

## Zadanie 4b – klastrovanie

Máme 2D priestor, ktorý má rozmery  $X$  a  $Y$ , v intervaloch od  $-5000$  do  $+5000$ . Tento 2D priestor vyplňte 20 bodmi, pričom každý bod má náhodne zvolenú polohu pomocou súradníc  $X$  a  $Y$ . Každý bod má unikátne súradnice (t.j. nemalo by byť viac bodov na presne tom istom mieste).

Po vygenerovaní 20 náhodných bodov vygenerujte ďalších 40000 bodov, avšak tieto body nebudú generované úplne náhodne, ale nasledovným spôsobom:

1. Náhodne vyberte jeden z existujúcich bodov v 2D priestore
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  $Y\_offset$

Vašou úlohou je naprogramovať zhukovač pre 2D priestor, ktorý zanalyzuje 2D priestor so všetkými jeho bodmi a rozdelí tento priestor na  $k$  zhukov (klastrov). Implementujte rôzne verzie zhukovača, konkrétne týmito algoritmami:

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

Vyhodnocujte úspešnosť/chybovosť vášho zhukovača. Za úspešný zhukovač 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.

V závere zhodnoťte dosiahnuté výsledky ich porovnaním.