

# Klasteryzacja

Celem laboratorium jest zapoznanie się z algorytmami klasteryzacji k-means, DBSCAN oraz klasteryzacją spektralną.

Na wynik działania algorytmu k-means wpływa głównie parametr

- liczba klastrów – `n_clusters`.

Z kolei na działanie algorytmu DBSCAN wpływają głównie dwa parametry, definiujące razem pojęcie gęstości:

- promień sąsiedztwa – `eps`,
- wymagana liczba punktów w sąsiedztwie, aby punkt został uznany za rdzeniowy – `min_samples`.

Na wynik działania algorytmu klasteryzacji spektralnej wpływ mają parametry

- liczba klastrów – `n_clusters`,
- współczynnik jądra RBF – `gamma`.

**Zadanie 1.** Celem zadania jest zbadanie, jak algorytmy k-means, DBSCAN oraz klasteryzacja spektralna radzą sobie ze zbiorami o różnej naturze. Wykorzystamy syntetycznie generowane zbiory:

- blobs
- circles
- moons
- ellipses

Zbadaj wpływ podanych parametrów algorytmów na kształt uzyskanych klastrów.

**Zadanie 2.** W zadaniu należy wykorzystać zbiór `banknotes`. Pierwsza kolumna zbioru wskazuje, czy banknot jest oryginalny czy sfalszowany. Pozostałe sześć kolumn opisuje cechy konkretnych banknotów, takie jak długość czy szerokość. Zastosuj algorytmy k-means i DBSCAN do pogrupowania banknotów w klastry. Ponieważ oba algorytmy wykorzystują pojęcie odległości euklidesowej, na początku znormalizuj wartości cech.

W przypadku algorytmu k-means oszacuj optymalną liczbę klastrów, korzystając z:

- wskaźnika sylwetkowego (ang. silhouette score),
- kryterium informacyjnego Akaikego (ang. Akaike Information Criterion).

*Uwaga.* Nie należy stosować reguły *elbow*.

W przypadku algorytmu DBSCAN sprawdź, czy wszystkie banknoty zostały przypisane do klastrów.

Czy wyniki klasteryzacji są zgodne z podziałem na banknoty oryginalne i sfalszowane? Oceń jakość otrzymanych klastrów, korzystając z miar <sup>1</sup>:

- homogeniczności (ang. homogeneity),
- zupełności (ang. completeness),
- V-miary (ang. V-measure).

---

<sup>1</sup><https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation>