2_p4.0	4.000000	0.090862	1.207203	94.205949	4	4.000000	0.174537	0.453937	688.670389	2	4.000000	0.597763	0.914824	297.214409	3			
512	E_LK_R_A3	nan	0.011245	1.922160	64.471455	4	nan	-0.019829	1.953980	295.183514	2	nan	0.652319	1.521006	160.211717	3			
513	E_LK_R_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.402757	0.256000	326.170702	3			
514	E_GG_R_A2_p0.1	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
515	E_GG_R_A2_p1.0	1.000000	0.000000	nan	0.000000	0	1.000000	0.000000	nan	0.000000	0	1.000000	0.000000	nan	0.000000	0			
516	E_GG_R_A2_p4.0	4.000000	0.000000	nan	0.000000	0	4.000000	0.000000	nan	0.000000	0	4.000000	0.000000	nan	0.000000	0			
517	E_GG_R_A3	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0			
518	E_GG_R_A4	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0			

Эксперименты и результаты

- Результаты презентующего (метрики)

Absenteeism										Heart Failure										Wholesale Customers									
combination	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found				
519	E_GD_R_A2_p0.1	0.100000	0.029581	1.930475	53.789611	4	0.100000	0.014547	1.980699	301.713253	2	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
520	E_GD_R_A2_p1.0	1.000000	0.081621	1.637193	74.137906	4	1.000000	0.140799	1.106786	382.209917	2	1.000000	0.016276	2.992422	221.515826	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
521	E_GD_R_A2_p4.0	0.000000	0.082896	1.718742	107.073532	4	0.400000	0.221485	0.451628	583.017434	2	4.000000	0.006019	7.510885	239.238766	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
522	E_GD_R_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.147157	0.000000	nan	0.000000	0	2	nan	-0.141467	19.233864	258.504442	3	0.100000	0.000000	nan	0.000000	0	
523	E_GD_R_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.147157	0.000000	nan	0.000000	0	2	nan	0.147157	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0
524	E_FD_R_A2_p0.1	0.100000	0.158455	1.159163	60.070831	4	0.100000	0.049644	1.782958	208.771517	2	0.100000	0.941768	0.711167	205.611433	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
525	E_FD_R_A2_p1.0	0.000000	0.174101	1.321177	84.689173	4	1.000000	0.246268	0.799172	357.397423	2	1.000000	0.931173	0.602115	149.834911	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
526	E_FD_R_A2_p4.0	0.000000	0.121822	1.394769	110.870805	4	0.400000	0.300680	0.345218	525.055904	2	4.000000	0.889655	0.980002	141.189039	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
527	E_FD_R_A3	nan	-0.023222	1.920995	64.592166	4	nan	-0.039669	1.907959	293.359276	2	nan	0.881526	1.298758	101.909068	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
528	E_FD_R_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.554222	0.138164	613.050513	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
529	E_3_A2_p0.1	0.100000	0.037482	1.920564	48.592179	4	0.100000	0.013849	1.981432	301.511278	2	0.100000	0.000000	0.242160	1.383809	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
530	E_3_A2_p1.0	0.000000	0.076762	1.771173	73.513626	4	1.000000	0.137816	1.132803	388.708458	2	1.000000	0.061693	1.592960	249.441426	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
531	E_3_A2_p4.0	0.000000	0.091146	1.6777940	103.455689	4	0.400000	0.216199	0.473975	632.293732	2	4.000000	0.1359854	0.506818	328.892145	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
532	E_3_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.147157	0.000000	nan	0.000000	0	2	nan	0.147157	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0
533	E_3_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.147157	0.000000	nan	0.000000	0	2	nan	0.147157	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0
534	E_4_A2_p0.1	0.100000	0.041272	1.908767	45.741066	4	0.100000	0.014951	1.980257	301.830732	2	0.100000	0.000000	0.242160	1.383809	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
535	E_4_A2_p1.0	0.000000	0.145774	1.411075	63.944725	4	1.000000	0.154955	1.102461	362.111347	2	1.000000	0.158589	1.586649	140.219535	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
536	E_4_A2_p4.0	0.000000	0.100532	1.602743	92.008898	4	4.000000	0.217565	0.440097	548.075442	2	4.000000	0.127679	1.324164	207.824579	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
537	E_4_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.147157	0.000000	nan	0.000000	0	2	nan	0.147157	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0
538	E_4_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.458791	0.346541	317.354517	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
539	E_5_A2_p0.1	0.100000	-0.001106	1.939180	56.255200	4	0.100000	0.002287	1.994347	297.947206	2	0.100000	0.000000	0.51631	1.650984	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
540	E_5_A2_p1.0	0.000000	0.077773	1.668956	78.310750	4	1.000000	0.129800	1.253162	408.527470	2	1.000000	0.096213	1.246435	153.526406	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
541	E_5_A2_p4.0	0.000000	-0.010629	1.575010	92.154439	4	4.000000	0.217565	0.505169	816.711878	2	4.000000	0.1696942	0.447168	669.080545	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
542	E_5_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.147157	0.000000	nan	0.000000	0	2	nan	0.147157	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0
543	E_5_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.267483	0.697116	342.642446	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
544	E_6_A2_p0.1	0.100000	0.137885	1.354360	58.992800	4	0.100000	0.068843	1.722071	240.219579	2	0.100000	0.786898	0.769982	249.317727	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
545	E_6_A2_p1.0	0.000000	0.112443	1.342935	84.180487	4	1.000000	0.213981	0.956159	377.166833	2	1.000000	0.718580	0.735778	257.603396	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
546	E_6_A2_p4.0	0.000000	0.013779	1.290661	92.538435	4	4.000000	0.1696942	0.447168	669.080545	2	4.000000	0.599113	0.911333	295.747140	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
547	E_6_A3	nan	0.016828	1.827726	62.650058	4	nan	-0.047149	1.885827	286.364788	2	nan	0.658204	1.395916	165.097471	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			
548	E_6_A4	nan	0.420000	0.000000	nan	2	nan	0.147157	0.000000	nan	2	nan	0.401695	0.258708	318.073105	3	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0			

