

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

Fakulta informatiky a informačných technológií
Ilkovičova 2, 842 16 Bratislava 4

Zadanie 4b – klastrovanie

Riešený problém

V 2D priestore, ktorý je v intervaloch od -5000 do +5000. Treba vyplniť najprv 20 náhodne vygenerovanými bodmi, každý bod má unikátne súradnice.

Následne vygenerujeme ďalších 20000 bodov podľa nasledujúcich kritérií:

1. Náhodne vyberte jeden zo všetkých doteraz vytvorených bodov v 2D priestore. Ak je bod príliš blízko okraju, tak zredukujete príslušný interval v nasledujúcich dvoch krokoch.
2. Vygenerujte náhodné číslo X_{offset} v intervale od -100 do +100
3. Vygenerujte náhodné číslo Y_{offset} v intervale od -100 do +100
4. Pridajte nový bod do 2D priestoru, ktorý bude mať súradnice ako náhodne vybraný bod v kroku 1, pričom tieto súradnice budú posunuté o X_{offset} a Y_{offset}

Následne musíme implementovať zhlukovač pre 2D priestor, ktorý zanalyzuje 2D priestor so všetkými 20020 bodmi a rozdelí ich na „k“ zhlukov (klastrov).

Treba implementovať zhlukovače s týmito algoritmami:

- k-means, kde stred je centroid
- k-means, kde stred je medoid
- aglomeratívne zhlukovanie, kde stred je centroid
- divízne zhlukovanie, kde stred je centroid

Musí sa vyhodnocovať úspešnosť/chybovosť zhlukovača. A za úspešný zhlukovač sa považujeme taký, v ktorom žiaden klaster nemá priemernú vzdialenosť bodov od stredu viac ako 500.

Vizualizácia: pre každý z týchto experimentov vykreslite výslednú 2D plochu tak, že označujete (napr. vyfarbíte, očísľujete, zakrúžkujete) výsledné klastre.

Základné informácie

Program vyhotovený v jazyku Python vo verzii 3.8.6

Program je spustiteľný ako .exe

Použité knižnice:

- random
- time
- numpy
- matplotlib (pip install matplotlib)
- math

Pre nastavenia používam globálne premenné:

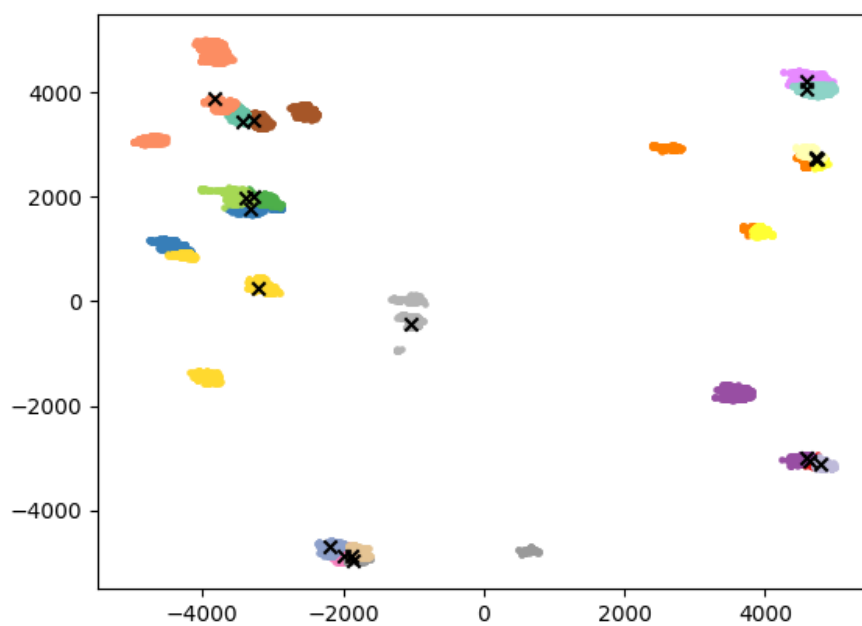
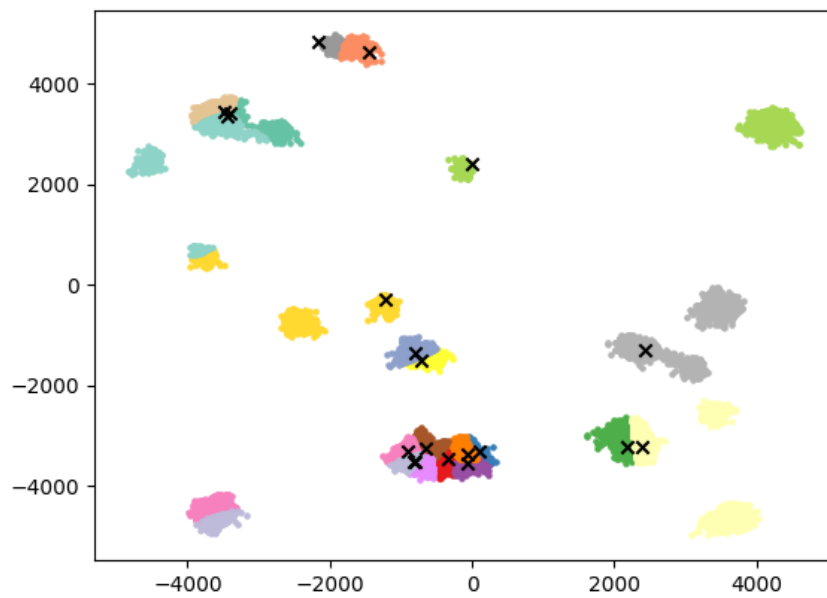
```
#NASTAVENIA
max_iterations = 300
global_interval = (-5000,5000)
global_count_random_points = 20
global_count_points_to_generate_using_offsets = 20000
global_x_offset_interval = (-100,100)
global_y_offset_interval = (-100,100)
global_number_of_cluster = 20
```

-vytvorený slovník s názvom „colors“, kde je 20 farieb, ak sa dá väčší počet klastrov treba zväčšiť toto pole aby bolo viac rozdielných ofarbení klastrov, prípadne implementovať funkciu na výber farby.

Opis riešenia

Opis algoritmov

Spôsob testovania



Zhodnotenie riešenia