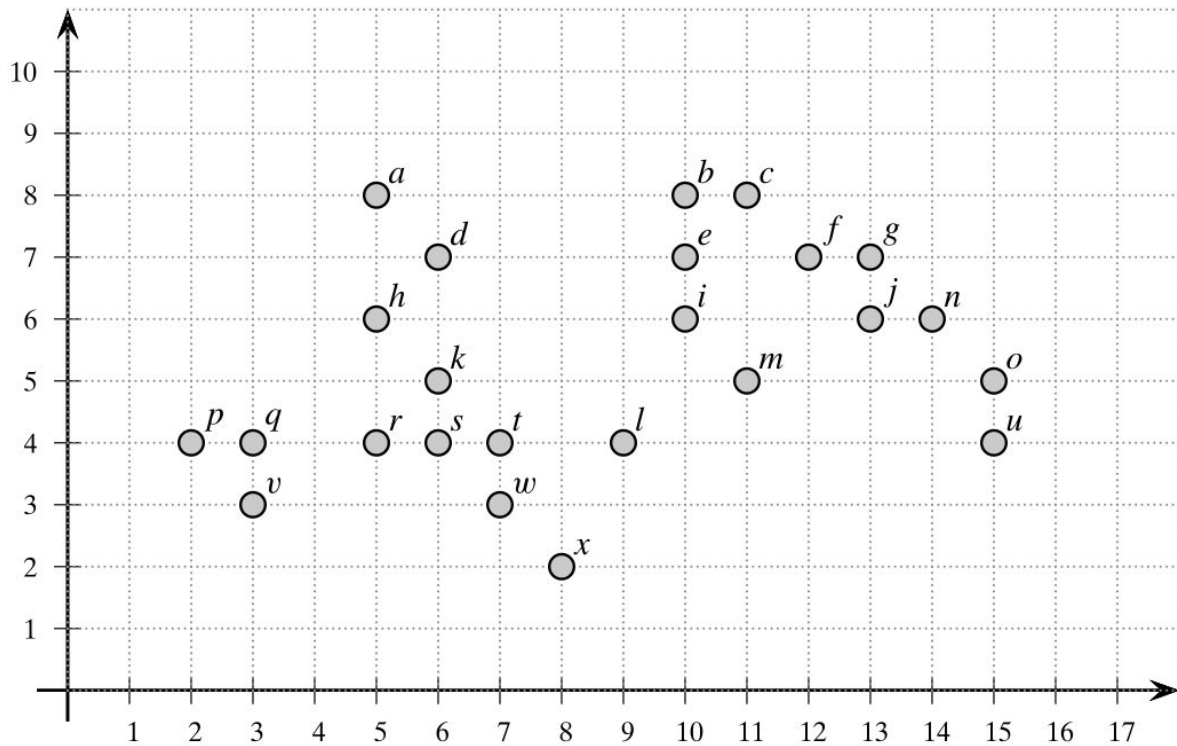


# Задание №1

Дан рисунок



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

А. Выпишите список всех основных точек

Точка	$N_\epsilon$	$ N_\epsilon  \geq \text{minPts}$
a	{a, d, h}	+
b	{b, c, e}	+
c	{b, c, e, f}	+
...	...	...

Основные точки: a, b, c, d, e, f, g, h, i, j, k, n, o, p, r, s, t, v, w

В. Покажите, является ли точка a прямо достижимой из точки d

$a \in N_\epsilon(d) = \{a, d, h, k\}$ ; d - основная точка => a прямо достижима из d

С. Покажите, является ли точка o достижимой по плотности из точки i. Если нет, то покажите на какой точке цепочка построения пути оборвалась.