Эксперименты и результаты

- Результаты презентующего (метрики)

Breast Cancer										make_circles				make_moons			
combination	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found		
122	Standard_KMeans	nan	0.343382	1.320510	267.691716	2	nan	0.353230	1.175177	174.943045	2	nan	0.490996	0.806426	409.607138	2	
123	E_LK_A2_p0.1	0.100000	0.616886	0.825071	561.099293	2	0.100000	0.301050	0.738522	403.577864	2	0.100000	0.229967	0.842450	353.601788	2	
124	E_LK_A2_p1.0	1.000000	0.594460	0.677435	580.393481	2	1.000000	0.298045	0.686538	434.050381	2	1.000000	0.225935	0.766275	388.800011	2	
125	E_LK_A2_p4.0	4.000000	0.578695	0.515718	648.792219	2	4.000000	0.320048	0.562335	529.551135	2	4.000000	0.252175	0.590430	504.735676	2	
126	E_LK_A3	nan	0.017390	1.840279	568.008804	2	nan	0.233692	0.997396	305.382579	2	nan	0.165489	1.156357	259.670591	2	
127	E_LK_A4	nan	0.347910	0.072083	1167.207078	2	nan	0.424196	0.375680	775.711314	2	nan	0.366976	0.376193	786.385542	2	
128	E_GG_A2_p0.1	0.100000	-0.008665	1.969116	566.870645	2	0.100000	0.002437	1.989573	299.411671	2	0.100000	0.433808	1.560562	271.017038	2	
129	E_GG_A2_p1.0	1.000000	0.042423	1.539810	573.381068	2	1.000000	0.105798	1.325537	334.729475	2	1.000000	0.321489	1.008346	300.954522	2	
130	E_GG_A2_p4.0	4.000000	0.075499	1.357647	638.149273	2	4.000000	0.228034	0.870745	390.313131	2	4.000000	0.298343	0.567256	376.284844	2	
131	E_GG_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.293779	1.703253	218.647293	2	
132	E_GG_A4	nan	0.152195	0.842421	811.444201	2	nan	0.316736	0.657233	443.384596	2	nan	0.377304	0.313439	487.386843	2	
133	E_GD_A2_p0.1	0.100000	-0.008665	1.969116	566.870645	2	0.100000	0.002437	1.989573	299.411671	2	0.100000	0.433808	1.560562	271.017038	2	
134	E_GD_A2_p1.0	1.000000	0.042423	1.539810	573.381068	2	1.000000	0.105798	1.325537	334.729475	2	1.000000	0.321489	1.008346	300.954522	2	
135	E_GD_A2_p4.0	4.000000	0.075499	1.357647	638.149273	2	4.000000	0.228034	0.870745	390.313131	2	4.000000	0.298343	0.567256	376.284844	2	
136	E_GD_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.293779	1.703253	218.647293	2	
137	E_GD_A4	nan	0.152195	0.842421	811.444201	2	nan	0.316736	0.657233	443.384596	2	nan	0.377304	0.313439	487.386843	2	
138	E_FD_A2_p0.1	0.100000	0.616886	0.825071	561.099293	2	0.100000	0.301050	0.738522	403.577864	2	0.100000	0.229967	0.842450	353.601788	2	
139	E_FD_A2_p1.0	1.000000	0.594460	0.677435	580.393481	2	1.000000	0.298045	0.686538	434.050381	2	1.000000	0.225935	0.766275	388.800011	2	
140	E_FD_A2_p4.0	4.000000	0.578695	0.515718	648.792219	2	4.000000	0.320048	0.562335	529.551135	2	4.000000	0.252175	0.590430	504.735676	2	
141	E_FD_A3	nan	0.017390	1.840279	568.008804	2	nan	0.233692	0.997396	305.382579	2	nan	0.165489	1.156357	259.670591	2	
142	E_FD_A4	nan	0.347910	0.072083	1167.207078	2	nan	0.424196	0.375680	775.711314	2	nan	0.366976	0.376193	786.385542	2	
143	E_LK_R_A2_p0.1	0.100000	0.499112	1.223504	572.625228	2	0.100000	0.207175	1.172556	381.256088	2	0.100000	0.150415	1.306159	342.122554	2	
144	E_LK_R_A2_p1.0	1.000000	0.487913	1.012095	605.394165	2	1.000000	0.222694	1.052177	424.762349	2	1.000000	0.167391	1.134655	393.906023	2	
145	E_LK_R_A2_p4.0	4.000000	0.490156	0.703981	690.576926	2	4.000000	0.281575	0.842279	530.116626	2	4.000000	0.236858	0.858302	520.894993	2	
146	E_LK_R_A3	nan	0.002424	1.979467	569.059031	2	nan	0.135888	1.447620	313.280887	2	nan	0.083343	1.612213	278.640781	2	
147	E_LK_R_A4	nan	0.340584	0.139664	1160.190630	2	nan	0.362242	0.656735	670.539763	2	nan	0.319878	0.657098	677.123358	2	
148	E_GG_R_A2_p0.1	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0	0.100000	0.000000	nan	0.000000	0	
149	E_GG_R_A2_p1.0	1.000000	0.000000	nan	0.000000	0	1.000000	0.000000	nan	0.000000	0	1.000000	0.000000	nan	0.000000	0	
150	E_GG_R_A2_p4.0	4.000000	0.000000	nan	0.000000	0	4.000000	0.000000	nan	0.000000	0	4.000000	0.000000	nan	0.000000	0	
151	E_GG_R_A3	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0	
152	E_GG_R_A4	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0	nan	0.000000	nan	0.000000	0	

