

### Ćwiczenie 3. Grupowanie

Poniższe polecenia należy wykonać dla zbiorów danych zawartych we wskazanych plikach.

Przydział danych do zespołów można znaleźć w pliku "przydziały.docx".

#### Polecenia:

- (2pt)** Dla danych z pliku "p#.csv" przeprowadzić "odręcznie" (tzn. na kartce, z wykorzystaniem arkusza Excela itp, ale nie w jęz. R) obliczenia za pomocą algorytmu k-średnich.
  - Pokazać szczegółowe wyniki (odległości, grupy, wartości centroidów) otrzymane w 3 kolejnych iteracjach algorytmu (podobnie jak to zostało pokazane w prezentacji). W obliczeniach użyć liczbę grup wyznaczoną na podstawie wykresów rozrzutu. Obliczenia należy rozpocząć od **rozwiązania początkowego**, które dane jest w pliku „initial centroids.txt” .
  - Wyznaczyć wartości miar:  $SSE$ ,  $SSB$ ,  $TSS$  i  $SSB/TSS$ .
- (2pt)** Przeprowadzić obliczenia z wykorzystaniem algorytmu k-średnich dla danych z pliku "f#.csv" używając języka R.
  - Na podstawie wykresu rozrzutu określić możliwą liczbę grup.
  - Obliczenia wykonać 5 razy (5 przebiegów algorytmu). Zapisać wartości  $SSB/TSS$  dla każdego przebiegu algorytmu. Wskazać wartość największą, najmniejszą oraz średnią po wszystkich przebiegach.
  - Zaznaczyć najlepsze uzyskane rozwiązanie na wykresie rozrzutu. Wskazać centroidy.
- (1pt)** Plik "zoo.csv" zawiera zmienne opisujące 7 typów zwierząt: ssaki, ryby, ptaki, bezkręgowce, owady, płazy i gady. Wyznaczyć grupy w zbiorze danych "zoo.csv" używając algorytmu k-średnich. Porównać otrzymane wyniki ze znaną klasyfikacją dostępną w pliku "zoo\_full.xlsx" (w ostatniej kolumnie "type"). Które typy zwierząt były najczęściej mylone?