есть цепочка  $i \rightarrow e \rightarrow f \rightarrow j \rightarrow n \rightarrow o$ ;  $i, e, f, j, n$  - основные точки  $\Rightarrow$  о достижима из  $i$

D. Покажите кластеры полученные алгоритмом DBSCAN и выпавшие точки.

C1: a, d, h, k, p, q, v, r, s, t, l, w, x

C2: b, c, e, f, g, i, j, n, m, o, u

## Задание №2

Даны следующие метрики:

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

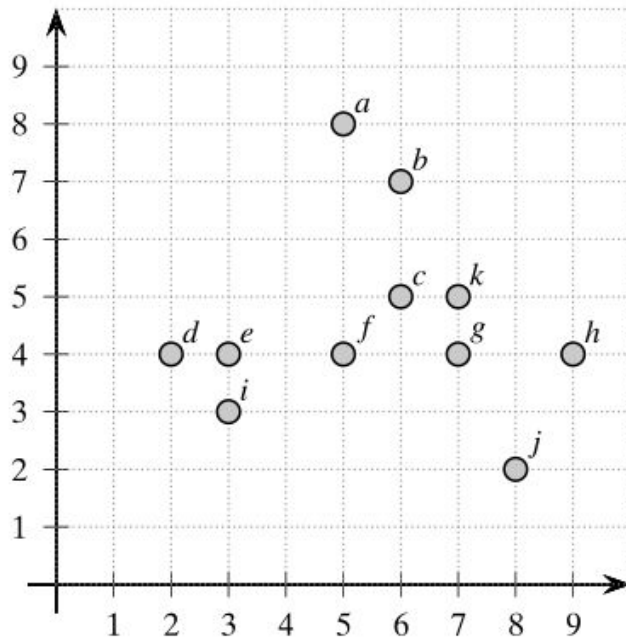
$$L_{\frac{1}{2}}(\mathbf{x}, \mathbf{y}) = \left( \sum_{i=1}^d |x_i - y_i|^{\frac{1}{2}} \right)^2$$

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

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

```
1. L_inf = lambda x, y: np.max(np.abs(x - y))
2. L_1div2 = lambda x, y: np.sum(np.abs(x - y) ** (1 / 2)) ** 2
3. L_min = lambda x, y: np.min(np.abs(x - y))
4. L_pow = lambda x, y: np.sum([2 ** i * (x[i] - y[i]) ** 2 for i in
    range(len(x))]) ** (1 / 2)
```

Для данных представленных следующим рисунком



```
1. X = np.array([[2, 4], [3, 4], [3, 3], [5, 4], [5, 8],
2.             [6, 5], [6, 7], [7, 4], [7, 5], [8, 2], [9, 4]])
```

Используя метод DBSCAN проведите кластеризацию при следующих параметрах, для всех случаев построить кластеры и отобразить основные точки, достижимые по плотности точки и выпавшие точки:

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

```
1. clustering = DBSCAN(eps=2, min_samples=5, metric=L_inf).fit(X)
2. print(clustering.labels_)
3. print(clustering.components_)
```

Принадлежность точек к кластеру (-1 - выпала): [-1 0 0 0 -1 0 0 0 0 0 0]

Основные точки:

```
[[5 4]
 [6 5]
 [7 4]
 [7 5]]
```

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

Принадлежность точек к кластеру (-1 - выпала): [0 0 0 0 0 0 0 0 0 -1 0]

Основные точки:

```
[[2 4]
 [3 4]
 [3 3]
 [5 4]]
```

```
[5 8]
[6 5]
[6 7]
[7 4]
[7 5]
[9 4]]
```

```
1. clustering = DBSCAN(eps=4, min_samples=3, metric=L_1div2).fit(X)
2. print(clustering.labels_)
3. print(clustering.components_)
```

- $\epsilon = 1$  и minPts = 6 и метрика (3)

```
1. clustering = DBSCAN(eps=1, min_samples=6, metric=L_min).fit(X)
2. print(clustering.labels_)
3. print(clustering.components_)
```

Принадлежность точек к кластеру (-1 - выпала): [0 0 0 0 0 0 0 0 0 0]

Основные точки:

```
[[2 4]
 [3 4]
 [3 3]
 [5 4]
 [6 5]
 [6 7]
 [7 4]
 [7 5]
 [9 4]]
```

- $\epsilon = 4$  и minPts = 6 и метрика (4)

```
1. clustering = DBSCAN(eps=4, min_samples=6, metric=L_pow).fit(X)
2. print(clustering.labels_)
3. print(clustering.components_)
```

Принадлежность точек к кластеру (-1 - выпала): [ 0 0 0 0 -1 0 0 0 0 0]

Основные точки:

```
[[3 4]
 [5 4]
 [6 5]
 [7 4]
 [7 5]
 [9 4]]
```

[6 7]  
[7 4]  
[7 5]  
[9 4]]