Эксперименты и результаты

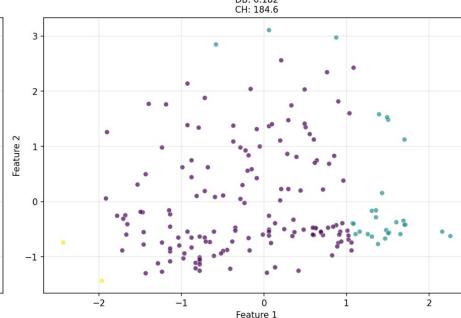
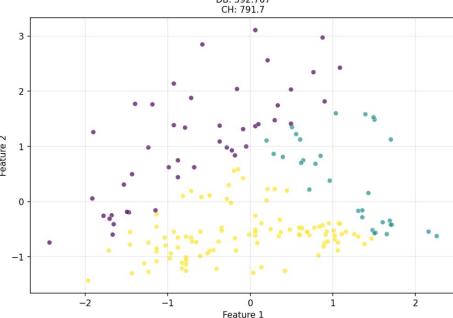
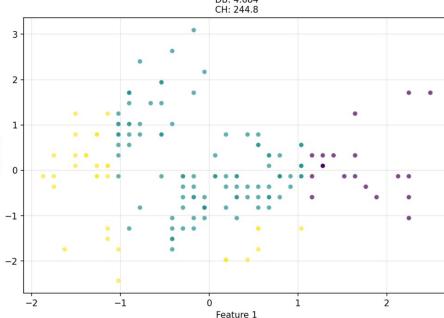
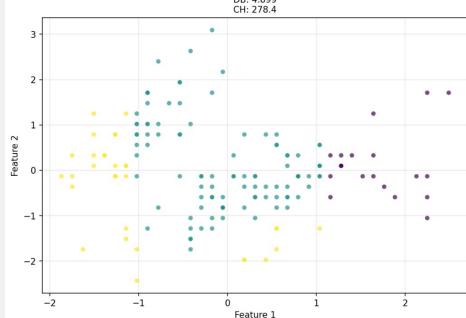
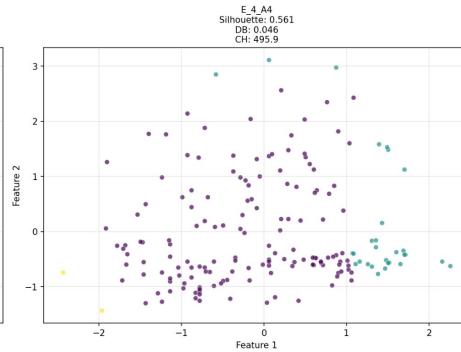
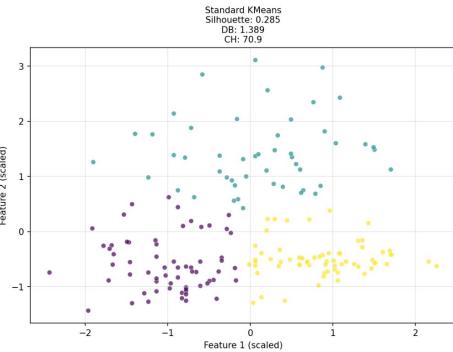
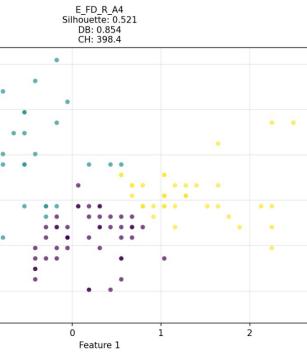
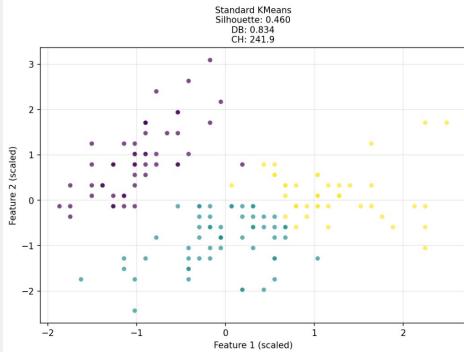
- Результаты презентующего (метрики)

