

**МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И.УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ**

**ОТЧЁТ
по практической работе №6
по дисциплине «Машинное обучение»
Тема: DBSCAN**

Студент гр. 6304

Преподаватель

Корытов П.В.

Жангиров Т.Р.

Санкт-Петербург

2020

1. Задание 1

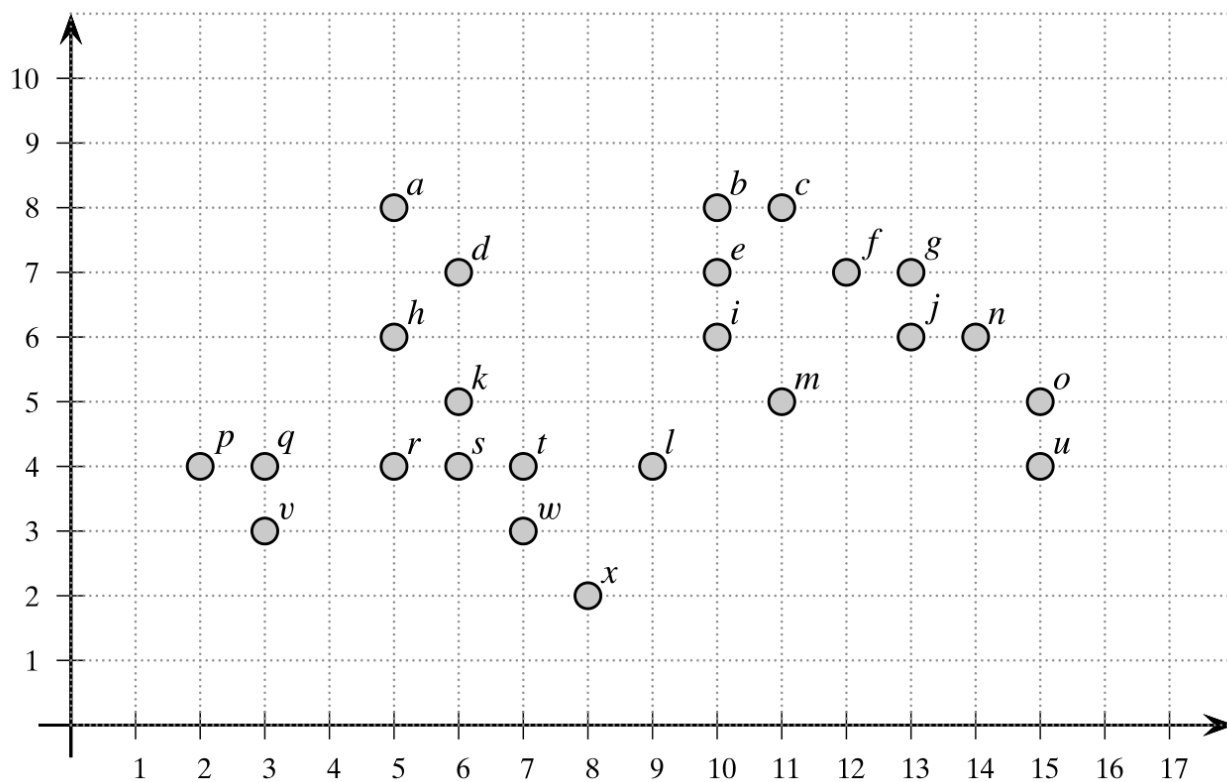


Рисунок 1 – Набор данных

$\epsilon = 2, minpts = 3.$

Таблица 1. Точки

x	$N_\epsilon(X)$	Core?
p	q, v	
q	p, v, r	+
v	p, q	
r	q, h, k, s, t	+
h	a, d, k, r	+
a	d, h	
d	a, h, k	+
k	d, t, s, r, h	+
s	k, r, t, w	+
t	l, w, s, r, k	+
w	s, t, x	+
x	w	
l	t	
b	c, e, i	+
c	b, e, f	+
e	b, c, f, i	+
f	e, c, g, j	+
g	f, j, n	+
i	e, b, m	+
j	f, g, n	+
n	g, j, o	+
m	i	
o	n, u	
u	o	

1. Основные точки — q, r, h, d, k, s, t, w, b, c, e, f, g, i, j, n
2. a — прямо достижима по плотности из d, т.к. $a \in N_\epsilon(d)$ и d — основная точка.
3. o — достижима по плотности из i, т.к.

$$i \rightarrow e \rightarrow f \rightarrow j \rightarrow n \rightarrow o, \quad (1.1)$$

где $\alpha \rightarrow \beta \equiv \beta \in N_\epsilon(\alpha)$ и α — основная точка.

