

Omówienie przykładu dla zadania *Klasteryzator*

Dane wejściowe:

3
0 1 4
euk
1.0, 1.0
1.0, 2.0
4.0, 3.0
4.0, 4.0
5.0, 8.0
5.0, 9.0

Interpretacja danych:

- dane są **3** centra
- centra ulokowane są w tym samym punkcie, co próbki o numerach **0, 1 i 4**
- zostanie użyta metryka **euklidesowa**
- klasteryzacji poddanych zostanie **6** próbek (zauważ, że współrzędne punktów dane są parami liczb, a więc wyznaczanie odległości dotyczyć będzie płaszczyzny)

Krok 1. Wyznaczenie odległości oraz przypisanie próbek do centrum:

- oblicza się odległość od każdego z punktów do każdego z centrów, np. w metryce euklidesowej odległość pomiędzy punktem P1 (1,0;2,0) a centrum C2 (5,0;8,0) wyniesie (w zaokrągleniu) 7,211, gdyż:

$$\sqrt{(5-1)^2+(8-2)^2} \approx 7,211$$

- wyniki tej fazy obliczeń obrazuje poniższa tabela (dla każdego z punktów kolor zielony wskazuje minimum z odległości do centrów)

	C0	C1	C2
P0	0,0000	1,0000	8,0623
P1	1,0000	0,0000	7,2111
P2	3,6056	3,1623	5,0990
P3	4,2426	3,6056	4,1231
P4	8,0623	7,2111	0,0000
P5	8,9443	8,0623	1,0000

- teraz każdą z próbek przypisuje się do tego centrum, które jest jej najbliższe – oznacza to, że uzyskamy następujące grupy:

C0 = [P0]

C1 = [P1, P2, P3]

C2 = [P4, P5]

Krok 2. Wyznaczenie nowych centrów:

- oblicza się skorygowane położenie centrów jako średnią arytmetyczną współrzędnych wszystkich punktów w grupie:
- tym samym centra przemieszczają się do następujących punktów:

$$C0 = \text{średnia}(P0) = (1,0; 1,0)$$

$$C1 = \text{średnia}(P1, P2, P3) = \text{średnia}((1,0; 2,0), (4,0; 3,0), (4,0; 4,0)) = (3,0; 3,0)$$

$$C2 = \text{średnia}(P4, P5) = \text{średnia}((5,0; 8,0), (5,0; 9,0)) = (5,0; 8,5)$$

Krok 3. Przypisanie próbek do nowych centrów --- koniec 1 iteracji:

- powtarza się czynności z kroku 1. i na nowo przypisuje się każdy z punktów do aktualnych centrów – tabela odległości przybierze tym razem postać następującą:

	C0	C1	C2
P0	0,0000	2,8284	8,5000
P1	1,0000	2,2361	7,6322
P2	3,6056	1,0000	5,5902
P3	4,2426	1,4142	4,6098
P4	8,0623	5,3852	0,5000
P5	8,9443	6,3246	0,5000

- powtórnie każdą z próbek przypisuje się do tego centrum, które jest jej najbliższe, co w efekcie spowoduje powstanie następujących grup:

$$C0 = [P0, P1]$$

$$C1 = [P2, P3]$$

$$C2 = [P4, P5]$$

- tym samym, spodziewana postać danych wyjściowych przybiera formę:

0 0

0 1

1 2

1 3

2 4

2 5