Breast Cancer								make_circles								make_moons							
combination	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found	p	silhouette_score	davies_bouldin_score	calinski_harabasz_score	n_clusters_found								
153	E_GD_R_A2_p0.1	0.100000	-0.09594	1.960401	566.609728	2	0.100000	0.002968	1.987101	299.710987	2	0.100000	0.624235	1.368304	211.544280	2							
154	E_GD_R_A2_p1.0	0.000000	0.173706	1.356245	560.495330	2	1.000000	0.114737	1.145639	322.290302	2	1.000000	0.564190	0.823854	243.068244	2							
155	E_GD_R_A2_p4.0	0.400000	0.140714	1.071740	636.439419	2	4.000000	0.286315	0.570449	376.230608	2	4.000000	0.598185	0.391568	340.525533	2							
156	E_GD_R_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.473662	1.521554	152.938200	2							
157	E_GD_R_A4	nan	0.236140	0.423914	971.615097	2	nan	0.465244	0.300746	473.708219	2	nan	0.749374	0.126154	620.483645	2							
158	E_FD_R_A2_p0.1	0.100000	0.740795	0.842207	440.839953	2	0.100000	0.386247	0.691640	431.157215	2	0.100000	0.283158	0.865962	343.786449	2							
159	E_FD_R_A2_p1.0	0.000000	0.731478	0.597414	484.061172	2	1.000000	0.381388	0.592380	502.991703	2	1.000000	0.271707	0.714456	416.734452	2							
160	E_FD_R_A2_p4.0	0.400000	0.726480	0.370769	596.636283	2	4.000000	0.433021	0.401960	739.170455	2	4.000000	0.333984	0.440596	675.864898	2							
161	E_FD_R_A3	nan	0.004878	1.958933	571.156282	2	nan	0.288040	0.995511	312.408707	2	nan	0.187642	1.238338	244.242082	2							
162	E_FD_R_A4	nan	0.480743	0.009003	1398.639479	2	nan	0.613509	0.189250	1505.630709	2	nan	0.548384	0.190574	154.912397	2							
163	E_3_A2_p0.1	0.100000	-0.099311	1.962263	566.646993	2	0.100000	0.002803	1.987857	299.621796	2	0.100000	0.570548	1.422507	241.807997	2							
164	E_3_A2_p1.0	0.000000	0.146115	1.392907	561.814311	2	1.000000	0.105952	1.212161	325.382583	2	1.000000	0.487094	0.882531	269.561164	2							
165	E_3_A2_p4.0	0.400000	0.102088	1.157649	632.537200	2	4.000000	0.250674	0.681678	376.983863	2	4.000000	0.486678	0.456970	347.939186	2							
166	E_3_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.428523	1.567148	183.199617	2							
167	E_3_A4	nan	0.184505	0.566842	862.726603	2	nan	0.379954	0.432745	445.624888	2	nan	0.606810	0.197915	519.969348	2							
168	E_4_A2_p0.1	0.100000	-0.007955	1.959098	566.551958	2	0.100000	0.003063	1.986658	299.763356	2	0.100000	0.654519	1.337729	186.826453	2							
169	E_4_A2_p1.0	0.000000	0.194093	1.329870	556.924025	2	1.000000	0.114492	1.117771	317.127448	2	1.000000	0.604193	0.793373	217.467474	2							
170	E_4_A2_p4.0	0.400000	0.155591	1.017991	637.223287	2	4.000000	0.292918	0.523964	362.255664	2	4.000000	0.647828	0.362352	315.781767	2							
171	E_4_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.505614	1.489278	137.558103	2							
172	E_4_A4	nan	0.261818	0.347860	1022.155778	2	nan	0.496230	0.245689	459.763790	2	nan	0.806364	0.097468	649.549529	2							
173	E_5_A2_p0.1	0.100000	0.013996	1.985241	566.965866	2	0.100000	0.002177	1.990795	299.256170	2	0.100000	0.319127	1.676345	269.929263	2							
174	E_5_A2_p1.0	0.000000	0.047885	1.805807	588.357298	2	1.000000	0.114338	1.380619	342.282742	2	1.000000	0.217908	1.086924	312.782513	2							
175	E_5_A2_p4.0	0.400000	0.071457	1.476569	654.903970	2	4.000000	0.234751	0.962229	403.658565	2	4.000000	0.213342	0.617139	402.081680	2							
176	E_5_A3	nan	0.000000	nan	0.000000	1	nan	0.000000	nan	0.000000	1	nan	0.189139	1.808950	228.281254	2							
177	E_5_A4	nan	0.154970	1.000431	824.797907	2	nan	0.314378	0.765927	455.843367	2	nan	0.275758	0.364897	501.669730	2							
178	E_6_A2_p0.1	0.100000	0.506855	1.081706	581.522030	2	0.100000	0.220386	1.031369	385.273373	2	0.100000	0.164337	1.138816	348.817587	2							
179	E_6_A2_p1.0	0.000000	0.494867	0.929549	606.776970	2	1.000000	0.229557	0.951918	417.358425	2	1.000000	0.174694	1.027056	386.829599	2							
180	E_6_A2_p4.0	0.400000	0.489087	0.67902	678.582912	2	4.000000	0.272659	0.786037	505.127181	2	4.000000	0.225249	0.805659	493.280087	2							
181	E_6_A3	nan	0.009120	1.914519	568.106336	2	nan	0.155107	1.300708	309.941977	2	nan	0.102984	1.449208	275.547776	2							
182	E_6_A4	nan	0.335029	0.0135848	1153.282845	2	nan	0.355259	0.603127	649.055551	2	nan	0.308115	0.604903	653.806970	2							

