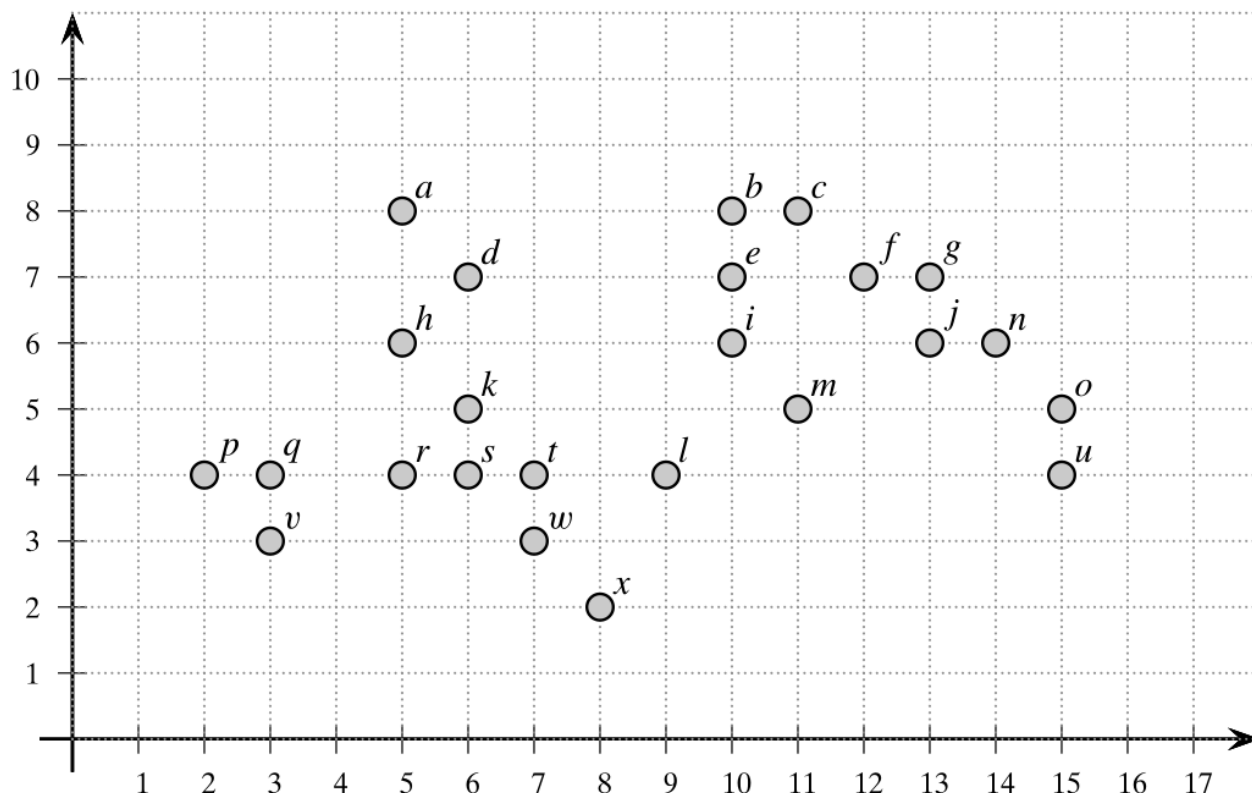


ЗАДАНИЕ 1



Допустим, что используется Евклидово расстояние, $\epsilon = 2$ и $\text{minPts} = 3$. Выполните следующие задачи:

1. Выпишите список всех основных точек
2. Покажите, является ли точка a прямо достижимой из точки d
3. Покажите, является ли точка o достижимой по плотности из точки i . Если нет, то покажите на какой точке цепочка построения пути оборвалась.
4. Покажите кластеры полученные алгоритмом DBSCAN и выпавшие точки.

1. x — основная точка, если в его соседстве есть более minpts точек.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	r	s	t	u	v	w	x	
core point	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	+	-	+	+	-

2. x — прямо достижима из u , если u входит в соседство x и является основной

a прямо достижима из d , потому что

$a \in N_{\epsilon}(d) = \{a, h, k, d\} +$

d — core point +

З х — достижима по плоности из точки у, если существует цепочка из точек, которые последовательно прямо достижимы.

о достижима по плотности из i, потому что есть цепочка:

i — e
e — f
f — j
j — n
n — o

4.

e=2 minpts=3

1. Найти главные точки

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	r	s	t	u	v	w	x
core point	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	-	+	+	-

Кластеры:

1 a d h k r q p v s t w x l
2 b c e i m f g j n o u

Выпавших точек нет.

ЗАДАНИЕ 2

Даны следующие метрики:

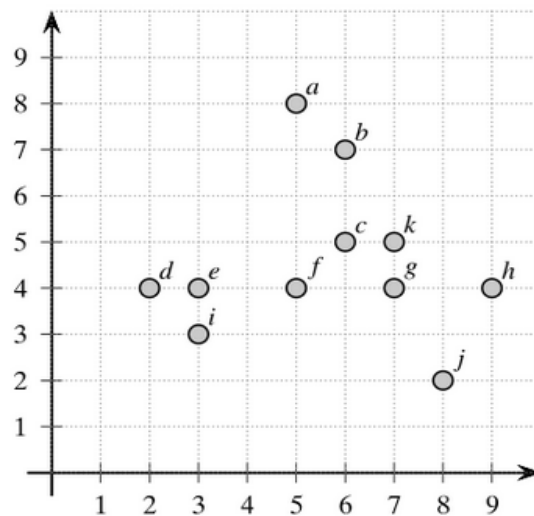
$$1. \quad L_{\infty}(\mathbf{x}, \mathbf{y}) = \max_{i=1}^d \{|x_i - y_i|\}$$

$$2. \quad L_{\frac{1}{2}}(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^d |x_i - y_i|^{\frac{1}{2}} \right)^2$$

$$3. \quad L_{\min}(\mathbf{x}, \mathbf{y}) = \min_{i=1}^d \{|x_i - y_i|\}$$

$$4. \quad L_{pow}(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^d 2^{i-1} (x_i - y_i)^2 \right)^{1/2}$$