4. Кластеры:

4.1. p, q, v, r, h, a, d, k, s, t, w, x, l

4.2. b, c, e, f, g, i, j, n, m, o

Выпавшая точка — и.

Дерево рекурсивного спуска представлено на рис. 2.

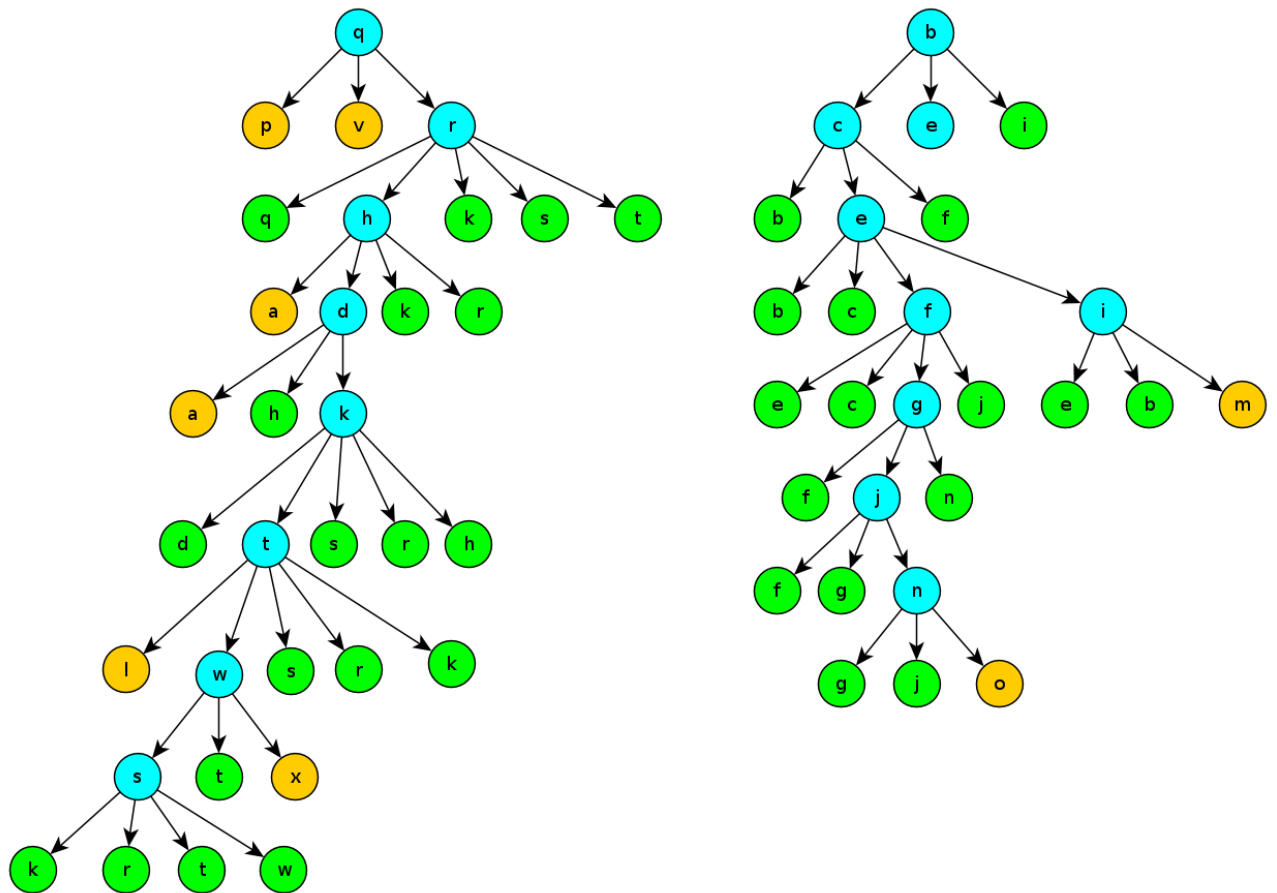


Рисунок 2 – DBSCAN

Желтые точки — обычные, синие — основные, зеленые — уже обработанные основные.