Эксперименты и результаты

- Результаты презентующего (сравнение)

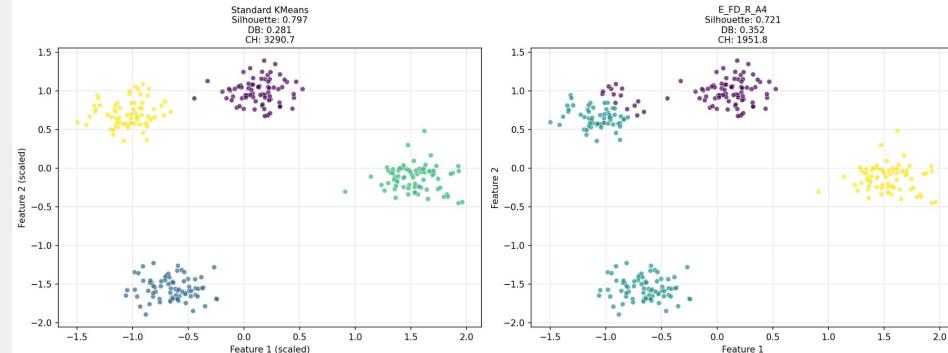
Clustering Comparison for Iris
Standard KMeans vs Top 3 Fuzzy Methods



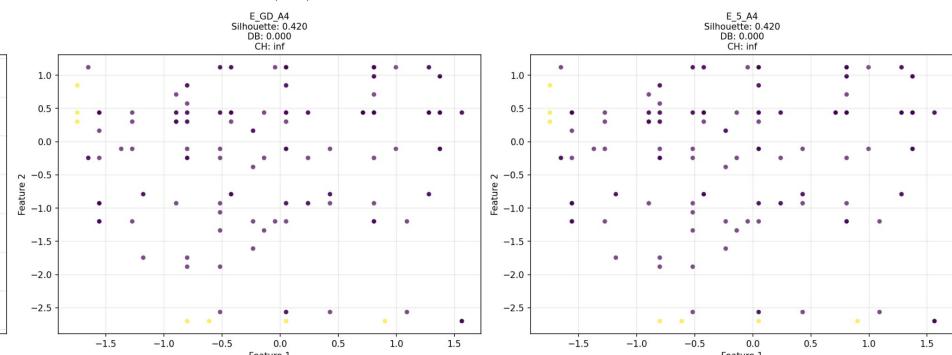
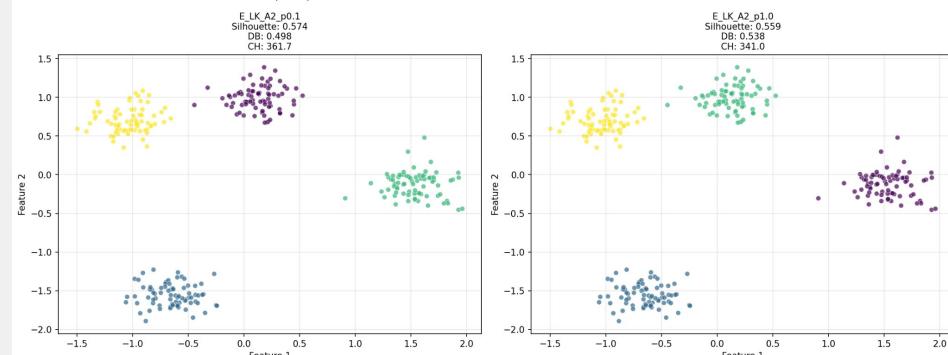
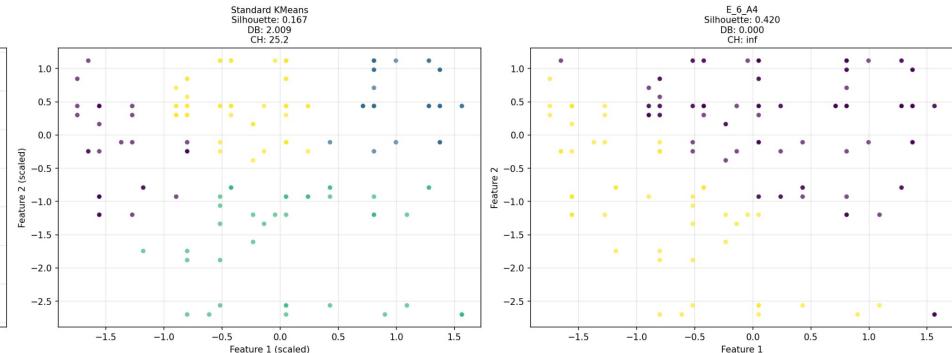
Эксперименты и результаты

- Результаты презентующего (сравнение)

Clustering Comparison for Blobs
Standard KMeans vs Top 3 Fuzzy Methods



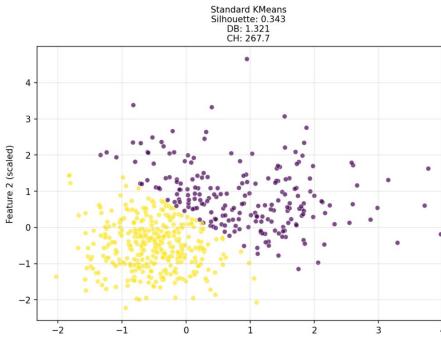
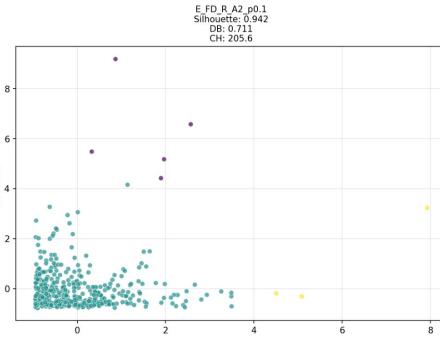
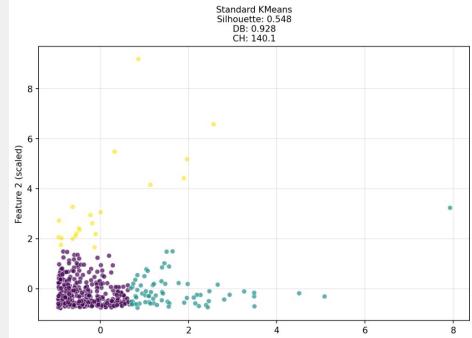
Clustering Comparison for Absenteeism
Standard KMeans vs Top 3 Fuzzy Methods