Для данных представленных следующим рисунком



Используя метод DBSCAN проведите кластеризацию при следующих параметрах:

- $\epsilon = 2$ и $\text{minPts} = 5$ и метрика (1)
- $\epsilon = 4$ и $\text{minPts} = 3$ и метрика (2)
- $\epsilon = 1$ и $\text{minPts} = 6$ и метрика (3)
- $\epsilon = 4$ и $\text{minPts} = 6$ и метрика (3)

Был написан скрипт для визуализации кластеризации.

Желтый — попал в кластеризации

Голубой — выпал из кластера

```
test_bases = {0: [2, 5, lambda x,y:np.max(np.abs(x-y))],
               1: [4, 3, lambda x,y:pow(np.sum(np.sqrt(np.abs(x-y))), 2)],
               2: [1, 6, lambda x,y:np.min(np.abs(x-y))],
               3: [4, 6, lambda x,y: np.sum([2**i * (x[i] - y[i]) **2 for i in
range(len(x))])] ** (1 / 2)]}

for i in range(len(test_bases)):
    clustering = DBSCAN(eps=test_bases[i][0],
                        min_samples=test_bases[i][1],
                        metric=test_bases[i][2]).fit(data)

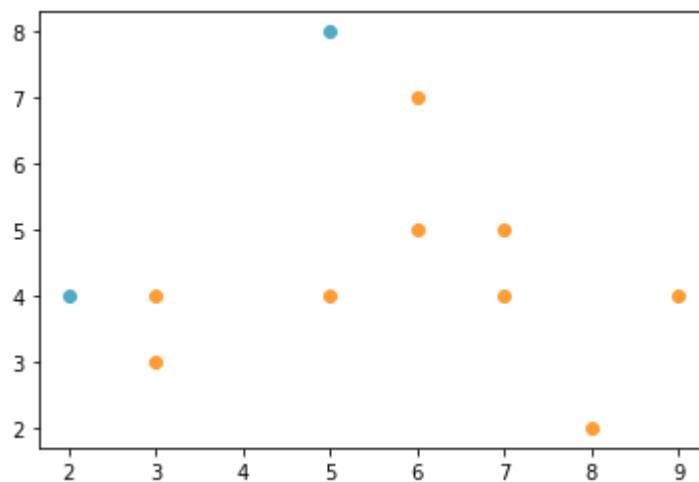
    print(set(clustering.labels_))
    print(len(set(clustering.labels_)) -1)
    print(list(clustering.labels_).count(-1) /len(list(clustering.labels_)))

    colors = ['#4EACC5', '#FF9C34', '#4E9A06']
    labels = clustering.labels_ + 1

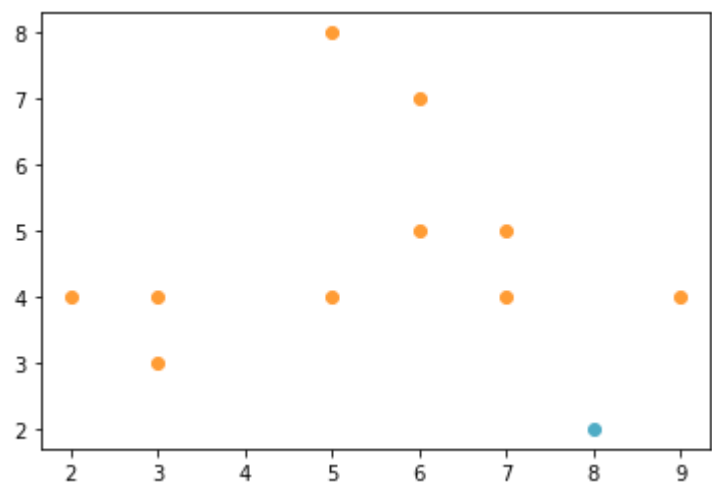
    for i, row in data.iterrows():
        plt.scatter(row[0], row[1], c=colors[labels[i]])

plt.show()
```

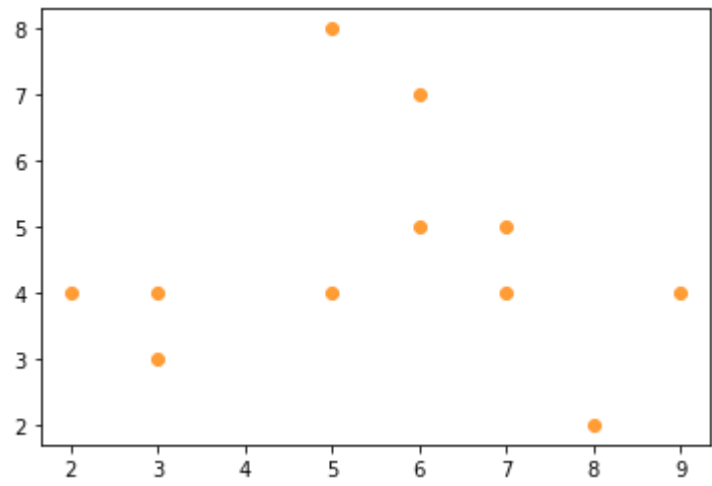
1 метрика



2 метрика



3 метрика



4 метрика