2. Задание 2

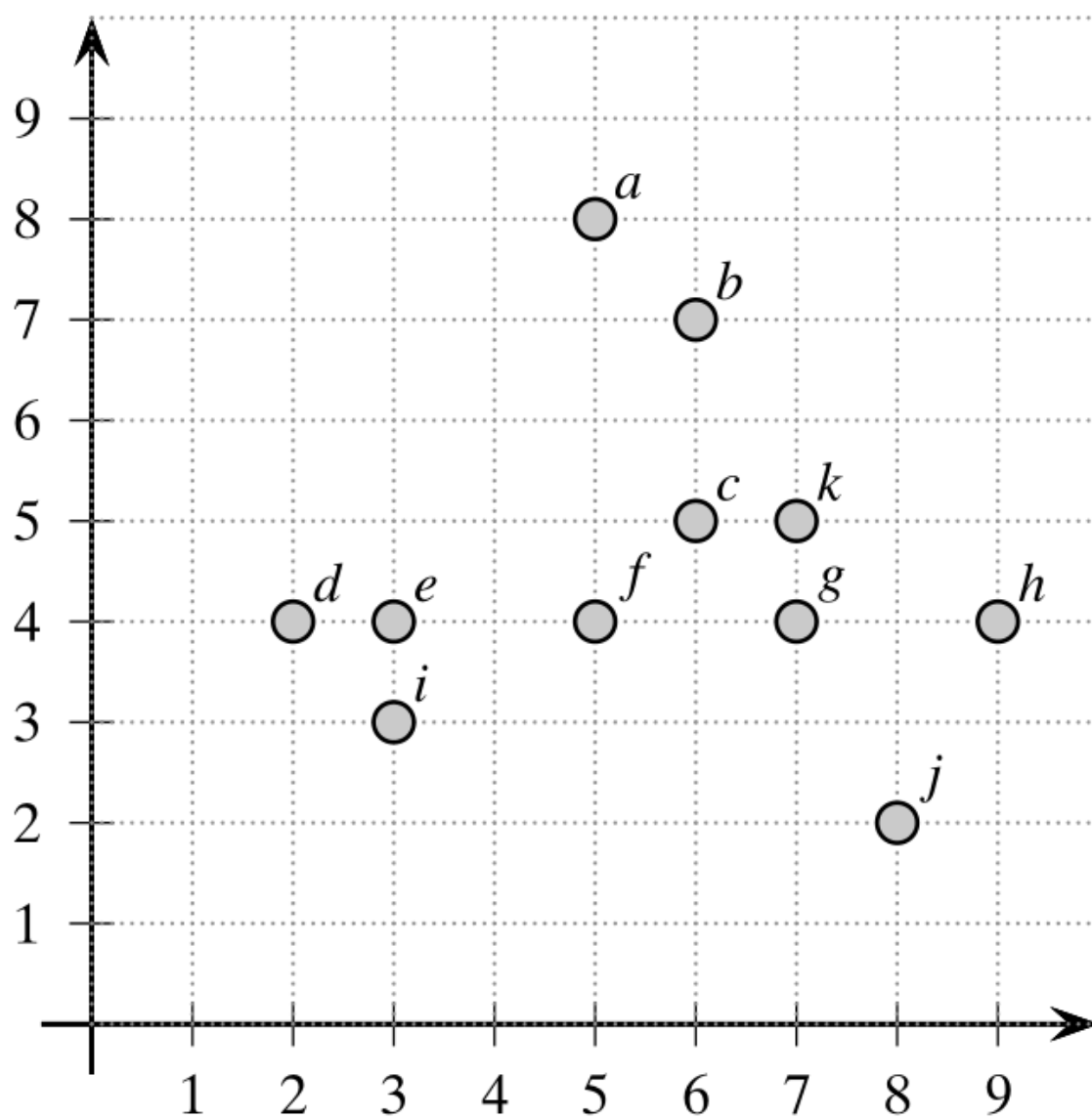


Рисунок 3 – Q2

2.1. L_∞

$$L_\infty(x, y) = \max_{i=1}^d \{|x_i - y_i|\}. \quad (2.1)$$

$\epsilon = 2, \text{minpts} = 5.$

Таблица 2. Результаты работы DBSCAN

х	$N_\epsilon(x)$	Основная?	Кластер
a	b		Выпала
b	a, c, k		1
c	b, k, g, f		1
d	e, i		Выпала
e	d, f, i		1
f	i, e, c, k, g	+	1
g	f, c, k, h, j	+	1
h	g, k, j		1
i	d, e, f		1
j	g, h		1
k	f, c, b, g, h	+	1