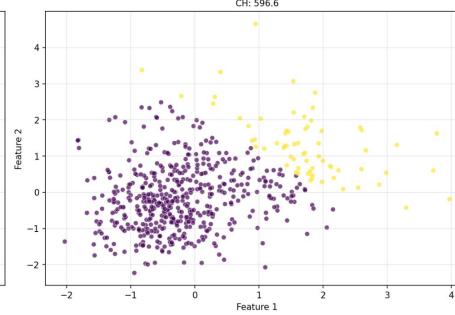
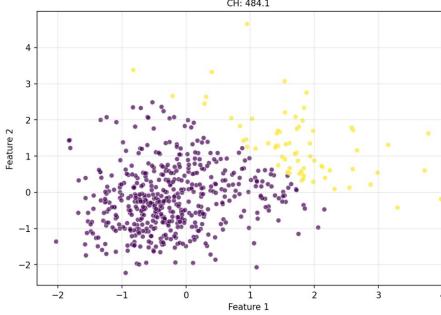
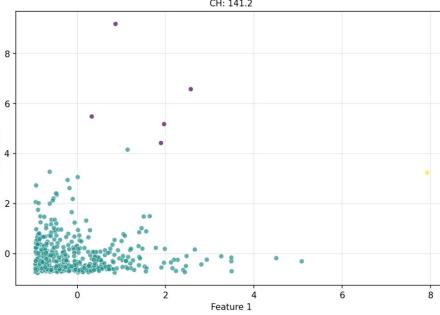
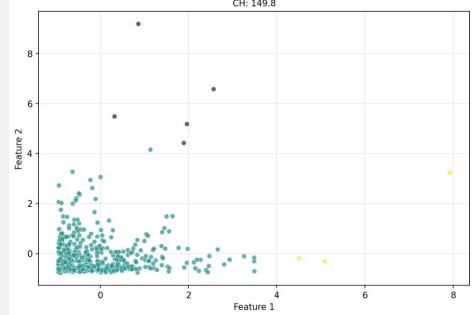
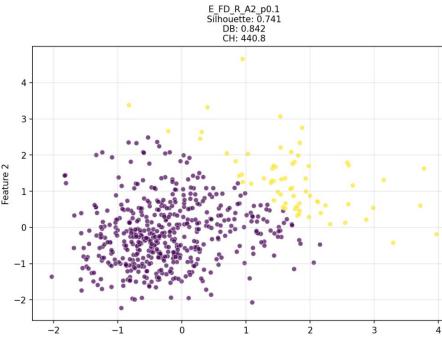
Эксперименты и результаты

- Результаты презентующего (сравнение)

Clustering Comparison for Wholesale_Customers
Standard KMeans vs Top 3 Fuzzy Methods



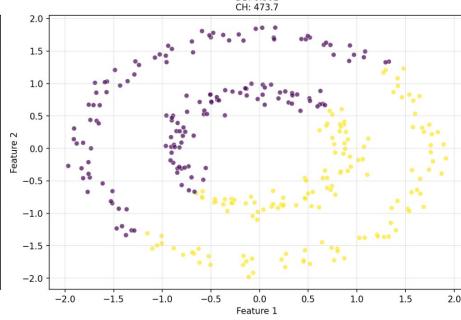
