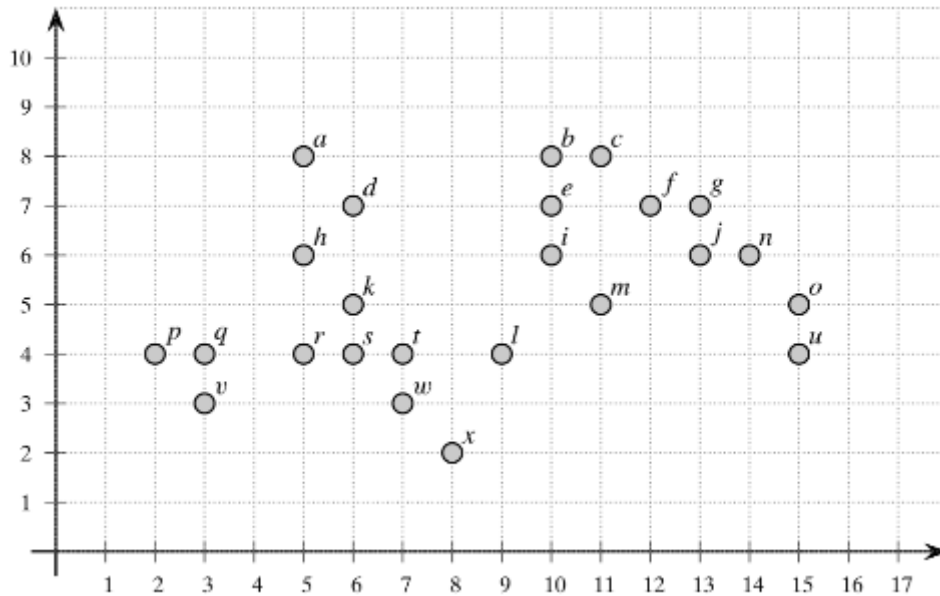


Практические задания №5. Григорьев И.С. 6304

Задание №1

Дан рисунок



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

1. Выпишите список всех основных точек

Точка x основная, если количество других точек, попадающих внутрь окружности с центром в этой точке и радиусом ϵ ($N_\epsilon(x)$), больше либо равно чем minPts .

Основные точки: $q, r, s, t, w, k, h, d, i, e, b, c, f, g, j, n$

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

Точка a прямо достижима из d , т.к. a принадлежит $N_\epsilon(d) = \{a, h, k\}$ и d — основная точка.

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

Является, т.к. есть набор основных точек, ведущих от i к o .

$$i \rightarrow e \rightarrow f \rightarrow j \rightarrow n \rightarrow o$$

4. Покажите кластеры, полученные алгоритмом DBSCAN и выпавшие точки.

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

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

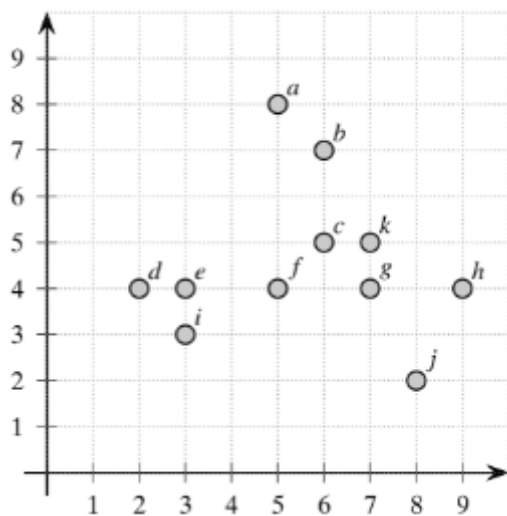
Выпавшая точка u

Задание №2

Даны следующие метрики:

1.
$$L_{\infty}(\mathbf{x}, \mathbf{y}) = \max_{i=1}^d \{ |x_i - y_i| \}$$
2.
$$L_{\frac{1}{2}}(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^d |x_i - y_i|^{\frac{1}{2}} \right)^2$$
3.
$$L_{\min}(\mathbf{x}, \mathbf{y}) = \min_{i=1}^d \{ |x_i - y_i| \}$$
4.
$$L_{pow}(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^d 2^{i-1} (x_i - y_i)^2 \right)^{1/2}$$

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 minPts = 6 и метрика (3)
- $\epsilon = 4$ и minPts = 6 и метрика (3)

Для всех случаев построить кластеры и отобразить основные точки, достижимые по плотности точки и выпавшие точки

```
def print_cl(cl):
    core = [labels[i] for i in cl.core_sample_indices_]
    print('Основные точки:', core)
    cl_num = len(set([i for i in cl.labels_ if i != -1]))
    for j in range(cl_num):
        in_cl = [labels[i] for i, x in enumerate(cl.labels_) if x == j ]
        print(f'Кластер {j + 1}: ', in_cl)
    noise = [labels[i] for i, x in enumerate(cl.labels_) if x == -1 ]
    print('Выпавшие точки:', noise)
```

```
cl = DBSCAN(eps=2, min_samples=6, metric=lambda x, y: np.max(np.abs(x - y))).fit(X)
print_cl(cl)
```

```
Основные точки: ['f', 'g', 'k']
Кластер 1: ['b', 'c', 'e', 'f', 'g', 'h', 'i', 'j', 'k']
Выпавшие точки: ['a', 'd']
```

```
cl = DBSCAN(eps=4, min_samples=4, metric=lambda x, y: np.sum(np.abs(x - y) ** 0.5) ** 2).fit(X)
print_cl(cl)
```

```
Основные точки: ['c', 'd', 'e', 'f', 'g']
Кластер 1: ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'k']
Выпавшие точки: ['j']
```

```
cl = DBSCAN(eps=1, min_samples=7, metric=lambda x, y: np.min(np.abs(x - y))).fit(X)
print_cl(cl)
```

```
Основные точки: ['c', 'd', 'e', 'f', 'g', 'h', 'i', 'k']
Кластер 1: ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k']
Выпавшие точки: []
```

```
cl = DBSCAN(eps=4, min_samples=7, metric=lambda x, y: np.min(np.abs(x - y))).fit(X)
print_cl(cl)
```

```
Основные точки: ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k']
Кластер 1: ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k']
Выпавшие точки: []
```