2.2. $L_{\frac{1}{2}}$

$\epsilon = 4, \text{minpts} = 3.$

Рассчитаны расстояния между всеми точками:

Таблица 3. Расстояния между точками по норме $L_{\frac{1}{2}}$

	a	b	c	d	e	f	g	h	i	j	k
a	0.00	4.00	7.46	13.93	11.66	4.00	11.66	16.00	13.32	17.49	9.90
b	4.00	0.00	2.00	13.93	12.00	7.46	7.46	12.00	13.93	13.32	5.83
c	7.46	2.00	0.00	9.00	7.46	4.00	4.00	7.46	9.90	9.90	1.00
d	13.93	13.93	9.00	0.00	1.00	3.00	5.00	7.00	4.00	14.93	10.47
e	11.66	12.00	7.46	1.00	0.00	2.00	4.00	6.00	1.00	13.32	9.00
f	4.00	7.46	4.00	3.00	2.00	0.00	2.00	4.00	5.83	9.90	5.83
g	11.66	7.46	4.00	5.00	4.00	2.00	0.00	2.00	9.00	5.83	1.00
h	16.00	12.00	7.46	7.00	6.00	4.00	2.00	0.00	11.90	5.83	5.83
i	13.32	13.93	9.90	4.00	1.00	5.83	9.00	11.90	0.00	10.47	11.66
j	17.49	13.32	9.90	14.93	13.32	9.90	5.83	5.83	10.47	0.00	7.46
k	9.90	5.83	1.00	10.47	9.00	5.83	1.00	5.83	11.66	7.46	0.00

Таблица 4. Результаты работы DBSCAN

х	$N_{\epsilon}(x)$	Основная?	Кластер
a	b, f		1
b	a, c		1
c	b, f, g, k	+	1
d	e, f, i	+	1
e	d, f, g, i	+	1
f	a, c, d, e, g, h	+	1
g	c, e, f, h, k	+	1
h	f, g		1
i	d, e		1
j			Выпала
k	c, g, k	+	1

2.3. L_{\min}

$$L_{\min}(x, y) = \min_{i=1}^d \{|x_i - y_i|\}. \quad (2.2)$$

$\epsilon = 1, minpts = 6.$

Таблица 5. Расстояния между точками по норме L_{\min}

	a	b	c	d	e	f	g	h	i	j	k
a	0.00	1.00	1.00	3.00	2.00	0.00	2.00	4.00	2.00	3.00	2.00
b	1.00	0.00	0.00	3.00	3.00	1.00	1.00	3.00	3.00	2.00	1.00
c	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	2.00	2.00	0.00
d	3.00	3.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	1.00
e	2.00	3.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	1.00
f	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	1.00
g	2.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00
h	4.00	3.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
i	2.00	3.00	2.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	2.00
j	3.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	0.00	1.00
k	2.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	2.00	1.00	0.00

Таблица 6. Результаты работы DBSCAN

x	$N_\epsilon(x)$	Основная?	Кластер
a	b, c, f		1
b	a, c, f, g, k		1
c	a, b, d, e, f, g, h, k	+	1
d	c, e, f, g, h, i, k	+	1
e	c, d, f, g, h, i, k	+	1
f	a, b, c, d, e, g, h, i, k	+	1
g	b, c, d, e, f, h, i, j, k	+	1
h	c, d, e, f, g, i, j, k	+	1
i	d, e, f, g, h, j	+	1
j	g, h, i, k		1
k	b, c, d, e, f, g, h, j	+	1

2.4. L_{\min}

$\epsilon = 4, \minpts = 6$.

Поскольку с $\epsilon = 1$ и той же метрикой все точки оказались в одном кластере, с $\epsilon = 4$ результат будет такой же.