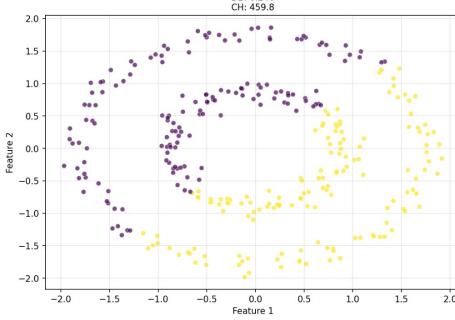
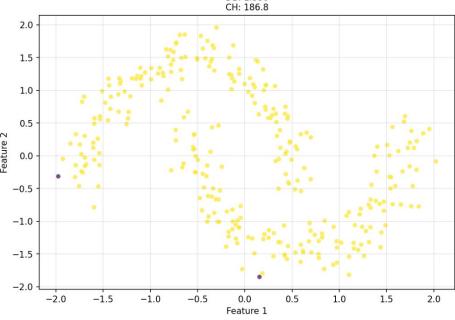
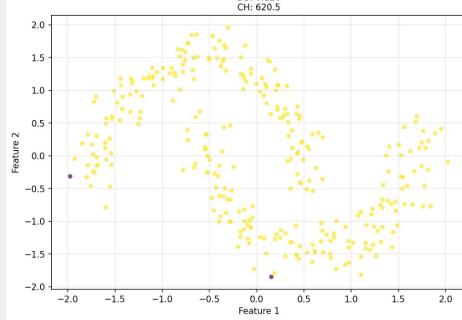
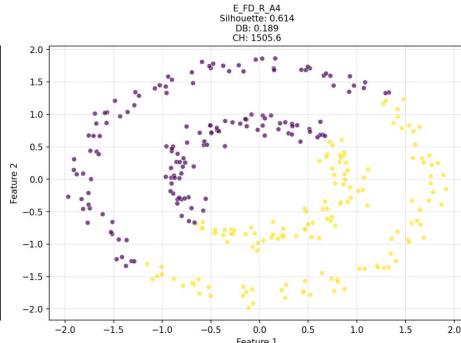
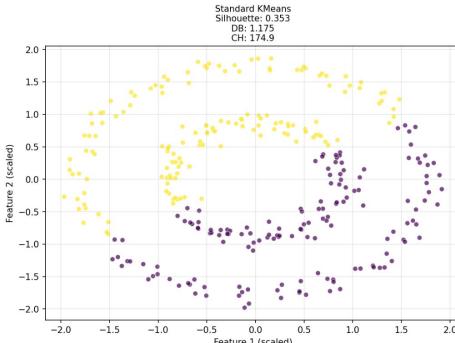
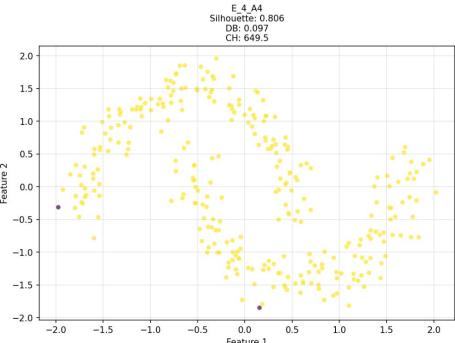
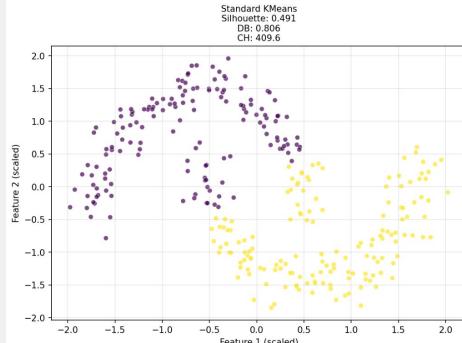
Clustering Comparison for Breast_Cancer
Standard KMeans vs Top 3 Fuzzy Methods



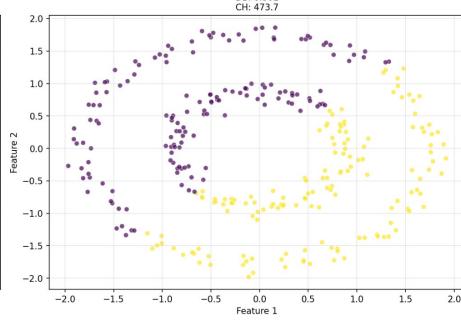
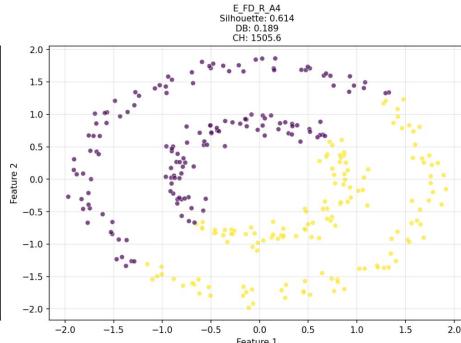
Эксперименты и результаты

- Результаты презентующего (сравнение)

Clustering Comparison for Moons
Standard KMeans vs Top 3 Fuzzy Methods



Clustering Comparison for Circles
Standard KMeans vs Top 3 Fuzzy Methods





Эксперименты и результаты



- **Результаты презентующего**

- (выводы)**

- По частоте попадания в топ-3 на основе коэффициента силуэта наиболее «успешными» оказались:

Агрегации: A_4 - 14 раз; $A_{2(4)}$ – 5 раз; $A_{2(0.1)}$ – 4 раза; $A_{2(4)}$ – 11 раз; A_4 – 10 раз; A_3 – 3 раза.

Нечеткие эквивалентности: EFD – 12 раз; E_4 – 7 раз; ELK – 3 раза; EGD – 3 раза; E_3 – 1 раз; E_5 - 1 раз; E_6 – 1 раз.

Пары: $A_4(EFD)$ - 4 раза; $A_{2(4)}(EFD)$ - 4 раза; $A_4(E_4)$ - 3 раза.

- По частоте попадания в топ-3 на основе индекса Дэвиса-Болдина наиболее «успешными» оказались:

Агрегации: A_4 - 22 раза; $A_{2(0.1)}$ – 3 раза; $A_{2(4)}$ – 1 раз; $A_{2(4)}$ – 11 раз; A_4 – 10 раз; A_3 – 3 раза.

Нечеткие эквивалентности: EFD – 8 раз; ELK – 6 раз; E_4 – 4 раза; E_3 – 3 раза; EGD – 2 раза; E_6 – 2 раза; E_5 - 1 раз.

Пары: $A_4(EFD)$ - 7 раз; $A_4(E_4)$ - 4 раза; $A_4(E_3)$ - 4 раза.





Эксперименты и результаты



- Результаты презентующего
(выводы)
 - По частоте попадания в топ-3 на основе индекса Калински-Харабаша наиболее «успешными» оказались:
 - Агрегации: A_4 - 26 раз.
 - Нечеткие эквивалентности: EFD – 13 раз; ELK – 6 раз; E_4 – 3 раза; E_6 – 2 раза; EGD – 2 раза.
 - Пары: $A_4(EFD)$ - 13 раз; $A_4(ELK)$ - 6 раз; $A_4(E_4)$ - 3 раза.
 - Для многих наборов данных композиции дают метрики сопоставимые или лучше, чем классический k-means с евклидовой метрикой.
 - Утверждение авторов о более надежных результатах для сферических данных правдиво.
 - Визуализация результатов показывает, что доверять только одной метрике, которая предложена авторами, нельзя. Таким образом оценка результатов должна происходить на основе нескольких факторов.





Спасибо за внимание!

**Если у вас есть какие-то
вопросы,
я готов на них ответить**

